Summer 8-18-2023

Scaling Up the Relevance of Land-Sea Connections in Coastal Bacteria Pollution Vulnerability

Bea E. Van Dam
University of Maine, bea.vandam@maine.edu

SCALING UP THE RELEVANCE OF LAND-SEA CONNECTIONS IN COASTAL BACTERIA POLLUTION VULNERABILITY

By

Bea Elizabeth Van Dam

B.S. University of Maine, 2011

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

(in Earth and Climate Sciences)

The Graduate School

The University of Maine

August 2023

Advisory Committee:

Sean M.C. Smith, Associate Professor of Earth and Climate Sciences, Advisor

Kate Beard, Professor of Spatial Informatics

Andrew Reeve, Professor of Earth and Climate Sciences

Lauren Ross, Assistant Professor of Civil and Environmental Engineering

Matthew Baker, Professor of Geography and Environmental Systems (University of Maryland, Baltimore County)

SCALING UP THE RELEVANCE OF LAND-SEA CONNECTIONS IN COASTAL BACTERIA POLLUTION VULNERABILITY

By Bea Elizabeth Van Dam

Dissertation Advisor: Dr. Sean M.C. Smith

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
(in Earth and Climate Sciences)
August 2023

Bacteria pollution closures of Maine's coastal shellfish harvest areas have substantial negative consequences for coastal businesses and communities. Sustainability solutions for Maine's shellfish harvesting areas and businesses require new types of knowledge and information to protect water quality and public health while avoiding unnecessary fishery closures. Coastal management agencies have interests in tools to support science-based management decision-making related to pollution and sustainability solutions for businesses and communities.

Prior research into land-sea connections has demonstrated uses of geographic information and statistical methods to facilitate management and science communication. Research in Maine has focused on identification and comparison of attributes influencing coastal conditions. Examinations of coastal settings based on proxy spatial data metrics for pollution sources, delivery, and residence time (SDR) attributes have demonstrated capacity to identify locations with varied pollution vulnerability when paired with water quality sampling data.

This research starts from the proof of concept from previous work and cogeneration of knowledge with stakeholders in Maine. Advancements include the strategic process for selecting and assembling proxy spatial data metrics, procedures to identify coastal pollution response units (CPRUs), and approaches used to document associations of CPRU settings with pollution problems. Outcomes include delineations of land-sea connection domains and identification of seven CPRU setting types. Results indicate similarity among locations derived from proxy metrics and bacteria sampling data based on selected pollution attributes and equal weighting of SDR attribute categories. Lands adjacent to tidal boundaries, "margin watershed areas" (MWAs), comprise 9.8% of the CPRU land area. However, MWAs were not found to increase the predictability of vulnerability to bacteria pollution.

Multiple information gaps are assumed to influence results and limit direct applications from the analyses, including: 1) Biases in bacteria sampling from management activities, 2) Static nature of proxy metrics describing land-sea connection processes, 3) Domain outlet specifications, 4) Influence of large river flows and ocean input, 5) Stochastic events, 6) Equal weighting of SDR pollution culprit categories. However, research outcomes provide a defensible framework for coastal pollution vulnerability evaluations, guidance for targeting pollution problems, and new information to support research and management decisions related to coastal planning and monitoring activities.

ACKNOWLEDGEMENTS

Funding support for this research has been provided by the Maine Agricultural and Forest Experiment Station (MAFES) Project No. ME0ME022209 and by the Maine Water Resources Research Institute (WRRI) Project No. 2018ME331B and FY21 Project No. 5520980-5409118-20.

A dissertation is never a solo endeavor, and the work presented here would not have been possible without the contributions of many people over the past several years. Firstly, I would like to thank my advisor, Sean Smith, for guidance and opportunities over nearly a decade of working together that have shaped me as a researcher. Grateful thanks are extended to my entire committee, whose input has challenged me and strengthened this research.

I would also like to gratefully acknowledge project partnership and support from the Senator George J. Mitchell Center for Sustainability at the University of Maine under the direction of David Hart. The Maine Department of Marine Resources Bureau of Public Health and Aquaculture were key stakeholders in this project whose collaboration provided expert advice and helped guide research inquiry in unexpected and fruitful directions. I would like to offer particular thanks to Bureau director Kohl Kanwit and growing area supervisors Bryant Lewis and David Miller.

Thank you as well to current and former colleagues and collaborators for input that shaped this research, including Brett Gerard, Sam Roy, Sohaib Alahmed, Hannah Horecka, Damian Brady (Darling Marine Center), and Stephen Jones (UNH), as well as other members of the Watershed Process and Estuary Sustainability Research Group (WPES) here at the University of Maine.

Finally, and especially, thank you to my family and friends for unwavering support throughout this process. I would not be where I am today without any of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

BACKGROUND AND INTRODUCTION

1.1. Introduction

Decision science, the scientific basis for decision-making when addressing complex or intractable problems, is a growing field (Marcot et al., 2012; Barnes et al., 2019; Yahara, 2021; Baker et al., 2022; R. Smith et al., 2022; Martin et al., 2023). Decision science in natural resource management targets the connection of data and knowledge of biophysical processes to sustainability goals and objectives. The inspiration for the research is often to parameterize problem spaces and identify optimal management solutions (Figure 1.1). Scientific knowledge and theory are thereby linked to policy needs and objectives to identify a problem domain, specify metrics to describe the problem, analyze relations between metrics, and develop decision support tools. The interdisciplinary research activities are completed in a process that involves knowledge co-generation with collaborators and stakeholders (Kates et al., 2001).

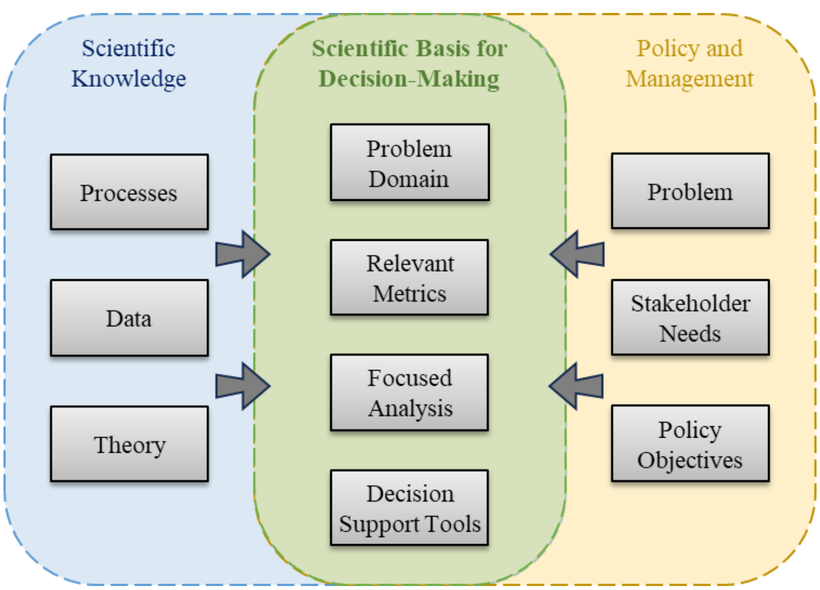

Figure 1.1. A decision science framework diagram.

Research supported by the National Science Foundation (NSF) has focused on strengthening of the scientific basis for decision-making related to sustainability solutions in New England in recent years (National Science Foundation, 2015, 2016). One body of work was attentive to pollution problems governed by complex land-sea connections and coupled human-environmental system dynamics with implications to shellfisheries and coastal communities.

The problem of coastal bacteria pollution and water quality in the northeastern United States is not new. The National Shellfish Sanitation Program (NSSP) was introduced in 1925 in response to outbreaks of illness related to consumption of shellfish that was ultimately attributed to sewage-polluted oysters (US Food and Drug Administration, 2019). The knowledge that water quality at the coast is greatly affected by land-sea connections is similarly well-established. Pollutants originating on the landscape are delivered to coastal waters through mechanisms such as precipitation runoff and wastewater conveyances, often remaining in those coastal waters until they are evacuated by tidal currents and freshwater runoff dynamics (Alonso Roldán et al., 2019; Brown et al., 2019). The Clean Water Act of 1972 was the landmark law regulating the discharge of pollutants into navigable waters in response to fundamental observations of pollution-related societal problems. The legislation had roots established earlier in the twentieth century in the 1948 Federal Water Pollution Control Act (US EPA, 2013) and evolved into regulatory responses to both point and nonpoint source pollution.

The importance of land-sea connections to water quality conditions and ecosystem sustainability has been generally recognized in policy through the framing of a "coastal zone" and related governance tools (Maine Department of Marine Resources, 2023a; NOAA Office for Coastal Management, no date). However, the delineation and characterization of such zones has thus far been jurisdictional rather than based on physical processes such as the distinction between overland and channelized runoff transport. Coastal management agencies such as the Maine Department of Marine Resources (MEDMR) remain constrained by the predictive capacity of current decision-making tools to assess how land-sea connections manifest in varied coastal settings.

The causes and effects of water pollution are commonly tied to the concept of vulnerability. The vulnerability of a place or community to pollution problems is complicated by land-sea connection dynamics coupled with human activities. Vulnerability in a specified domain is governed by complicated interactions of processes that act upon heterogeneous physiographic spaces. Human activities on the landscape and in coastal waters may both contribute to and ameliorate these dynamics.

Turner et al. (2003) identify several elements important to performing vulnerability analyses relevant to sustainability. These include consideration of the following: a) sequencing of multiple stressors, b) sensitivity of coupled systems to exposure to hazards, and c) nested scales of hazards, coupled systems, and dynamics. They also suggest emphasis on elements useful to vulnerability analyses for sustainability purposes, including the following measures: i) elevation of stakeholder input to identify responses to pollution outcomes that should be avoided, ii) profiling of unequal vulnerability of system subcomponents, and iii) identification of suspect causal structures that affect vulnerability, and iv) development of methods for assessments and tests. Accordingly, the research described in this dissertation focuses on the development of the problem domain, metrics, and spatial data and the formation of vulnerability analyses to create an expert system framework for scientific assessment of coastal pollution affecting estuarine shellfisheries to support management decision-making.

1.2. Research and Management Problems

Estuarine bacteria pollution is a multifaceted problem, making it complex to manage. It is first a biophysical problem. Maine’s clams, mussels, oysters, and scallops are filter feeders that derive sustenance by drawing water through their gills and trapping and ingesting plankton and other microscopic particles present in the water. As such, they are predisposed to ingestion of potentially harmful substances such as microplastics (Woods et al., 2018; Zhao et al., 2018), toxic algae (Thomas et al., 2010; Clark et al., 2019), toxic chemicals, and bacteria (Jones, 2011). A particular concern related to shellfish harvested for human consumption are feces-borne pathogens that may be present in coastal waters in association with fecal coliform bacteria and can cause severe illness in consumers (Jones, 2011). Because of these dangers, bacteria pollution represents a coastal sustainability problem that threatens a traditional food source and harvesting industry with economic implications for harvesters and coastal communities.

The MEDMR is responsible for overseeing the safety of intertidal shellfish resources, including making the decision to close harvesting areas when potentially harmful levels of bacteria are in the water. Policy challenges can arise due to competing goals between reopening shellfisheries quickly to minimize impact on harvesters and ensuring closures are long enough for bacteria evacuation so that consumer health is not compromised. Bacteria pollution is therefore also a knowledge and information problem. Since bacteria testing cannot be performed at all times in all places, managers must have adequate knowledge and information about varied coastal settings to make informed, scientifically-based decisions surrounding bacteria pollution closure timing.

1.3. Background and Study Area

The Maine coastline is at the western boundary of the Gulf of Maine, a ~93,000 km^2 body of water on the North Atlantic Ocean (Figure 1.2). Maine's coast is deeply indented, a result of the interaction between bedrock controls and Late Wisconsinan glaciation that induced isostatic sea level adjustments (Schnitker, 1974; Kelley, 1987; Borns et al., 2004; Uchupi and Bolmer, 2008; Kelley, Belknap and Claesson, 2010). The modern coastline features a total of ~5,600 km of tidal shoreline (including islands) over less than 400 km of straight-line distance between the borders with New Hampshire and Canada (Kelley, 1987). Nestled among the many embayments and tidal rivers of the rocky coastline is an extensive series of sand- and mudflats formed from reworked outwash and marine sediment, which are regularly inundated and exposed by semidiurnal tides with mean ranges varying from 2.5 m at Kittery in the south (NOAA station #8419870) to 5.6 m at Eastport in the northeast (NOAA station #8410140). These intertidal zones are home to a variety of wildlife species, including molluscan shellfish.

Shellfishing, like logging, boat building, and lobstering, is a quintessential traditional Maine industry with social and economic impacts in the coastal communities that practice it and the state at large (Ellis and Waterman, 1998). The harvesting of mussels, oysters, and various species of soft- and hard-shell clams provides a low barrier income source for more than 1500 licensed harvesters who ply Maine's tidal flats, bringing tens of millions of dollars of direct revenue (Evans et al., 2016; Maine Department of Marine Resources, 2019b). Maine's commercial soft-shell clam fishery, the most valuable of the state's mollusk fisheries, accounted for over 70% of total 2015 US domestic soft-shell landings and 79% of total US commercial landings value at more than $22.8 million (National Marine Fisheries Service, 2016; Maine Department of Marine Resources, 2019b).

In addition to the direct benefits to harvesters and communities, income from shellfishing has a cascading effect on the economy. In 2006, a year in which direct output value of all Maine molluscan shellfishing was $29.9 million and direct labor income to harvesters was $21.5 million, the estimated total economic impact on the Maine economy was $56 million (Athearn, 2008). Shellfishing also represents an aspect of cultural heritage on the Maine coast, where harvesters may be engaging in a multi-generational family tradition.

Bacteria pollution-related closures of mud flats can reduce the economic vitality of shellfishing businesses. Most Maine shellfishers are independent harvesters (Evans et al., 2016). It is not an industry in which harvesters draw salaries or can even be sure of steady paychecks, amplifying the effects of

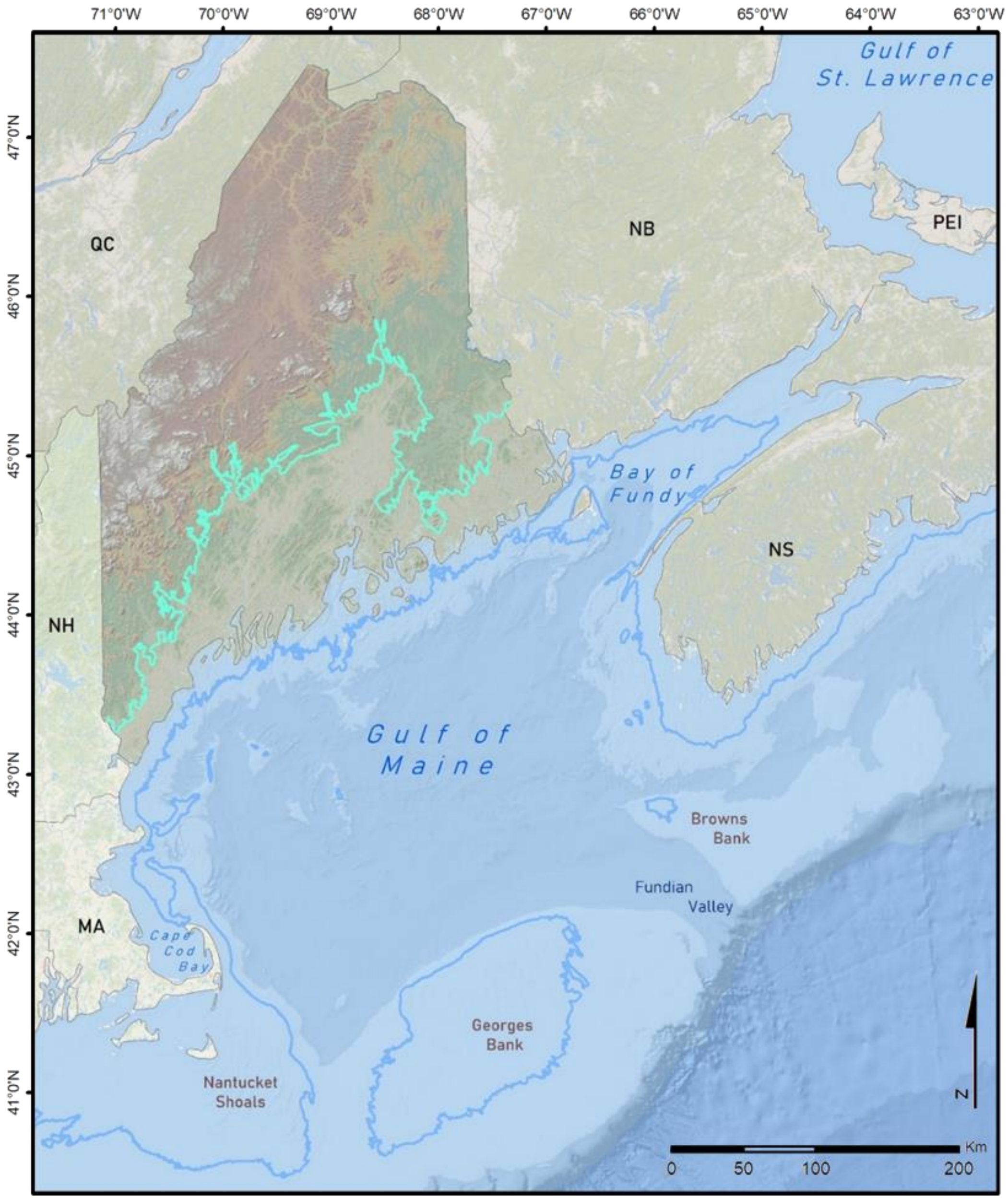

Figure 1.2. Gulf of Maine and surrounding land, with marine high stand transgression line within Maine (foam green) and modern 60 meter bathymetry contour (blue) representing general approximation of shoreline at marine low stand. (Map background adapted from Esri World Ocean Base, other data from Maine GeoLibrary).

closures on their daily earnings. The work is physically demanding – "hard on the whole body" – and it is not uncommon for shellfish harvesters to seek treatment for chronic pain, leading to an estimated opiate addiction rate of at least 10-20% among harvesters (McGreavy, 2015). This further intensifies the closure effects on coastal individuals and communities. The implications of closures to shellfishing businesses and communities provides a strong rationale for management decision tools to guide deployment of monitoring resources and avoid unnecessary closures.

Bacteria and pathogens originate at point and non-point sources on the landscape and within estuaries, and the hydrologic connection between landscape and coastal waters facilitates the transport of these pollutants into shellfishing areas via runoff (Jones, 2011). Identification of particular sources has been accomplished through DNA source tracking in individual Gulf of Maine estuaries. A DNA marker analysis of fecal coliform bacteria in one southern Maine system found ubiquitous incidence of avian droppings throughout the sampling period and spikes in human DNA at development-adjacent sites during months with peak seasonal populations (Sims and Kaczor, 2017). This study and others that are similar indicate that while bacteria sources within a watershed can be multiple and diverse, some may be positively correlated with particular land use characteristics and thus predictive of vulnerability to high bacteria levels in their associated embayments.

1.4. Prior Work

1.4.1. Management

Coastal managers must be aware of the varied coastal conditions in the state's widely scattered shellfishing flats in order to meet the objective of keeping temporary pollution closures to the minimum lengths required to ensure consumer safety. This management requirement inherently involves consideration of the factors influencing bacteria production, transport, and estuarine residence time. The sustainability targets can be advanced through attention to the elements identified by Turner et al. (2003), as well as through consideration of how coastal attributes driving land-sea connections operate independently and collectively to produce water pollutions outcomes.

The MEDMR does not have capacity to continuously monitor bacteria levels at every mud flat in the state. Managers proactively and automatically execute temporary mudflat closures after high-magnitude rain events based on general knowledge of the links between precipitation runoff and estuarine bacteria pollution and with National Shellfish Sanitation Program (NSSP) practices and standards. A closure length of two weeks after a precipitation event of two inches or more in twenty-four hours is a general rule of thumb.

Mapping of shellfish closure areas and monitoring outcomes has been a consistent method through which coastal management agencies have communicated to the public (Figure 1.3). The maps that are now online generally respond to agency knowledge of conditions and history of bacteria sampling. They contain some assumptions related to land-sea connection processes anchored on associations of bacteria sources and delivery mechanisms. Outflows of wastewater from pipe outfalls and increased runoff production and delivery from urban landscapes are two prominent examples (US Food and Drug Administration, 2019). Ambitions to advance the consideration of land-sea connection processes in vulnerability analyses inspired recent research on pollution management decision tools in Maine coastal areas. Part of that pursuit is related to expanded capacity to consider water pollution attributes operating independently and collectively. Another part relates to interest in a more detailed delineation of the coastline relative to land-sea connection processes influencing pollution outcomes and potentially guiding management solutions.

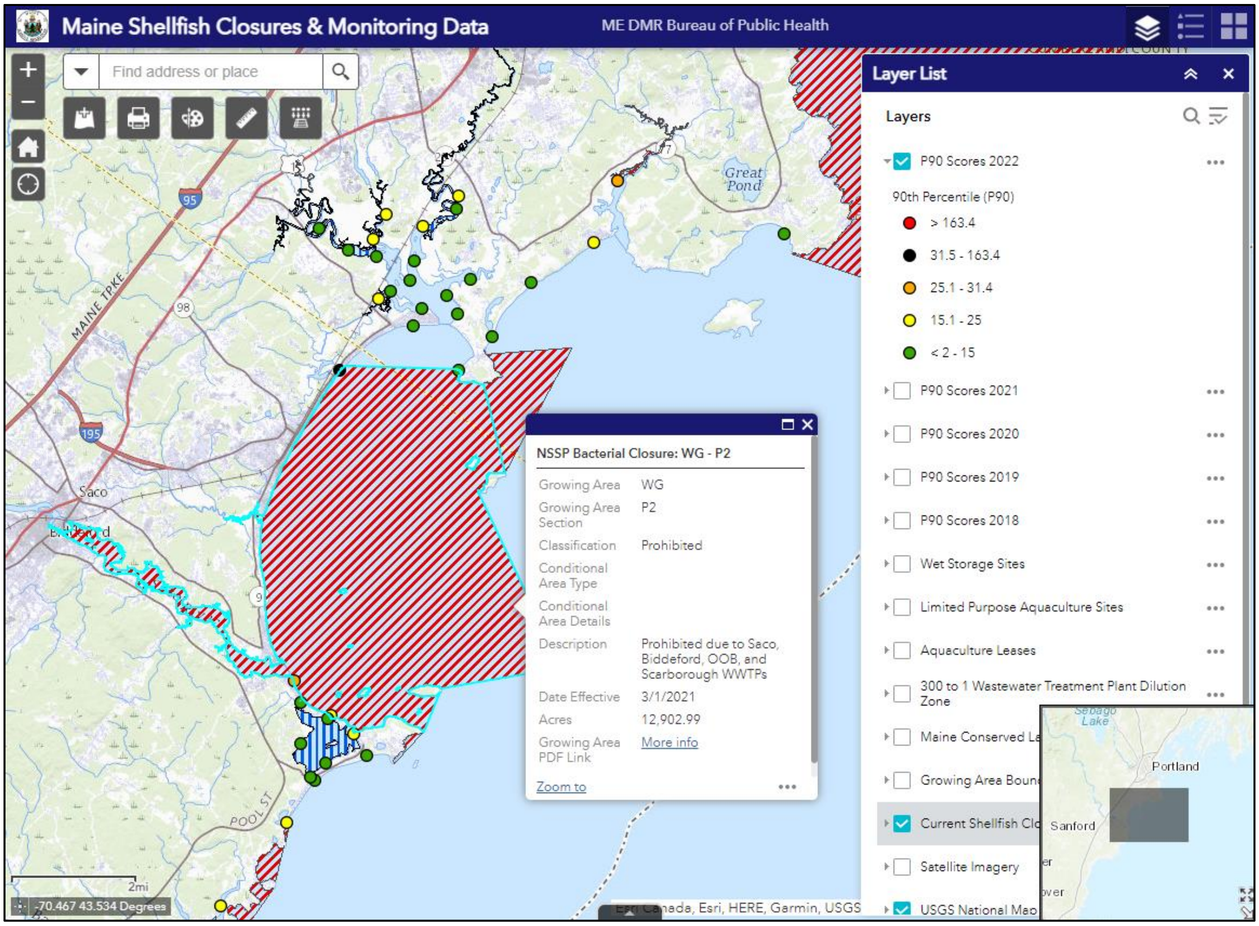

Figure 1.3 Screen capture of Maine Dept. of Marine Resources (MEDMR) online shellfish closures and monitoring data map showing the area of Saco Bay, south of Portland, ME. Red hatched areas indicate growing areas where shellfishing is prohibited; growing area WG-P2 highlighted here is closed due to the influence of discharge from wastewater treatment plants (WWTPs) from three surrounding towns. Blue hatched areas indicate growing areas conditionally open to shellfishing, and coastal areas with no hatching have no harvesting restrictions. Circles represent MEDMR bacteria monitoring locations. Colors indicate 90th percentile (P90) fecal coliform counts for samples taken at each site over the year 2022 as shown in the legend at upper right. (Maine Department of Marine Resources, 2023c)

1.4.2. Research

Research into land-sea connections and identification of coastal settings has taken varied forms depending on research goals and associated coastal management needs. Research in the Maryland portion of Chesapeake Bay with a goal of better defining the relations between climate factors and fecal coliform bacteria focused on linear regressions and the clustering of temporal data groupings. Results revealed that precipitation and air temperature are good predictors of bacteria levels with seasonally variable strengths of climate-bacteria relations (Leight et al., 2016).

Detailed modeling has been used in locations such as Hawai'i, where coupled groundwater and coral reef models incorporating nutrient flux and wave power were developed and calibrated for individual sites to identify priority management targets for reef resilience (Delevaux et al., 2018). Hydrodynamic numerical modeling of pollutant particle transport in Galveston Bay, TX revealed that the timing of pollutant release into estuarine waters in relation to storm events is an important factor in determining pollution flushing times.

In contrast to the detailed modeling approach, Bartley et al. (2001) tested the applicability of a single proxy metric for marine and terrestrial forcing processes, coastline complexity (map-view tortuosity), to classify coastal settings. Their case study in Baja California suggested that a global coastline dataset could be used to identify clusters of coastline type based on complexity at multiple scales. Other unsupervised cluster analyses based on different sets of embayment morphometry and hydrologic attributes resulted in the identification of nine emergent estuarine setting types for large estuaries in the contiguous United States (Engle et al., 2007) and the Australian state of Tasmania (Edgar et al., 2000).

Other setting identification has been performed using decision tree analysis. Hume et al. (2007) proposed a multi-level classification tree for New Zealand estuaries for conservation management combining expert judgment with environmental factors at scales ranging from global climate variations to local catchment properties. Van Niekerk et al. (2020) similarly developed a rules based ecosystem classification tree for large and small estuaries in South Africa based on regional setting, estuary morphometry and hydrodynamics, and biota. Both studies identified twenty or more estuary settings, and each relied on expert judgment and interpretation for developing classification rules. Decision trees have also been developed as management support tools for identifying how important land-based drivers of pollution are in a particular system for prioritizing conservation strategies (Fredston-Hermann et al., 2016).

Research in coastal Maine over the past decade has assessed coastal landscape conditions and developed decision support tools in accordance with management needs related to bacteria pollution (Taylor, 2018a). The previous work has had a primary goal of identifying a scientific basis for making management decisions regarding bacteria pollution closures along the Maine coast. Researchers and stakeholders identified a number of available coastwide spatial datasets to serve as proxies for processes involved in bacteria production (sources), delivery methods/efficiency from landscape to estuary, and residence time before eventual evacuation from an estuary (Smith et al., 2016; Roy et al., 2018; Taylor, 2018b). These data layers were overlaid against delineated watersheds of 535 rivers and streams draining to Maine's tidal coastline and a *k*-means unsupervised cluster analysis was performed to group the 535 watersheds into a limited number of Landscape Pollution Response Units (LPRUs) with similar metric values (Gerard, 2018). Watersheds in each cluster could be expected to behave broadly similarly during and after two-inch precipitation events, allowing MEDMR to use data from monitored watersheds to make informed management decisions about an unmonitored watershed in the same cluster.

The LPRU identification served as a successful proof of concept for a coast-wide bacteria pollution vulnerability analysis using source, delivery, and residence time (SDR) proxy metrics to identify similarly-behaving landscape unit settings (Figure 1.4). The study identified five LPRUs and found they

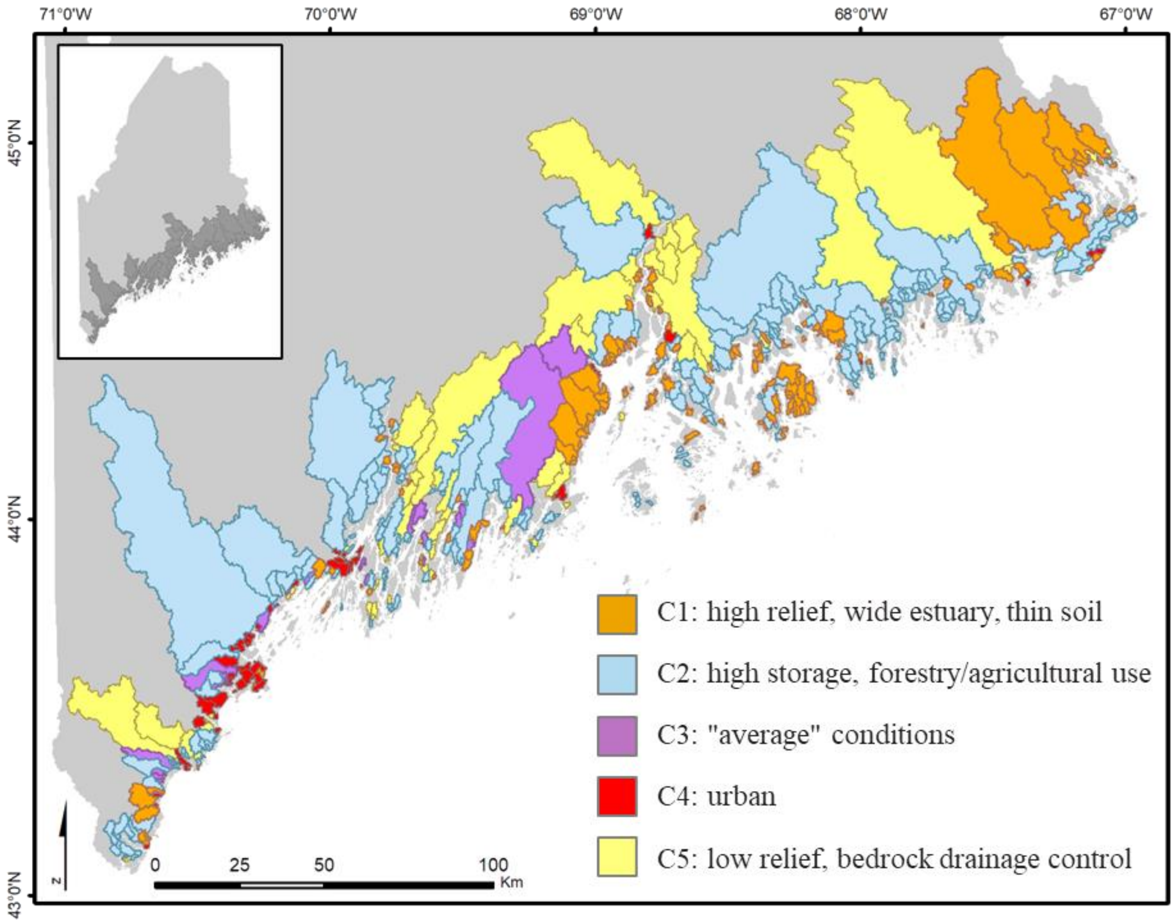

Figure 1.4. Map of Maine Landscape Pollution Response Units (LPRUs). Prior *k*-means cluster analysis of 535 nontidal watersheds based on source, delivery, and residence time proxy metrics resulted in the identification of five LPRUs for coastal Maine (adapted from Smith et al. 2016).

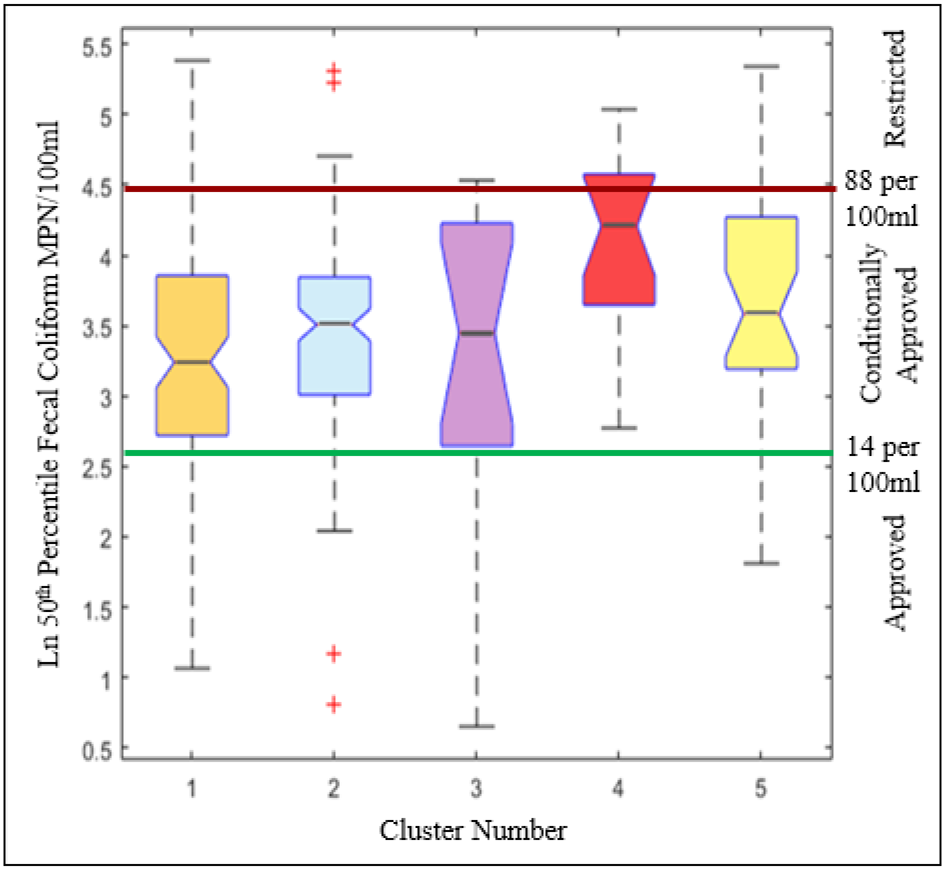

Figure 1.5. Median fecal coliform bacteria counts at watershed outlet monitoring stations, by Landscape Pollution Response Unit (adapted from Smith et al. 2016).

exhibited "statistically different bacteria pollution responses" (Smith et al., 2016; Gerard, 2018) (Figure 1.5). Regressions of single SDR proxy metrics against bacteria counts revealed that some metrics showed better correlations with high bacteria levels than others. However, no individual metric can be looked at as a reliable predictor of vulnerability to bacteria pollution in all locations. This implies that bacteria vulnerability in a coastal setting is a result of the interaction of multiple attributes related to land-sea connection processes. The research outcomes also illuminated a potentially important gap in the consideration of attributes in land areas immediately adjacent to tidal waters.

1.4.3. Research and Management Gaps

Outcomes from research in Maine and other locations revealed several prominent gaps related to domain delineations, attribute selection, and analytical approaches. A major limitation of prior coastal Maine LPRU identification was the omission of ~2,800 km^2 of coastal land not included in the delineated non-tidal watersheds. The immediate coastal landscapes between non-tidal watershed outlets, which are referred to here as "Margin Watershed Areas" (MWAs or "margins"), do not support perennial stream networks but instead contribute runoff directly as overland flow. These areas include parts of some of the state's most densely populated cities and towns, including the Portland peninsula (Figure 1.6).

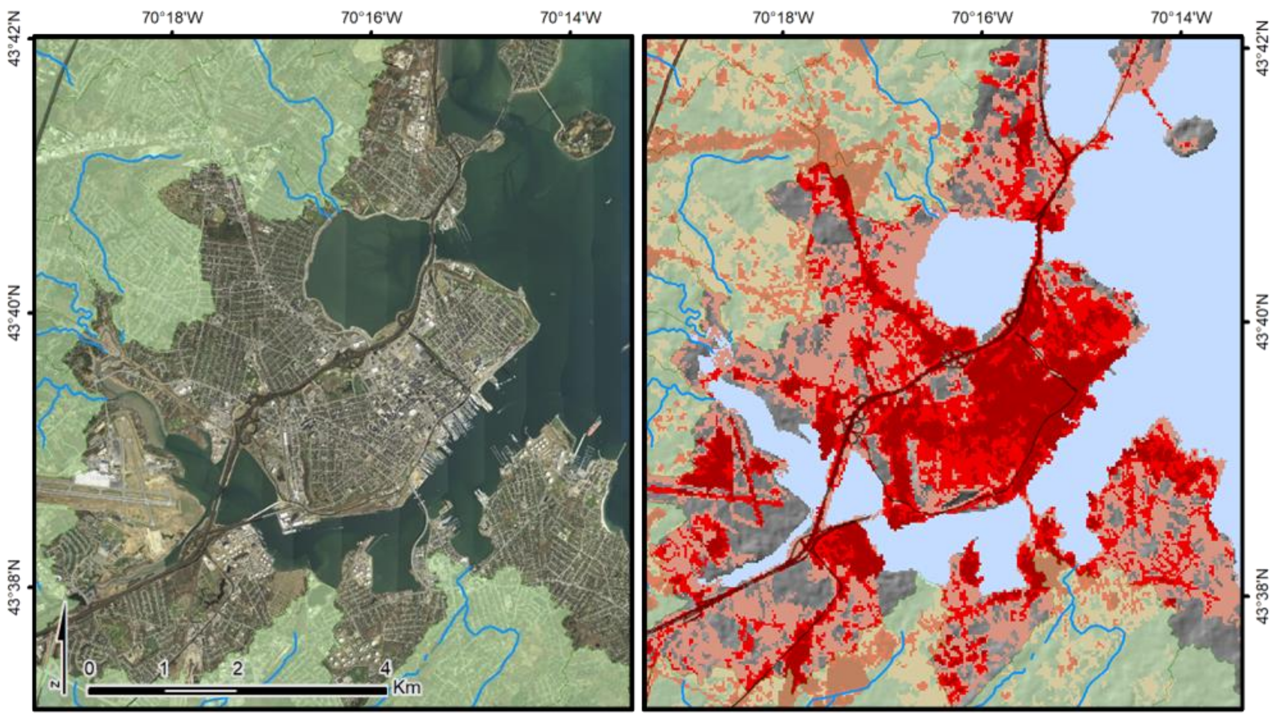

Figure 1.6. Maps of Portland peninsula showing extensive development within Margin Watershed Areas (MWAs). MWAs, the coastal regions between the boundaries of nontidal watersheds (green) and coastal waters, contribute surface runoff directly as overland flow. Red shades indicate low (lightest), medium, and high (darkest) intensity developed spaces.

1.5. Research Questions

The research described in this dissertation seeks to advance the capacity to guide coastal management decisions related to pollution affecting shellfish industries. The focus is on identification of land-sea connection attributes driving water pollution problems in varied coastal settings in Maine, selection of metrics to evaluate the drivers, and development of an expert system to evaluate vulnerability of Maine coastal settings to pollution problems.

The approach relies on spatial data and existing knowledge of land-sea connection processes related to pollution sources, watershed delivery mechanisms, and estuary residence time to guide shellfishery pollution management decisions in a sustainability solutions framework (Kates et al., 2001). Geographic information system (GIS) resources and machine learning approaches are used to parameterize coastal settings and identify land-sea connection conditions, identifying locations with similar attributes related to pollution problems, then relating each of the identified settings to bacteria pollution sampling results. The goal of this research is thereby the development of a scientifically based approach to identifying coastal setting types and where they occur in Maine. It then evaluates the associations between land-sea connection attributes operating independently and collectively in the identified settings with long term pollution problems documented from bacteria sampling by MEDMR. The research is carried out using spatial and statistical analysis, machine learning, and widely available spatial information from customized terrain analysis and widely available datasets.

The central question targeted by this research is, “How can an expert system domain, metrics, and analysis approach be framed to identify coastal settings and land-sea connections influencing pollution problems affecting shellfish harvesting areas?” An expert system designed to assess the vulnerability of varied coastal settings to bacteria pollution requires that several related practical and theoretical research questions be addressed. This provides the strategic objectives of the research summarized in this dissertation. These have been framed in the research questions listed below.

1. What are the appropriate proxy metrics related to source, delivery, and residence time that capture the processes leading to bacteria pollution problems in estuaries?
2. How can coupled land-sea connection settings be compared using spatial data information to evaluate their relative vulnerability to bacteria pollution?
3. What role do the coastal margin watershed areas play in bacteria pollution vulnerability (i.e., does proximity of certain pollution-related factors to the estuary matter)?
4. What are the implications of the estuary outlet pour line placement on the outcome from the expert system to identify coastal vulnerability in varied settings?

The dissertation is organized into five chapters summarizing research activities aligning with the above four research questions. Chapter 1 has been an introduction to the coastal bacteria pollution problem in the Gulf of Maine and land-sea connection research surrounding it. Chapter 2 describes the identification and parameterization of attributes describing pollution source, delivery, and residence time processes in coastal settings. Chapter 3 describes the development and demonstration of an expert system for aggregating and clustering coastal attributes to identify settings with similar vulnerability to bacteria pollution. Chapter 4 describes the research focused on associations of coastal pollution attributes and settings with pollution problems. An overarching summary of observations, outcomes, and conclusions is provided in Chapter 5. Explanations for the research outcomes and suggestions for future work are also provided.

CHAPTER 2

COASTAL MAINE LAND-SEA CONNECTION DOMAIN DEVELOPMENT

2.1. Introduction

There is an extensive body of literature on water pollution problem causes, predictions, and solutions in coastal areas (Dheenan et al., 2016; Tyre et al., 2023; Mallin et al., 2001; Shahidul Islam and Tanaka, 2004; Roy et al., 2018, 2016). Identifying the vulnerability of different coastal areas to water pollution problems requires the selection, creation, organization, and analysis of spatial information related to land conditions, estuary morphometry, and precipitation runoff patterns. Data and analyses typically relate to land conditions, estuary morphometry, and precipitation runoff patterns (Leight et al., 2016; Delevaux et al., 2018; Alonso Roldán et al., 2019).

Pollution problems in estuarine waters are fundamentally a relation between system inputs that generate a pollution concentration, processes that transport pollutants, and tidal hydrodynamics that influence the residence time of contaminated water. A proof of concept study using proxy spatial data metrics to represent land-sea connection attributes influencing pollution problems was demonstrated with a focus on 535 coastal Maine streams and rivers discharging into the Gulf of Maine (Smith et al., 2015, 2016; Smith, 2016). Outlet locations were based on National Hydrography Dataset (NHD) data. Coastal settings were identified based on the nontidal watersheds and corresponding attributes derived from proxy spatial data metrics representing source, delivery, and residence time (SDR) processes influencing pollution outcomes. The approach resulted in the identification of five setting types, referred to as "landscape pollution response units" or LPRUs (Figure 1.4) (Smith et al., 2016).

The LPRU analyses revealed several research gaps. One was in the specification of attributes influencing land-sea connections governing pollution problems. Another was related to a major spatial data gap in the land areas considered, specifically 2,771 km^2 of coastal land area situated between the boundaries of the nontidal watersheds and the Gulf of Maine shoreline. This portion of the coastal landscape forms a relatively narrow band along the immediate near-coast that lacks stream networks and contributes runoff to the tidal zone directly as overland flow over short distances (Van Dam, Smith and Beard, 2019). This portion of the landscape has been referred to as "margin watershed areas" (MWAs) in the context of this research. MWAs along the Maine coast comprise both ecologically important areas, with over 10% classified as wetlands, and some of the state's most densely populated and paved areas in close proximity to tidal waters, including the entire Portland peninsula.

This chapter describes the process of developing strategic revisions to the LPRU approach to evaluating settings with varied land-sea connection attributes influencing coastal pollution. Specific targets include new attributes describing SDR processes influencing pollution outcomes and new procedures for delineation of the domains of coastal settings under consideration. The settings identified with the revised approach will be referred to as Coastal Pollution Response Units (CPRUs). The delineations of the CPRUs provide a basis for comparison of "estuary units" under evaluation to capture the influence of varied coastal landscape areas and the SDR processes governing pollution vulnerability.

Contributing areas within an "estuary unit" that establishes the grain scale of CPRU analysis include nontidal stream and river watersheds that were a focus of the LPRU analysis as well as margin watershed areas and tidal embayment areas defined by estuary outlet pour lines (Figure 2.1). The intellectual structure of this research is framed around the pollution closure problem, knowledge of processes influencing the problem, gaps in the delineation and attribution of estuary units to define settings, and the design of solutions to convert knowledge of pollution drivers into actions for stakeholder applications based on a sustainability science framework (Kates et al., 2001; Turner et al., 2003; Maxwell, Hubbell and Eisenhauer, 2019; Weaver and Miller, 2019; Canfield, Mulvaney and Chatelain, 2022). Accordingly, four primary functional objectives frame the research summarized in this chapter to support

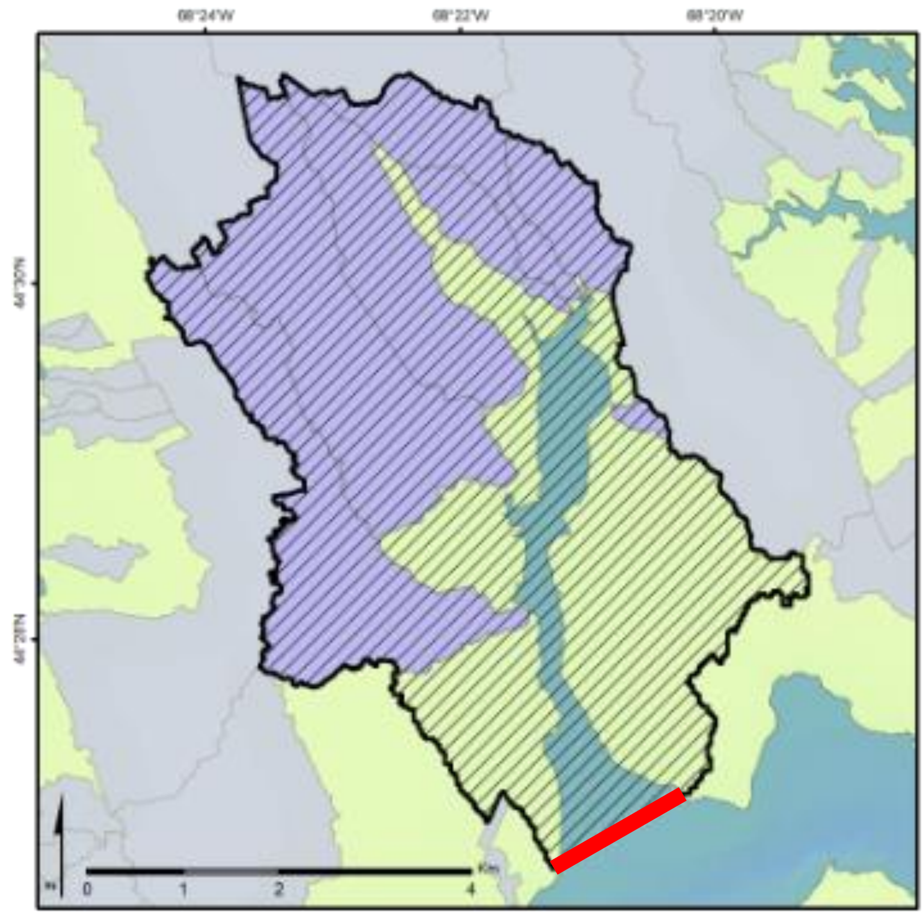

Figure 2.1 Example of an "estuary unit," the grain scale of analysis for this research. The single unit indicated with bold borders and hatching consists of nontidal watershed areas (purple), margin watershed areas (yellow), and the estuarine embayment (blue) above a defined outlet line (red). Jordan River estuary, Trenton, ME.

the overarching objective of establishing protocols to define complete land-sea connection domains (Figure 2.2).

The four functional research objectives are:

1. Identification of surface water runoff flow paths for Maine’s coastal areas.
2. Delineation of estuary units that include non-tidal watersheds and margin watershed areas based on flow path delineations, as well as estuary embayment areas.
3. Identification and assembly of spatial datasets relevant to surface water runoff and pollution sources and delivery.
4. Application of proxy source and delivery metrics to nontidal watersheds and margin watershed areas to identify and compare pollution vulnerability and culprit causes.

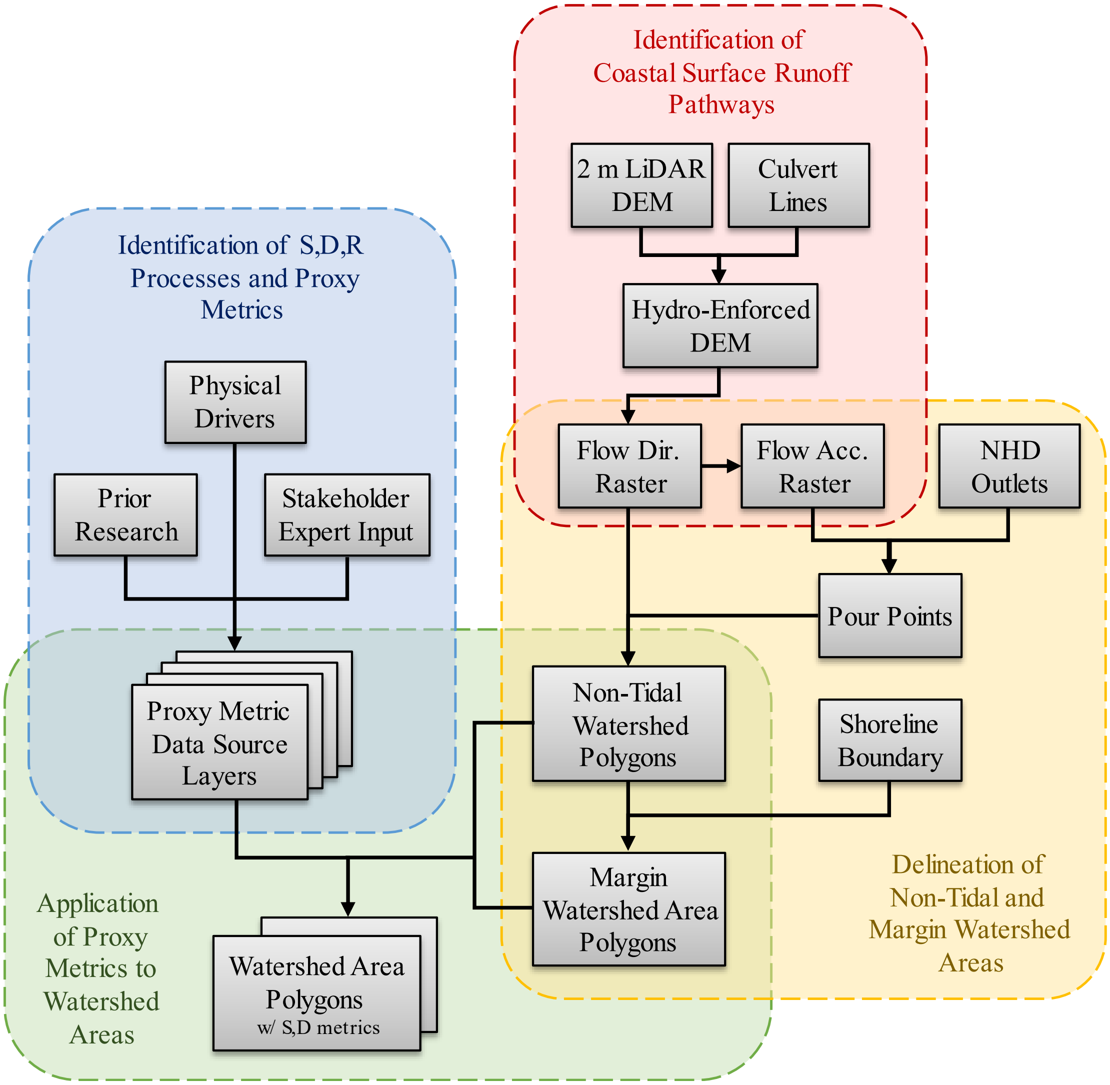

Figure 2.2. Work flow of functional objectives for definition and parameterization of the land-sea connection analysis domain.

2.2. Methods

2.2.1. Flow Pathways

The foundation for understanding coastal land-sea connections is dependent on the determination of where and how surface runoff sources influence conditions in specific estuarine water bodies. Assuming surface water runoff is the prominent bacteria transport vehicle, this necessitates identifying the problem domain and mapping the coastal landscape areas that contribute surface flows to estuaries along the tidal coastline. Sources for data layers used here are reported in Table B.1 (APPENDIX B).

Landscape surface flow path delineation is predicated on the observation that with rare exceptions, water flows down slope. Surface water runoff seeks to follow a pathway with the steepest energy gradient from areas of high hydraulic head to areas of low hydraulic head. For overland flow the pressure gradient is negligible and flow is driven by gravitational potential energy. Although flow resistance from surface roughness and obstructions can have influence on the energy gradient, the steepest gradient is typically the steepest downhill surface slope (Drucker and Williams, 2003).

Multiple algorithms have been developed to partition flow across digital elevation models (DEMs), modeled land surfaces represented by gridded elevation data (Peckham 1998; Jenson 1985; Tarboton 1997) (Figure 2.3a). One of the more commonly applied approaches is the D8 method, where each DEM cell's elevation is compared to those of its eight immediate neighbors with all flow routed out of the cell along the steepest descending slope (O'Callaghan and Mark, 1984) (Figure 2.3b, c). From D8 flow direction, a flow accumulation algorithm calculates the cumulative number of upstream cells that direct flow into each cell, facilitating visualization of flow networks (Jenson and Domingue 1988) (Figure 2.3d).

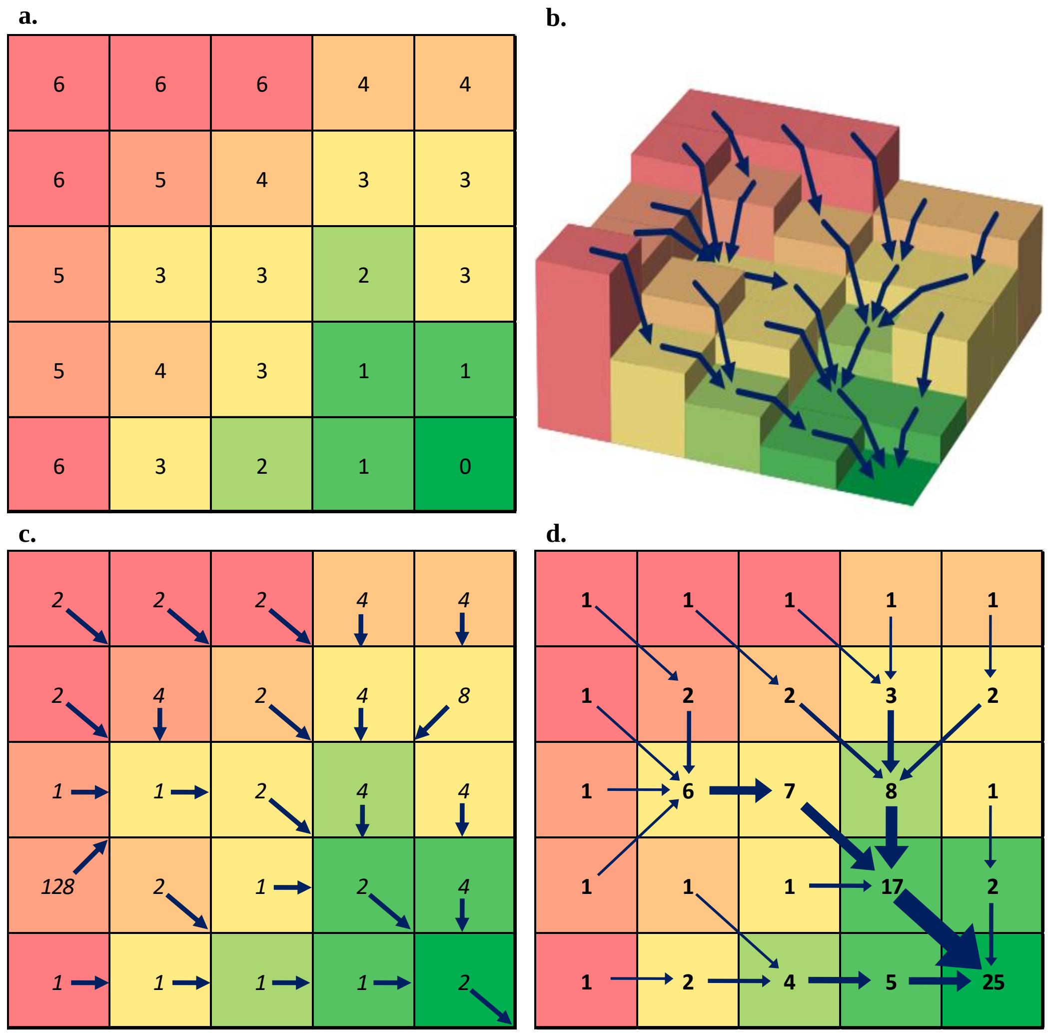

Figure 2.3. Illustration of D8 flow paths across a gridded digital elevation model (DEM). Elevations are represented by cell color throughout the figure. a) Two-dimensional map view of a DEM, with cell elevations labeled numerically. b) Three-dimensional perspective view showing arrows representing steepest flow paths from each DEM cell to one of its eight (fewer for edges) neighboring cells. c) Flow direction map for the DEM, with arrows and numeric values representing flow directions out of each cell. ArcGIS represents calculated flow directions as powers of 2, with 2^0 indicating grid east, 2^1 indicating grid southeast, 2^2 indicating grid south, and so forth for the eight directions. d) Flow accumulation map for the DEM, with arrow weight and numeric values representing total area contributing surface flow to a given cell.

2.2.1.1. Terrain Data Assembly

Boundaries: The official Maine Town and Townships Boundaries layer was used as the definitive shoreline for this research. Per the source layer metadata, this boundary corresponds to the mean high tide elevation. In practice, the actual location of the data layer boundary relative to tidal stage is not entirely consistent along the coast. Areas where the shoreline has been corrected to high-resolution satellite and elevation data are well-fitted, while original lines digitized from 1:24,000 USGS 7.5” quadrant maps have some lateral variation. For medium (e.g., York, Webhannet, Scarborough Rivers) to large (e.g. Saco, Kennebec, Penobscot Rivers) river estuaries, the inland extent of this boundary line is the river’s tidal limit, which may fall at a higher elevation.

An outer boundary of the study region into the Gulf of Maine waters was set as a simplified shell corresponding generally to the seaward jurisdictional boundary of the Maine Coastal Zone (Maine Department of Marine Resources, 2023a). The study boundary encompasses major islands off the Maine coast, with Isle Au Haut and Frenchboro at the outer limit (Figure 2.4). Some small islands beyond this limit, including Isle of Shoals, Monhegan Island, and Matinicus and Ragged Islands, were omitted despite hosting permanent or seasonal populations.

Elevation Data: Detailed flow path delineation for this project was performed using a two meter (2 m) bare earth DEM from aerial LiDAR. No hydro-conditioning and limited hydro-enforcement were performed on these data before public release (see next section). In order to ensure adequate flow path delineation for current and future near-coast research needs and check the 535 existing nontidal watershed boundaries and outlets, for which some minor discrepancies had been discovered, the inland extent of the flow path delineation area was established by creating a 1,500 m buffer around the margin area identified by Smith et al. (2016) (Figure 2.4). Due to the large size of this DEM ($2x10^9$ raster cells and 63.64 GB) and limited available computing power, it was split into twelve smaller rasters for delineation.

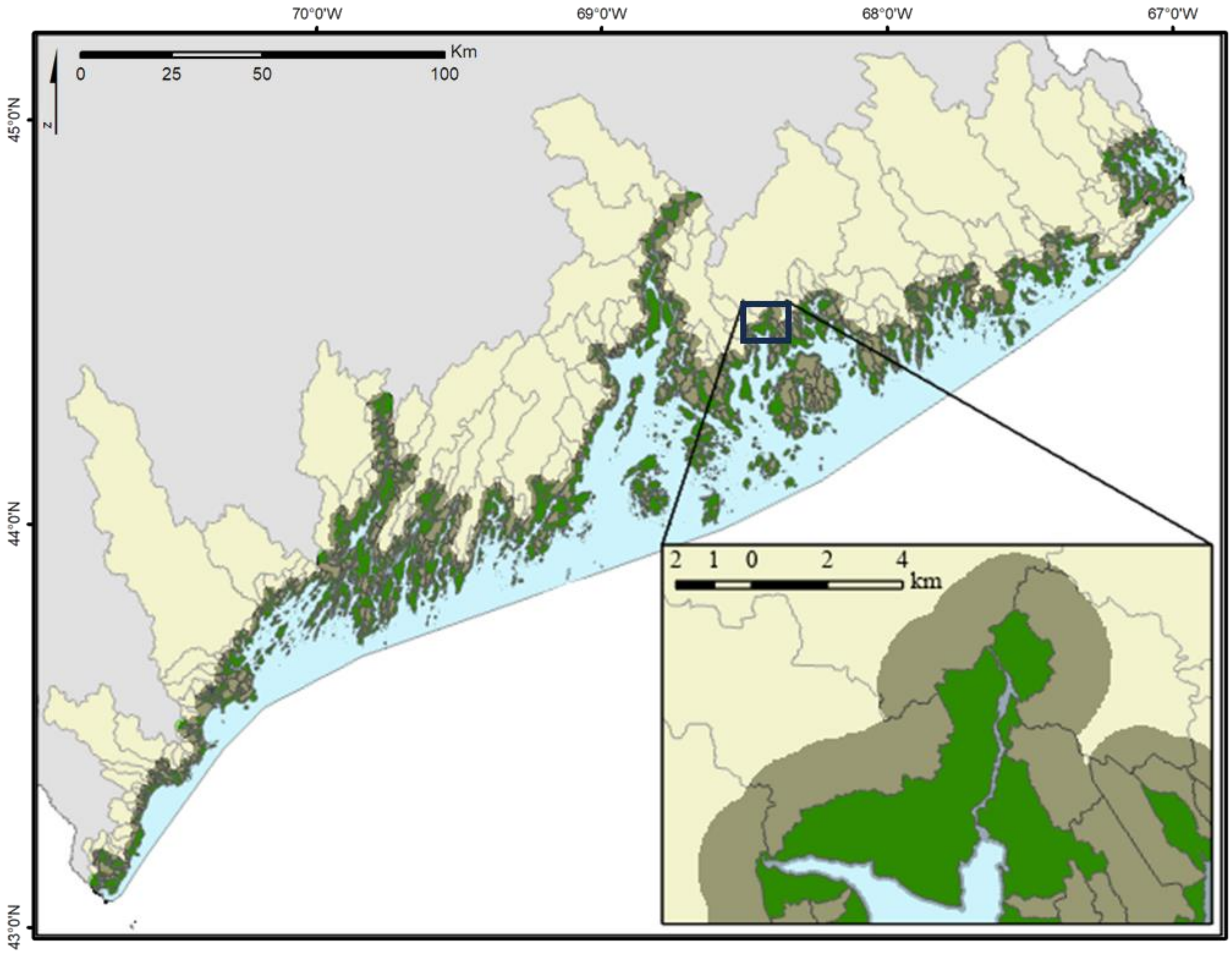

Figure 2.4. Map of Maine coast showing extent of two-meter digital elevation model (DEM) used for updated flow path delineation. The DEM was clipped to a 1.5 km buffer (shaded) inland of the margin watershed area (green), which was defined as the area between shoreline and the previously delineated 535 nontidal watersheds (yellow polygons) of Smith et al. (2016), as well as a 100 m buffer seaward of the shoreline. The coastal waters considered in the study domain are also shown (light blue polygon).

2.2.1.2. Data Preparation

Flow path delineation cannot immediately be performed on raw digital elevation models, particularly the increasingly high resolution remotely sensed DEMs now available that capture microtopography and other small storage features. The pre-processing steps of hydro-conditioning and hydro-enforcement must first be performed to remove surface irregularities and artificial impediments to flow that result in interruptions to natural hydrologic connectivity or erroneous flow delineations. Pre-processing more than 5,200 km^2 of 2 m DEM was performed using ArcMap hydrology tools and routines as described in Van Dam (2022) (APPENDIX A).

Hydro-Conditioning: Because the D8 and other flow direction algorithms execute computations cell by cell, any cell lower than all its neighbors (a "pit" cell) cannot find a downhill slope and fails to calculate a flow direction. To counteract this, bare earth DEMs must go through a process of hydro-conditioning in which all pits or depressions are filled to their spill-over levels, the elevation of the lowest neighboring cell, ensuring that every cell has a monotonically-decreasing path to a watershed outlet (Jenson and Domingue, 1988). This largely involves filling small natural local depressions captured in the DEM that would fill with water and spill over during two inch rainfall runoff events (Van Dam and Smith, 2023).

Flow Path Enforcement: A limitation of flow path delineation from surface DEMs is the inability of LiDAR and other remote sensing technology to identify or detect flow paths below solid overhanging surfaces, including under bridges and through road-spanning culverts. The locations appear as artificial dams in the DEM, causing backfills, often extensive, during hydro-conditioning and often leading to incorrect delineations as flow is routed along roadside ditches or other alternative paths (Poppenga et al., 2014). It is necessary to "burn in" culverts to correct the problem, lowering a line of cells through the artificial impediment in the DEM to allow the flow direction algorithm to correctly route flow through to the downstream end.

This process was complicated by missing culvert location data for most roads along the Maine coast. The Maine Department of Transportation (DOT) has publicly-available crossing point data for three structure types (i.e., bridges, cross culverts, and large culverts) along state roadways. However, these represent only a fraction of the total number of road-crossing culverts in the study region, making it necessary to first identify culvert locations along local (non-DOT) roads, railroads, and other engineered structures before proceeding with the burning process. A common solution to this problem is to find intersections in existing spatial data layers for roads and stream networks, but alignment issues between the available NHD stream lines and LiDAR-derived DEM terrain led to inaccurate imposition of flow paths in the study domain (Poppenga, Gesch and Worstell, 2013). Additional culvert lines were manually placed and burned into the coastal Maine DEM following the methodology of Van Dam (2022).

2.2.2. Watershed Drainage Divides

Watersheds constitute the physical spaces in which runoff processes governed by fundamental drivers deliver surface water and transported constituents, including pollutants sourced on the landscape, to coastal waters in spatially predictable patterns. As such, they represent meaningful units for consideration during spatial analyses and ecosystem management decision-making (Montgomery, Grant and Sullivan, 1995; Buzzelli, 2020). Described here is the partitioning of the landscape into nontidal watersheds that deliver runoff through single channelized outlets and margin watershed areas that deliver runoff as overland flow.

2.2.2.1. Watershed Delineation

The D8 flow accumulation data for a domain allows the visualization of flow patterns and networks and the proper placement of "pour points" coincident with watershed outlets (O'Callaghan and Mark, 1984). A watershed algorithm uses a selected pour point in conjunction with a flow direction raster to identify all cells that contribute flow to that outlet, which can then be converted to a polygon feature representing the watershed boundary (Esri, 2016). Pour points can be placed anywhere in a DEM to delineate the contributing drainage area to a particular cell, but in practice they are most often placed at physically meaningful locations such as stream outlets. In this case, they are positioned at the confluence of nontidal channel networks and tidal water bodies.

The U.S. Geological Survey's National Hydrography Dataset (NHD) "Flowline" dataset was used as the definitive set of rivers and streams reaching the Maine coast for this research. While the NHD includes watershed boundaries at multiple scales, these were not delineated at sufficiently high resolution nor with sufficient accuracy to be usable for our analyses, leading to the need to delineate our own watersheds from the outlet points to ensure proper association of contributing watershed areas to the coast's many small inlets. During the analyses of Smith et al. (2016) the dataset contained 535 stream lines that intersected with the coastline boundary, leading to the delineation of 535 nontidal watersheds. Subsequent densification of the dataset by the USGS added a further 1,660 small coastal streams, for a total of 2,195 nontidal watershed outlets for the current analysis.

The term "pour point" denotes the outlet of a watershed, i.e. the point from which surface flow "pours" out of an area (Esri, 2016). When delineating a nontidal watershed (i.e., the landscape contributing flow to a stream above its tidal limit) from a DEM, the pour point is located where a stream meets the shoreline boundary layer. Using the updated NHD dataset as the authoritative set of coastal Maine streams, the theoretical implementation of this would be the intersection of the NHD flow lines with the shoreline of the Maine townships boundary layer. However, NHD lines do not align perfectly with the DEM-derived flow direction and accumulation rasters used for watershed delineation. Accordingly, it was necessary to manually place a pour point for each of the 2,195 NHD outlets at the intersection of the shoreline boundary and the nearest major flow accumulation line present in the D8 flow accumulation raster. Pour points were used in conjunction with the hydrologically enforced 2 m flow direction raster to update the boundaries of the 535 original nontidal watersheds and delineate 1,660 new small nontidal watersheds that fell within the original margin areas.

2.2.2.2. Margin Watershed Area Delineation

The updated Margin Watershed Area (MWA) comprises the coastal landscape falling between the shoreline and the boundaries of the 2,195 nontidal watersheds and was delineated as a polygon clipped to those two data layers. Unlike stream and river watersheds that function as ‘natural units’ (Dungan et al., 2002) with distinct individual boundaries and surface flow that exits through single outlets, surface water runoff flow paths from margin watershed areas do not converge into single channel outlets but rather enter estuarine waters as non-channelized overland flow or shallow concentrated flow, where the total margin area contributing runoff to a coastal embayment is dependent on the placement of the outlet line for that embayment. The MWA polygon was split into a regular grid of 30 x 30 m polygon “cells” to support parameterization of margin watershed areas for later aggregation and analysis using variable embayment outlet locations.

2.2.3. SDR Parameterization

An exhaustive accounting of all the bacteria pollution along the Gulf of Maine coast would require direct bacteria source tracking, comprehensive modeling of the overland and channelized flow that transports the bacteria, and detailed three-dimensional modeling of the complex, temporally variable flow paths through tidal embayments. Some of these types of information have been produced independently for limited subsets of the problem domain (Sims and Kaczor, 2017; Cronin, Smith and Fisher, 2022; Casella et al., 2023; Alahmed, Ross and Smith, 2022; Ross et al., 2021), but their acquisition for the entire coast is beyond the current capacity of the researchers working in this domain or their partner stakeholder agencies. In order to understand relations between landscape conditions and estuarine bacteria pollution levels to support scientific decision-making, both nontidal watershed units and margin watershed areas required parameterization with proxy metrics from spatial data corresponding to the bacterial SDR culprit categories.

Proxy metrics are used across the sciences as stand-ins for processes or parameters that cannot be measured directly or for which obtaining data would be prohibitively difficult (Sutton et al., 2009; Wei Luo et al., 2009; Chen and Nordhaus, 2011; Maina et al., 2012; Schilling et al., 2014; Demattê et al., 2020; Jennings, McCormack and Sheane, 2020; Royer-Gaspard, Andréassian and Thirel, 2021; Stirpe et al., 2021; Roydhouse et al., 2022). The choice of specific proxy metrics for analysis was guided by theory and knowledge of processes related to coastal bacteria SDR, prior research findings in the Gulf of Maine coastal domain, and expert stakeholder input through collaboration engagement with Maine DMR managers to identify processes relevant to decision science objectives. Metric selection was constrained by the public availability of relevant spatial data. General rationale for selecting source, delivery and residence time metrics are outlined in this section.

<u>Source Metric Categories:</u> Source proxy metrics are chosen to capture processes related to the generation of bacteria and pollutants. The pathogens under particular consideration for this research are sourced from the feces of warm-blooded animals, including humans, with sewage being a known problem. Source categories fall into two types, point sources and nonpoint (distributed) sources. One set of selected source proxy metrics encompasses known outfall point sources, which can deposit waste directly into estuarine waters. A second set relates population distribution in the landscape to the nonpoint source generation of waste from humans and their pets. A distinction here was made between year-round

residents captured by census data and the large seasonal (primarily summer, for the coastal zone) tourist population addressed as the distribution of addressable structures, including vacation homes that may stand empty during the off-season. A third set relates land use and cover to other nonpoint sources aggregated at the landscape scale. These sources include anthropogenic influences such as agriculture as well as source processes related to wild animal populations.

Delivery Metric Categories: Pollutant delivery metrics are proxies for the transport of bacteria and pollutants from the landscape into estuaries through surface water runoff. These processes can be divided generally into two primary categories related to the generation of runoff during precipitation events and the efficiency of movement of that runoff across the landscape. Runoff generation, which governs the capacity to wash pollutants from their sources in the landscape, is addressed using proxy metrics related to soil types that determine partitioning of precipitation between infiltration and runoff during storm events. (A land use metric considered in the Sources category, percent developed/urban, also serves as a proxy metric for total impervious surface in a landscape). Runoff efficiency is addressed with proxy metrics related to the density of channelized and engineered drainage networks, landscape slope and terrain that governs drivers of overland flow, and proximity of generated runoff to coastal waters.

Residence Time Metric Categories: Residence time metrics are proxies for processes that control the length of time polluted water remains in an estuary unit. The processes driving pollutant evacuation fall primarily into two broad categories, freshwater forcing from landscape sources and physical forcings from external sources such as tidal action, currents, and wind. For freshwater forcing, proxy metrics were selected to account for embayment size and the ratio of runoff volume to estuary volume during high-magnitude precipitation events. External forcings are more challenging to capture with proxies at a coastwide scale due to the complex, three-dimensional and temporally variable movement of water within embayment settings. Three plan-view morphology measures were selected as proxies for processes that affect external forcings and within-estuary mixing. Estuary openness, a measure of how enclosed an embayment is in relation to the waters beyond outlet, was chosen to address the ease of exchange of estuary water with the waters of the greater Gulf of Maine. Estuary circularity, a measure of the compactness vs branching of an embayment, was chosen as a proxy for the relative amount of mixing within an estuary. Outlet bearing, a measure of embayment orientation, was selected to account for the effect of prevailing wind and wave direction on estuary forcings.

2.2.4. Proxy Spatial Data Metrics

Watershed area metrics: Proxy spatial data metrics related to bacteria sources and delivery can be preemptively applied to nontidal watershed and margin cell polygons for later selection and aggregation, preventing the need to re-calculate metric values from the various source data layers every time an estuary unit is delineated. Most of these metrics are applied using simple overlays of watershed polygons against their respective data layers to calculate the fraction of overall polygon area within the relevant metric data class or total counts of points within the polygon; the latter is easily divided by polygon area to calculate density. Exceptions to the simple overlays are described below.

Geomorphically-derived (non-engineered) drainage density within watershed polygons was calculated using stream networks extracted from the 2 m flow accumulation raster due to underrepresentation of first-order and zero-order flow paths in available NHD data. Due to lack of data about channel head locations across the study domain, three sets of polylines representing stream networks based on source area to channel initiation of 0.30, 0.20, and 0.05 km^2 were delineated from the flow accumulation raster following the methodology of Van Dam (2022). Total stream network length within each polygon was calculated and divided by the corresponding watershed drainage area. Geomorphic + Engineered drainage density was similarly calculated by adding in the total length of the road network to the stream network.

Population data were derived from aggregated data at the census block level, the finest level of detail available. Population totals within census blocks were converted to population densities, which were then aggregated for watershed area polygons using area-weighted overlays of census block polygons. Performing these overlays required the assumption that population density within individual census blocks was uniform, which is not always true in a rural state with geographically large blocks.

<u>Estuary area metrics:</u> Residence time proxy metrics for estuary areas require the specific embayment polygon to be delineated before they can be calculated. While this precludes the possibility of pre-parameterizing the estuary space with proxy metrics, calculations can be described. Estuary mean depth and volume are calculated by overlaying the delineated embayment polygon against bathymetric raster data. Estuary openness is calculated as the ratio of the length of the outlet line (i.e., the width of the embayment mouth) to the perimeter of the embayment. Estuary circularity is calculated as the ratio of the embayment area to the area of a circle with the same perimeter using the Polsby-Popper score (Polsby and Popper, 1991; Cox, 1927). Outlet bearing is calculated as the azimuth of an outward-pointed line perpendicular to the estuary outlet line.

2.3. Results

2.3.1. Flow Pathways

Flow path delineation for this project resulted in the creation of several new data layers spanning the length of Maine's coastal landscape at two meter resolution. These runoff pathway layers provide the necessary data for nontidal watershed and margin watershed area delineation for this research, as well as for subsequent delineations of estuary units.

- *Culvert polylines* – for non-State roads without DOT culvert point data, this data layer represents the most complete set of georeferenced culvert locations for coastal Maine.

- *Hydrologically enforced bare earth DEM* – this raster, the product of burning culvert polylines into the raw 2 m DEM for coastal Maine, served as the base layer from which high resolution flow path data were derived.
- *D8 flow direction raster* – this layer, derived from the hydrologically enforced DEM, contains flow direction data necessary for delineating watershed areas at higher resolution and with more accuracy than existing NHD watersheds.
- *D8 flow accumulation raster* – this layer, derived from the D8 flow direction raster, contains flow pattern data (location and relative amount) necessary for the placement of "pour points" (outlets) for delineating watershed areas.

A total of 6,097 DOT culvert points fell within the footprint of the coastal Maine DEM. An additional 29,056 culvert lines were manually placed and burned into the DEM. Without this extensive hydro-enforcement, runoff from sections of the landscape would be erroneously routed into the wrong watersheds and thus wrong coastal estuaries (Figure 2.5). In the two most extreme examples, inland regions of 30.1 and 232.4 km^2 were incorrectly assigned due to single missing breach lines until proper flow paths were enforced.

2.3.2. Watershed Delineations

In the approximately five years between initial LPRU nontidal watershed delineation by Smith et al. (2016) and the re-delineation of nontidal watersheds for this research, the number of coastal stream and river outlets represented in the National Hydrography Dataset increased from 535 to 2,195. The updated nontidal watershed dataset accounts for a total of 16,625 km^2 of drainage area all falling within the borders of the State of Maine, with median watershed area 0.52 km^2, minimum 0.005 km^2, and maximum 1,660.5 km^2.

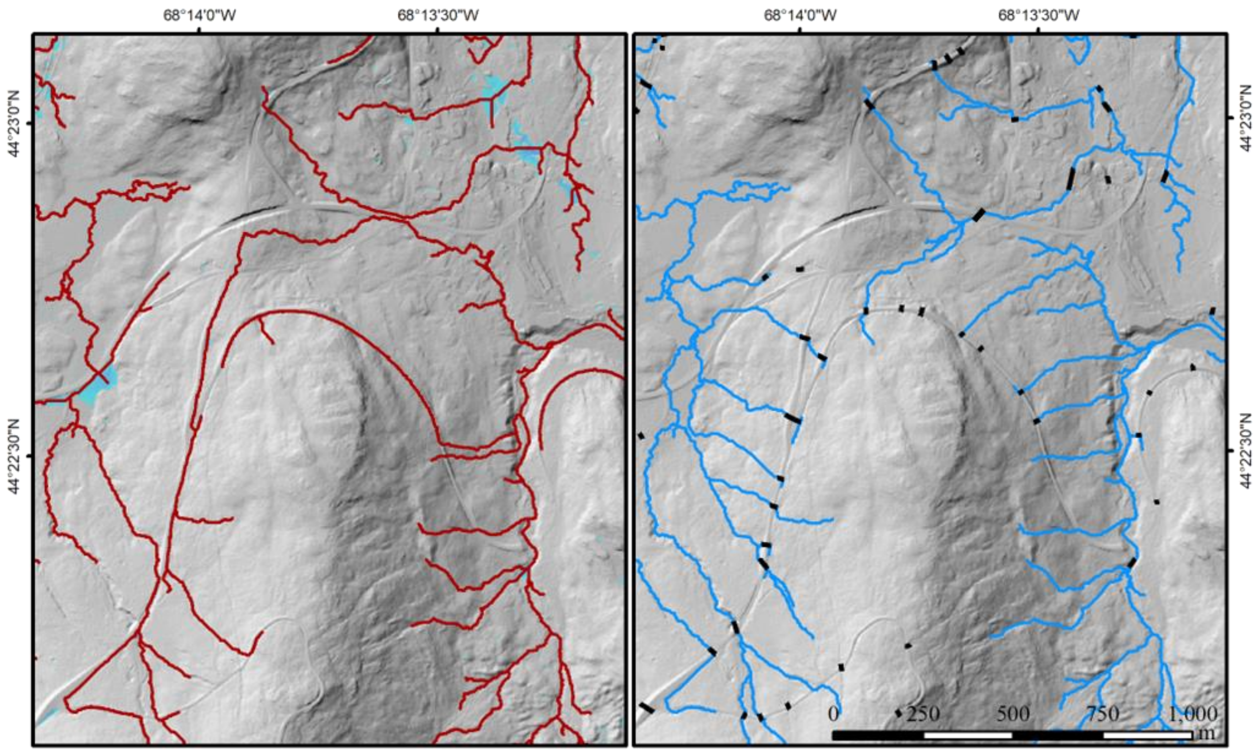

Figure 2.5. Flow lines at Cadillac Mountain on Mount Desert Island delineated from non-hydro-enforced (left) and hydro-enforced (right) versions of two meter resolution digital elevation model (DEM). Missing culverts in the non-hydro-enforced DEM result in artificial fills (teal regions) forming behind some road crossings and in runoff (red flow lines) from the northwest side of the mountain being incorrectly routed along the mountain access road to the Kebo Brook stream network to the east, ultimately entering coastal waters at Cromwell Cove southeast of Bar Harbor. Hydro-enforcement of the DEM by burning in road-crossing culvert lines (black) results in delineated flow pathways (blue lines) that more accurately correspond to the actual channel network on the ground, with runoff from the northwest side of the mountain correctly routed into the Duck Brook watershed that enters coastal waters northwest of Bar Harbor, more than 3 km from Cromwell Cove and separated by a headland.

Because the 2 m DEM used for delineations extended only 1.5 km inland from the original margin boundaries, any of the 535 previously nontidal watershed boundaries that extended farther inland remained unchanged. Of a total of 8,640 km of boundary length for those watersheds, 4,195 km across 417 nontidal watersheds were beyond the re-delineation area and were retained unaltered. Where discrepancies from original delineations appeared, they were most often caused by the addition of a road-crossing culvert that had been missed during the original delineation process, swapping generally tenths of a square kilometer from one nontidal watershed to another (Figure 2.6).

2.3.3. Margin Watershed Areas

The addition of 1,660 coastal nontidal watersheds associated with NHD densification resulted in a decrease in apparent margin watershed area in the coastal domain by 964 km^2, or almost 35%, as all new watersheds fell within what had previously been considered MWA (Figure 2.7). The total of 1,807 km^2 of MWA was split into ~2.1 million 30 m by 30 m polygon "cells" for parameterization with SD proxy metrics (Figure 2.8). Where the grid of cells intersects with nontidal watershed or shoreline boundaries, MWA cell drainage area is as low as 10.0 m^2. A small proportion of polygons (<2.67%) along nontidal watershed boundaries were merged to alleviate technical issues with selection and aggregation arising from their particular shapes and dimensions (Chapter 3), resulting in drainage areas greater than 900 m^2.

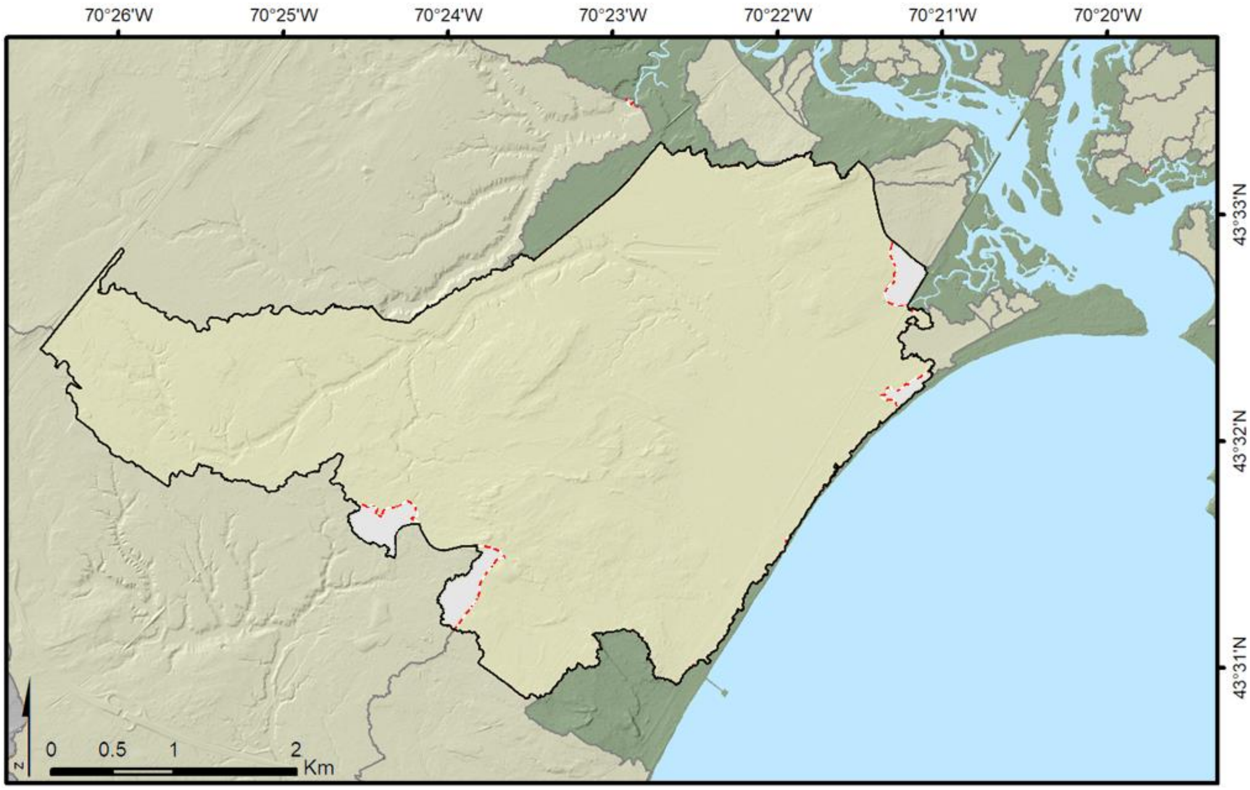

Figure 2.6. Typical minor changes between Smith et al. 2016 (dashed red) and re-delineated (solid black) boundary lines for the Jones Creek nontidal watershed, Scarborough River estuary. The Jones Creek watershed (yellow) absorbed a total of 0.42 km^2 (white) that had previously been delineated into adjacent nontidal watersheds (tan) or margin watershed areas (green).

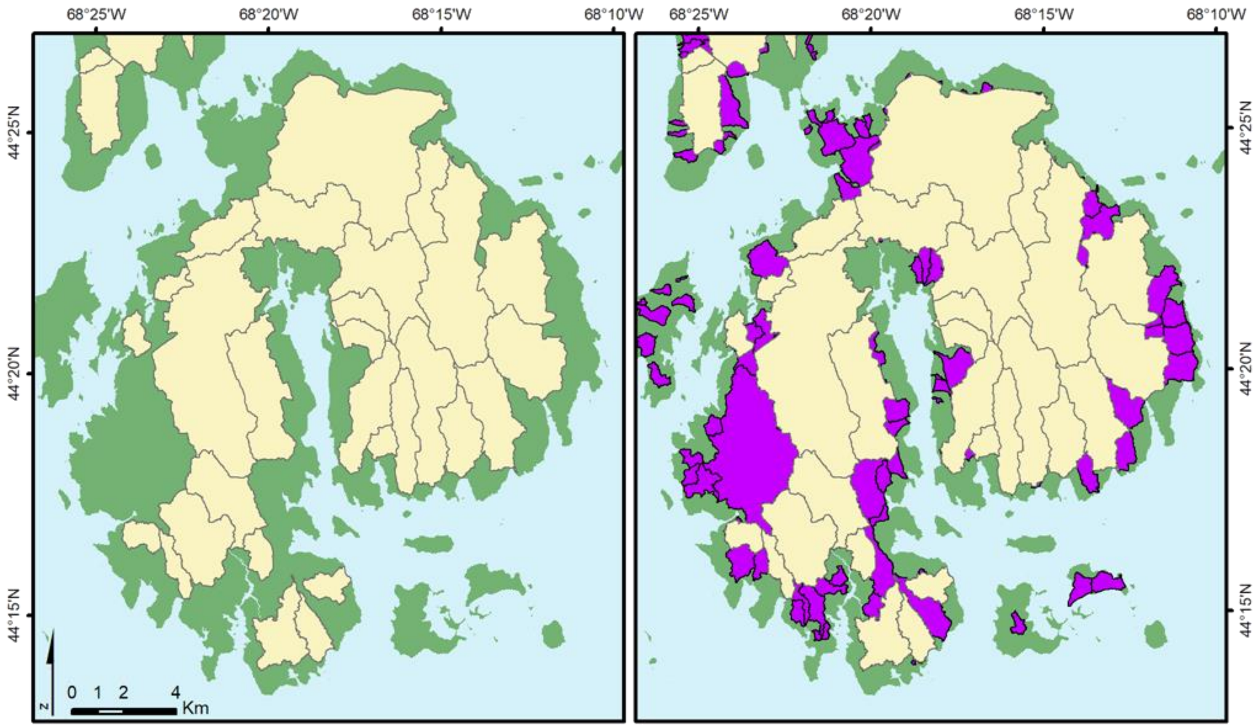

Figure 2.7. Maps showing nontidal and margin watershed areas on Mount Desert Island. These appear here as delineated for "landscape pollution response unit" analysis by Smith et al. (2016) (left) and after updated delineations including additional stream outlets (right). Coastal margin watershed area extents (green) are defined by the absence of previously delineated (yellow) and newly-delineated (purple) nontidal watersheds.

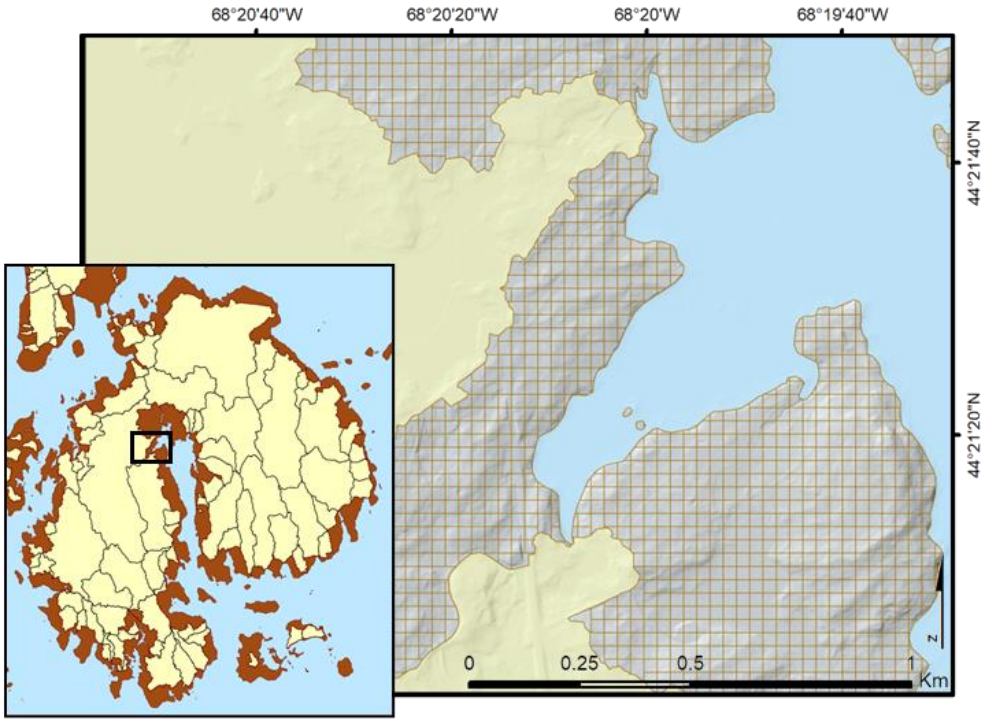

Figure 2.8. Partitioning of Margin Watershed Areas (MWAs). The MWA, shown in brown in the inset map of Mount Desert Island, was divided into a regular grid of 30 meter polygon cells to accommodate parameterization with proxy metrics. Cell size and alignment was chosen to correspond with the 30 m resolution National Land Cover Database raster, the coarsest proxy metric data raster used in the analysis.

2.3.4. SDR Parameterization and Proxy Metrics

Spatial data proxies were identified for seven general types of coastal processes within the three SDR pollution culprit categories: point sources (S), population-related nonpoint sources (S), land use-related nonpoint sources (S), surface water runoff generation (D), runoff delivery mechanisms (D), freshwater forcing in estuaries (R), and controls on estuary circulation patterns (R). A full list of processes and associated metrics selected for analysis with references (Table B.2) and publicly available sources for the data used in proxy metric calculation (Table B.3) can be found in APPENDIX B.

All nontidal watersheds and MWA cell polygons were parameterized with a total of 36 proxy metrics within source and delivery culprit categories. Size distributions of polygons have been presented in sections 2.3.2 and 2.3.3. Drainage density values including both terrain-based and engineered drainage structures range as high as 33.98 km^{-1} among nontidal watersheds (drainage area (DA) = 6.45 km^2) and 178.5 km^{-1} among full-size (DA = 900 m^2) MWA cells. Median values are 5.2 km^{-1} and 0 km^{-1} for nontidal watersheds and MWA cells, respectively. Among nontidal watersheds population density ranges as high as 2,280.54 persons/km^2 (DA = 0.20 km^2) with median value 25.85/km^2 and mean value 66.86/km^2, while among full-size MWA cells it reaches 96,989.5/km^2 with median value 20.93/km^2 and mean value 104.02/km^2. Highest values for both watershed area types occur on or adjacent to the densely populated Portland peninsula in southern Maine.

All but three proxy metrics related to proportions of land use (e.g., fraction rural) and soil drainage class (e.g., fraction poorly drained) are represented along the full range from 0% to 100% coverage among nontidal watersheds. The exceptions are fraction developed (maximum 97.2%), fraction farmed (86.2%), and fraction moderately well drained (96.5%). All land use and drainage class metrics were represented along their full range among MWA polygons.

2.4. Discussion

The accomplishment of the four functional objectives targeted by the research presented in this chapter establishes the spatial data necessary to facilitate setting identification and pollution vulnerability analyses in the coastal Maine problem domain. The outcomes provide a roadmap for researchers seeking to parameterize spatial domains related to runoff-borne constituents. Delineation of surface runoff flow paths in the problem domain resulted in the creation of highly detailed flow maps and watershed boundary delineations for partitioning runoff from land areas into tidal embayments that are components of estuary units. This process included the development of protocols for the identification of road-crossing culvert locations from digital elevation models in the absence of other available spatial data. The research outcomes also provide the first comprehensive treatment of coastal margin watershed areas with delineations based on physical runoff processes rather than previous approaches to the coastal zone that relied on jurisdictional boundaries.

Methodological choices in the design of this research were driven by deficiencies of the National Hydrography Dataset, the only existing hydrography data layers encompassing our entire study domain, for representation of features at the scale of small coastal watersheds. The highest resolution catchment data available for the state of Maine at the time hydrographic data preparation was completed for this study were HUC-12 hydrologic units (Seaber, Kapinos and Knapp, 1987) from the NHDPlusV2. HUC polygons are convenient units of analysis and have been used in Maine for setting identification from aggregated proxy metric data in upland settings (Gerard, 2018). However, the HUC-12 dataset is not high enough resolution to include a separate watershed polygon for each small stream meeting the coast, instead featuring individual polygons for larger rivers and coastal “margin” catchments that span multiple adjacent embayments and encompass several streams. This limitation led Smith et al. (2016) to delineate nontidal watersheds from LiDAR DEMs, a choice mirrored in this research.

The latest generation of the NHD, NHDPlus High Resolution (NHDPlus HR), does not have this limitation (Buto and Anderson, 2020). Each segment in the Flowlines stream network has an associated sub-catchment polygon. Sub-catchments along a stream network may be selected and aggregated to create a single nontidal watershed polygon for a stream meeting the coast. This generation of the NHD was not released for Maine until after delineations from LiDAR for this research had been completed but would have provided an attractive alternative to extensive DEM preparation and delineations for defining watershed boundaries. However, a separate remaining limitation of the NHDPlus HR is inconsistent stream line density for different areas of the state depending on data source resolution. The use of NHD stream networks to calculate drainage density values gives inconsistent results compared to DEM-derived networks as a result, supporting the use of the latter for this research. Additionally, the NHDPlus HR stream network still underrepresents the number of channelized drainage networks reaching the Maine coast, so the total margin area delineated from the current data continues to be an overrepresentation of its true extent.

Several lessons were learned during the course of coastal domain identification and parameterization. The first is a cautionary note about diminishing marginal returns and increased costs in the use of increasingly high resolution data sources (Zhang and Montgomery, 1994). The delineation of flow paths from 2 m LiDAR DEMs produced the highest resolution surface flow maps currently available for the coastal Maine landscape, compared to the 1:24,000 and 1:12,000 scale stream lines of the USGS NHDPlus dataset. Examinations of the role of DEM resolution in the accuracy of stream network and watershed delineations have revealed that high-resolution (1 m and 2 m) DEMs provide greater accuracy in hydrologic delineations than lower-resolution (10+ m) DEMs (Li and Wong, 2010; Lidberg et al., 2017). However, this came at the expense of considerable time and effort addressing the issue of hydro-enforcement. Separate delineation of flow for sections of the coastal domain using lower resolution 1/3 arc-second (~10 m) DEMs does not capture small-amplitude features at the same level of detail. However, it does produce substantially similar overall flow path patterns at the landscape scale with only a fraction of the hydro-enforcement being necessary. While the use of high-resolution LiDAR DEMs is appropriate for areas and analyses where high levels of detail are needed, a lower-resolution DEM may have sufficed for the specific objectives of this project.

The process of parameterizing the Gulf of Maine coastal domain also revealed data gaps related to flow path delineation, watershed outlet identification, and spatial data for proxy metrics. The accuracy of overland flow path delineation was potentially limited by two primary data gaps, both related to human engineering. The lack of location data for culverts on non-DOT (i.e., local) roads added considerable time to the length of the project and introduced potential sources of error (both of omission and commission), as potential road crossing locations had to be identified and delineated manually through DEM interpretation and, where available, photographic evidence. There is also a dearth of available data across the study domain for storm sewers, which remove precipitation runoff from the land surface and reroute it through subsurface engineered networks. In many cases, storm drain runoff will be discharged from an open pipe back into the nearby stream system or water body (or a constructed ditch or swale which connects with same) into which the runoff would have flowed pre-development, allowing it to complete largely the same journey to the tidal coastline and generally not affecting watershed boundaries. In more complex cases, storm sewer systems may be spatially extensive or boundary-crossing, requiring more work to identify network connectivity. Deficient representation of these networks results in a potentially distorted view of delineated flow paths as they relate to where concentrated flow reaches coastal waters.

Residence time-related processes were the most difficult of the pollution culprit categories to capsulize with proxy metrics. Metrics intended to address external forcing processes relied on plan-view morphometry, which do not fully capture the three-dimensional morphology of the estuarine water body. An additional factor not captured was the influence of outer islands or peninsulas beyond delineated embayment outlets that might moderate the effects of ocean swells and similar energetic forcings.

Other metrics were considered for this research but ultimately omitted due to lack of available data or concerns about data reliability. Septic systems, most commonly found in rural and semirural areas, are known to be potential contributors to water quality issues including fecal contamination in nearby water bodies when poorly sited or maintained (Reay, 2004; Sowah et al., 2014; Withers et al., 2014; Geary and Lucas, 2019). However, publicly available compiled data on septic system locations for the state of Maine was not available at the time of this study. Available biota-related data layers such as seabird nesting islands and shorebird areas were considered as source proxy metrics. Bird DNA markers were found in feces at 100% of sampling sites throughout the year during source tracking at Goosefare Brook in southern Maine (Sims and Kaczor, 2017). Gulls have also been implicated as a common fecal bacteria source in shellfish growing areas in New Zealand (Campos, Kelly and Banks, 2023) and at sites within the Great Lakes in the US and Canada (Staley and Edge, 2016; Brown et al., 2017), although at a lower level than other sources including human waste. Despite their prevalence in DNA tracking results, avian habitat was omitted from this study due to expert input from MEDMR shellfish program managers over concerns with reliability of the available spatial data for coastal Maine.

2.5. Conclusions

This chapter addressed the process of defining a land-sea domain relevant to bacteria pollution vulnerability in coastal estuaries. This process involved first the separation of the coastal landscape into categories of physical space based on differences in surface water runoff processes. A key scientific advancement of this portion of the research was the consideration of margin watershed areas immediately adjacent to the coast based on flow path delineation and runoff modality rather than jurisdictional boundaries, filling a research and management data gap in the Gulf of Maine. The flow path delineation process also highlighted the importance of proper flow routing for accurate connectivity of landscape areas to tidal embayments and resulted in the highest-resolution flow path data currently available for the coastal Maine domain. Selection of spatial data layers to serve as proxy metrics for processes related to bacteria pollution sources, delivery, and residence time revealed gaps among publicly available data layers and highlighted the importance of collaboration with expert stakeholders for identification and vetting of metric categories and data sources. The research produced a workflow for domain parameterization for identification of culprit processes and proxy metrics, coastal surface runoff pathways, watershed drainage areas, and proxy metrics representing attributes relevant to pollution problems. The outcomes provide a roadmap for researchers and managers to approach coastal management problems related to waterborne pollutants that can be applied to both coastal and inland landscapes.

CHAPTER 3

DIAGNOSTIC FRAMEWORK FOR COASTAL SETTINGS IDENTIFICATION

3.1. Introduction

Bacteria pollution problems in coastal estuarine systems are governed by physical processes related to pollution production in a landscape, transport of pollutants through coastal watersheds into tidal waters, and delays in the evacuation of pollution from estuarine waters through the combined influence of freshwater discharge and tidal exchange (Smith et al., 2022). Previous analyses using stream and river watersheds along the western Gulf of Maine indicate that proxy metrics derived from spatial data representing pollutant source, delivery, and residence time (SDR) processes can be used to identify and characterize coastal settings (Smith et al., 2016; Roy et al., 2018). The identification of varied conditions based on attributes related to pollution dynamics is important for coastal resource managers such as the Maine Department of Marine Resources (MEDMR) who support public health by ensuring that shellfish harvested from estuary mud flats are not contaminated by bacteria and are safe for consumers. This is accomplished primarily by temporarily closing harvest areas after high-magnitude precipitation events. Knowledge of the coastal settings associated with individual mud flat harvest areas provides a basis for the development of customized management strategies related to coastal planning, deployment of monitoring resources, and the design of setting-specific harvest closure policies.

Research designed to identify and distinguish coastal settings ideally clarifies the conditions of coastal landforms and attributes governing pollution problems in a manner useful to targeted management objectives such as shellfish sanitation and public health. Merriam-Webster (2023) defines a “setting” as the “time, place, or circumstances under which something occurs or develops,” which aligns with the goal of developing tools to support coastal management decisions in response to pollution problems affecting shellfisheries under varied conditions. The term has been widely used as a spatially framed description of physical conditions and in other cases to describe human activities, both of which are relevant to coastal pollution management and solutions (Lichter, Zviely and Klein, 2010; Van Den Berg, Jorgensen and Wilson, 2014; Lischeid et al., 2017).

There has been interest in dividing and classifying Maine's extensive coastline into different settings for at least 180 years. A four-compartment schema based on coastal morphology was first published by the state's inaugural State Geologist (Jackson, 1837). Kelley's (1987) comprehensive coastal census formalized this four-compartment schema based on principal component analysis of landscape physiography that identified marsh, ledge, and mudflat-dominated components. Fractal analysis of coastal shoreline plan-form complexity further supported the classifications by Kelley (Tanner, Perfect and Kelley, 2006). The broad distinctions in coastal settings included broad, arcuate beaches in the southwest, deeply indented coastline in the south central, island-bay complex in the north central, and rocky cliffs in the northeast. These coastal conditions are familiar to Maine residents and visitors, but it is important to note that they do not prescriptively describe conditions at individual sites within each compartment. Kelley (1987) cautions that "[a] ledge-dominated [site] from the northeast [compartment], for example, more closely resembles ledge environments in other compartments than marsh or flat-dominated [sites] from the northeast."

Pathogenic water pollution problems can occur in local areas. High bacteria counts or closures sometimes occur at monitoring sites immediately adjacent to sites with safe bacteria counts (Maine Department of Marine Resources, 2021). This observation indicates that attempts to define and classify coastal settings in relation to bacteria pollution problems must work at local rather than regional scales. It is necessary that the scales of consideration account for the physiographical and land use characteristics of individual estuaries or sub-estuaries and their contributing watersheds.

A range of approaches to classify estuarine setting types have been used around the world. The research efforts have ranged from primarily embayment-focused to holistically considerate of land-sea connections in coastal areas. Cluster analyses of moderate to large coastal estuaries in Tasmania and the United States have identified nine validated estuarine setting types based on overlapping sets of embayment specific geomorphic and hydrologic variables using unsupervised *k*-means clustering (Edgar et al., 2000; Engle et al., 2007). Multi-level hierarchical decision tree classifications for different estuary settings have been developed and applied in New Zealand (Hume et al., 2007) and South Africa (Van Niekerk et al., 2020). The classification approach broadly combined expert judgment rules for estuary morphologies and land cover characteristics to create ecosystem classification subtypes. Simenstad et al. (2011) incorporated ecoregion and ecosystem data, hydrogeomorphology, and land cover into a six-tier classification hierarchy within the Columbia River estuarine system in Washington.

When working in a coastal study domain with a tidal shoreline as complex as in Maine, a substantial ontological challenge that precedes setting identification is the establishment of a working definition for an embayment (Smith and Mark, 2003). Terrestrial watersheds have discrete boundaries and single outlet points through which gravity-driven surface water discharges to estuaries, making them ideal ‘natural units’ for spatial analyses (Dungan et al., 2002). In contrast, the outer boundary of tidal estuary embayments are not as clearly defined. The physical significance of an embayment outlet line used in the context of a delineated setting identification is to provide a partition between two bodies of water such that the embayment above the outlet and its contributing watershed can be considered as a unit. However, water flow within an estuary system is driven by multiple processes including freshwater forcing, tidal actions, and wind, leading to complex flow paths, including net movement of water inward from beyond the outlet in some locations within the study domain (Ross et al., 2021; Alahmed, Ross and Smith, 2022; Hillyer et al., 2022; Bailey et al., 2023).

Some published resources, such as the US Geological Survey's (USGS) National Hydrography Dataset (NHD), include embayment limits. These are primarily at the outlets of large embayments and where riverine estuaries meet a broadly sweeping coastline, without subdivisions useful for identifying settings to enable pollution prediction surrounding individual mud flats where shellfishing occurs. The USGS Geographic Names Information System (GNIS) contains points for almost 1,000 named "bay" type features along Maine's tidal coastline but poses the opposite problem, often defining very small embayments without consistency of naming convention. Consultation with coastal researchers familiar with Gulf of Maine morphometry suggests that there has been no definitive mapping of coastal Maine embayments, nor a robust definition or delineation rules for "embayment" in this ria coastal setting (J. Kelley, pers. comm.).

The project objective being pursued is the identification of coastal setting types using SDR proxy metrics for estuary units delineated from a static set of outlet lines as Smith et al. (2016) had done for the study domain with nontidal watersheds to identify Landscape Pollution Response Units (LPRUs). However, outcomes from the stakeholder engagement focused on estuary units indicate that a flexible system to identify coastal settings without *a priori* determination of what a relevant coastal unit looks like is a necessary and important research component. Two primary objectives thereby guided the process of setting identification for pollution management decision support tool development: 1) the development of an intellectual framework for delineating estuary units comprising nontidal watersheds, margin watershed areas, and embayment areas with associated SDR proxy metrics using any embayment outlet line drawn on a map, and 2) the establishment of rules for identifying a set of archetypical coastal setting types, or Coastal Pollution Response Units (CPRUs), for the Gulf of Maine study domain.

3.2. Methods

3.2.1. Stakeholder Collaboration

Stakeholder engagement activities were organized to establish criteria for embayment delineations on the Maine coast. A meeting to address the topic was convened with faculty and graduate students from the University of Maine and University of New Hampshire and shellfish program managers from the MEDMR, a key stakeholder group for this research with expertise and insight into the coastal sites under consideration. In a collaborative mapping exercise using a large print-out of the Maine coast, researchers and stakeholders collectively attempted to identify outlets to delineate embayment areas of particular importance to shellfishing management and ecological research (Figure 3.1).

There was a general consensus of what an embayment looks like relative to coastal geomorphic conditions and clear agreement on the outer limits of large embayment complexes and riverine estuary mouths delineated with a set of outlet lines similar to that of the NHD. However, there was surprisingly little agreement on outlet placement for sub-embayments within larger complexes and other edge cases, primarily due to uncertainty about surface water flow path connectivity at the local scale. Management-focused delineation suggestions were sometimes at odds with what coastline morphology would suggest due to focus on site-specific factors and field observations.

The exercise failed to establish comprehensive rules for delineating smaller subdivisions to identify settings at a more local level useful for management of individual mud flats. As a result, it was necessary to develop an alternate approach to the problem that did not rely on the predetermination of a static set of estuary outlets for setting delineation.

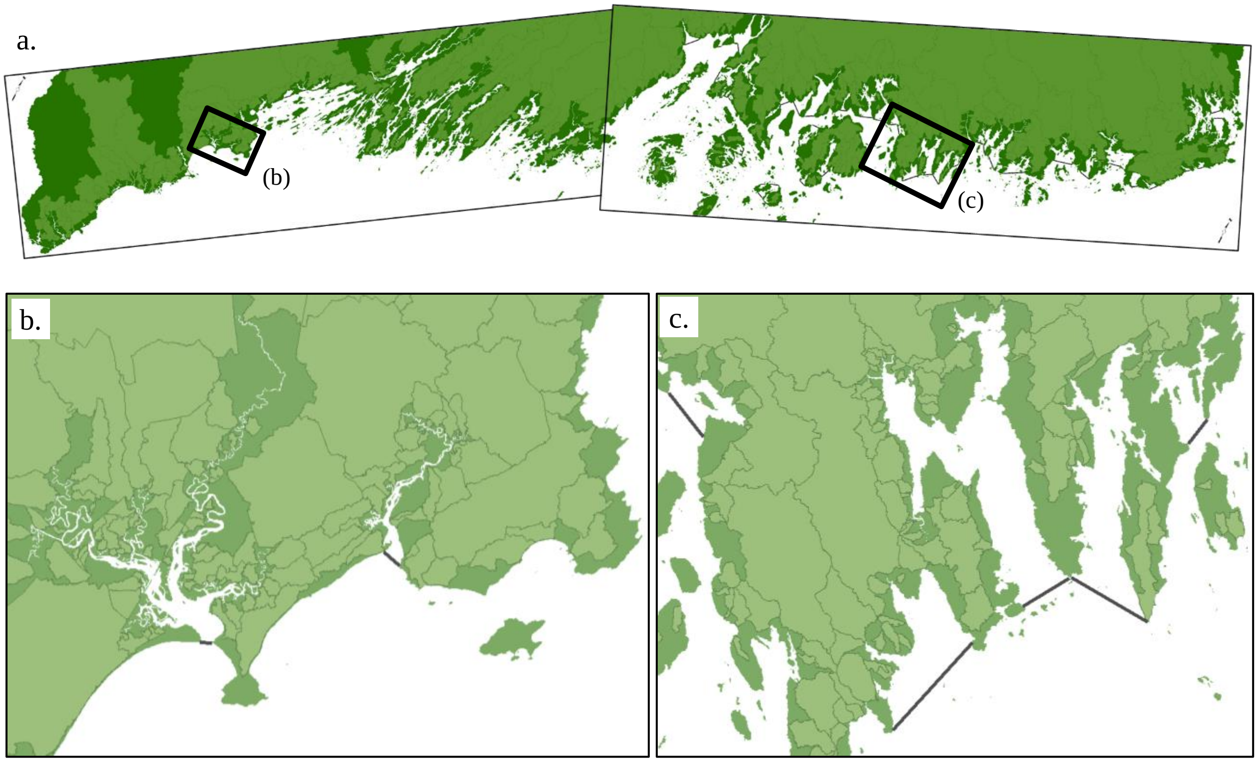

Figure 3.1. Results of embayment outlet mapping exercise for study domain. a) Coastal researchers, students, and stakeholders from the Maine Department of Marine Resources drawing on a 3.05 m (10’) map of the Maine coast collaboratively placed embayment outlets (black lines) where b) riverine estuaries met the sweep of the coastline and c) between the seaward ends of long peninsulas, but could not establish comprehensive scaling rules for the delineation of nested sub-embayments.

3.2.2. Classification System Conceptualization

In order to meet research needs, any setting identification system framework had to fulfil a set of primary design requirements (Simenstad et al., 2011). These requirements suggest the design of an expert system framework in which a knowledge base is combined with decision-making rules to convert relatively simple user input into expert knowledge about a setting (Jackson, 1986; Shu-Hsien Liao, 2005). The problem lends itself to a geographic information system (GIS) based approach, where a user-drawn outlet line on a map is used in conjunction with a spatial database to automatically delineate an embayment and contributing terrestrial watershed area (an "estuary unit"), aggregate relevant SDR proxy metric data for the delineated unit, and then apply a set of setting identification rules to sort the unit into one of several CPRUs (Figure 3.2). This is implemented in the form of the software tool developed for this research referred to as the "Estuary Builder" which runs as an add-on tool for the popular ArcGIS mapping program.

3.2.3. Study Domain and Estuary Units

The study domain for this research is the coastal landscape and waters of the state of Maine on the western Gulf of Maine. Maine's ~5,600 km of tidal coastline is drained by 2,195 coastal stream or river networks with outlets recognized in the USGS NHD "Flowline" dataset, as well as six large rivers (Saco, Androscoggin, Kennebec, Penobscot, Piscataqua, and St. Croix) that drain the interior of the state. The latter two rivers also mark the state's southwestern and northeastern coastal borders, respectively. These six rivers and their downstream areas are a data gap remaining from the LPRU analyses of Smith et al. (2016). They are omitted from this analysis due to their very large size in comparison to other watersheds under consideration. Between the boundaries of the 2,195 non-tidal watersheds and the Gulf of Maine shoreline is a total of 1,807 km^2 of "Margin Watershed Area" (MWA) that does not support the natural formation of stream channel networks such that all precipitation runoff occurs as overland flow.

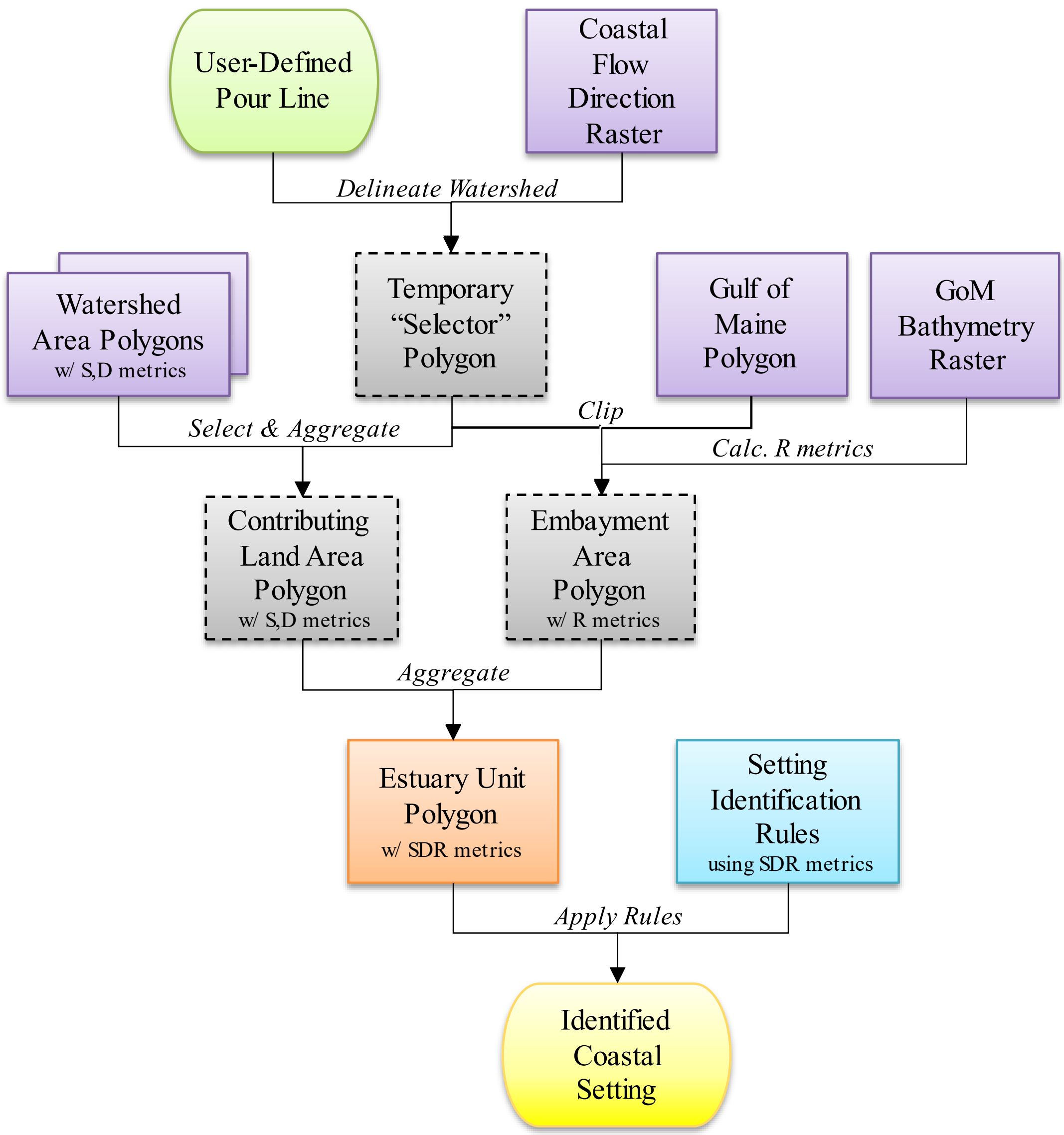

Figure 3.2. Simplified flowchart illustrating steps the expert system tool uses to identify coastal setting type for estuary unit delineated from user-defined outlet line. S, D, and R refer to bacteria source, delivery, and residence time, and GoM refers to the Gulf of Maine domain.

The seaward boundary of the study domain is a simplified shell encompassing the many islands off the mainland coast, with an outer extent at Isle Au Haut (Figure 3.3).

Polygons delineated from two meter resolution LiDAR digital elevation models (DEMs) by Van Dam (Chapter 2) identify 2,195 nontidal watersheds gridded 30 m cell polygons for MWAs. The coastal land areas within the nontidal watersheds and MWA sub-components of each “estuary unit” were parameterized with proxy spatial data metrics related to bacteria sources in the landscape and delivery efficiency processes, and additional proxy metrics relating embayment morphometry to flushing processes that affect estuarine residence time were identified.

Previous LPRU delineations used 2 m DEMs as the base layers for delineating watershed areas. However, their large file size led to slow processing times and strained available computer resources during geoprocessing operations. As a result, a lower-resolution one-third arcsecond (1/3”, ~10 m) DEM from NOAA incorporating both landscape and seafloor elevations throughout the study domain was selected to act as the base layer for embayment delineation and watershed area selection in the expert system tool. Due to lack of in-estuary flow pattern data for most of the study domain and limitations of static map-based geoprocessing tools for watershed delineation that require unidirectional, converging flow paths, the decision was made to route flow out of embayments based on bathymetry by treating the sea floor as a landscape surface. To ensure consistency with flow paths from 2 m data, raised ridge lines representing nontidal watershed boundaries and road-crossing culvert breach lines delineated during data preparation summarized in Chapter 2 were burned into the 1/3” DEM and D8 flow direction and accumulation rasters were prepared from the hydro-enforced DEM following methodology of Van Dam (2022).

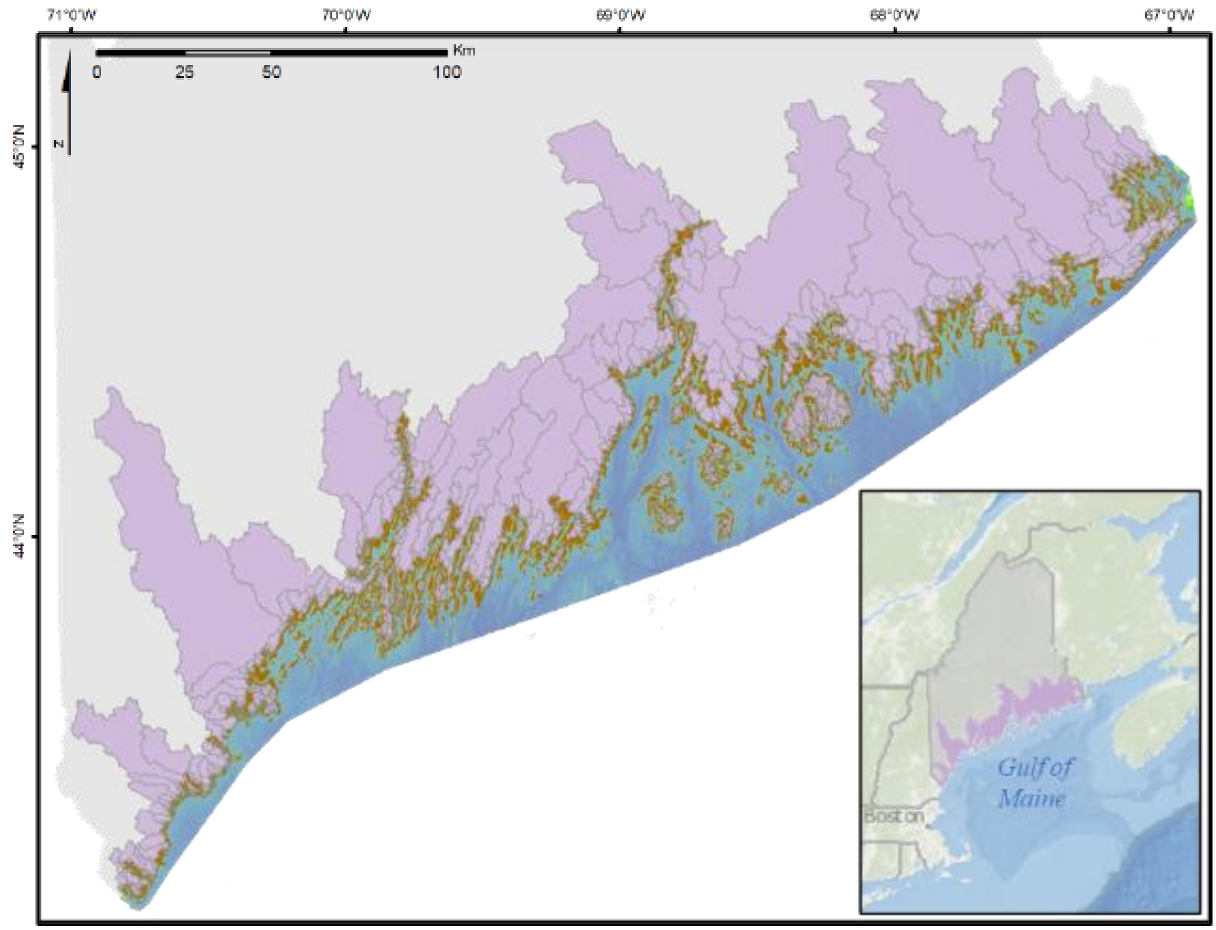

Figure 3.3. Coastal Maine study domain for coastal setting diagnostic analysis. The domain comprises coastal nontidal watersheds (purple), margin watershed areas (brown), and coastal waters (blue) along the western Gulf of Maine.

3.2.4. Estuary Units - Data Aggregation and Classification

A spatial data tool was developed for Esri's ArcMap (ArcGIS Desktop) to aggregate spatial data, delineate estuary units, and identify the estuary setting classification. Automated routines were written as a "script tool" in the Python (version 2.7) computer language. The tool incorporates functions from the ArcPy library that replicate individual geoprocessing functions found in ArcGIS software (Esri, 2022, 2021a). This allows a series of operations to be strung together to perform more complex tasks such as selecting and combining data layers intersecting a delineated watershed. Version 0.18.2 of the scikit-learn module containing machine learning and statistical tools was used for setting identification workflows (Pedregosa et al., 2011; Buitinck et al., 2013). During the course of the project it was announced by Esri that ArcMap would be discontinued and State of Maine agencies and the University of Maine System began to move to the replacement software, ArcGIS Pro. A parallel version of the code was adapted for the Python (version 3.8) computer language used by ArcGIS Pro as a result. Scikit-learn version 1.0.2 was used for this update.

Estuary units are delineated from user-defined outlet lines and a spatial database in a sequence of geoprocessing steps. Users may either load a data layer containing pre-drawn outlet lines or interactively draw one or more desired estuary outlets on a map. As an initial step, a new polygon type shapefile for output data is created from a template included in the tool geodatabase. This output layer is structured with an attribute table containing a column for each SDR proxy metric to be calculated and a row for each estuary unit outlet.

For each outlet line, a temporary "selector" watershed polygon is delineated from the D8 flow direction raster (Van Dam, 2022). This polygon comprises the embayment area above the outlet and all landscape area that contributes flow to the embayment, with an upper limit at the edge of the DEM 1.5 km inland of the MWAs. The selector polygon is used to select all nontidal watershed and MWA cell polygons it intersects within the delineated area. Because each of these watersheds is pre-parameterized with proxy metrics that can be accessed by the tool script, the tool can aggregate the attribute data of the selected polygons to calculate whole-watershed, margin-only, and highest high tide-only values for source and delivery proxy metric categories. Proxy metric values and a merged watershed polygons representing the entire terrestrial contributing area for the delineated estuary unit are then added to the output shapefile. The full set of SDR proxy metrics calculated by the spatial data tool and their aggregation methods can be found in Table C.1.

A small subset of residence time proxy metrics related to estuary morphometry cannot be aggregated from pre-parameterized data and must be calculated at the time of tool run. The selector polygon is clipped to include only the embayment area to calculate these metrics. Mean estuary depth is calculated by overlaying the resulting embayment polygon against bathymetric raster data (Esri, 2021b). Mean depth is multiplied by embayment polygon area to calculate total estuary volume. Estuary openness is calculated as the ratio of the length of the embayment outlet line to the length of tidal shoreline within the embayment. Estuary circularity, a measure of compactness, is calculated as the ratio of the embayment area to the area of a circle with the same perimeter (Polsby and Popper, 1991; Cox, 1927). Outlet bearing is calculated as the azimuth of an outward-pointed line perpendicular to the user-defined estuary outlet line.

Setting identification rules are based on unsupervised Gaussian mixture model (GMM) clustering of estuary units delineated with the initial, data-aggregation version of the spatial data tool. Model fitting was performed on a matrix of SDR proxy data for 500 delineated units using the scikit-learn Python library. Clustering rules were then integrated into the second iteration of the tool.

Unsupervised machine learning methods do not require labeled data or target variables and are used to find patterns and groupings in datasets. The unsupervised learning approach was deliberately chosen for the setting analysis over supervised approaches because MEDMR bacteria sampling techniques produce bias in the bacteria count record. This produces several outcomes that affect the analysis. First, areas with known long-term pollution problems are closed to shellfishing and never sampled. Second, sites where one test result triggers a temporary closure are retested until they are safe to reopen, expanding the dataset in certain locations. A third and potentially more important reason relates to equifinality (Beven and Freer, 2001). Two watershed units with very different SDR characteristics – for example, one with a source problem and one with a residence time problem – may coincidentally have similar bacteria counts as their SRD processes collectively create similar outcomes but for entirely different reasons. Accordingly, it is important that setting types reflect the particular combinations of SDR processes independent of bacteria counts at a particular site.

Past setting identification analyses using proxy metrics for Maine watersheds have used *k*-means unsupervised clustering (Smith et al., 2016; Gerard, 2018). The *k*-means clustering algorithm uses iterative expectation-maximization steps to find the centroids of *k* clusters and partitions points to the nearest centroids in *d*-dimensional parameter space. A limitation of this clustering method is that it tends to create circular clusters with the same radius regardless of underlying data structure. In contrast, the Gaussian mixture model (GMM) unsupervised clustering algorithm with similar calculation steps relaxes some of the limiting assumptions of the *k*-means algorithm such as allowing full covariance to identify clusters with different shapes or densities. This provides a modest improvement for many real-world data scenarios (Vanderplas, 2016). The unsupervised approach was thereby pursued using GMM to better account for outliers and a larger range of cluster constellations.

A principal component analysis (PCA) was first performed on standardized (centered by subtracting the mean and scaled to unit variance) metric values to identify linear combinations of metrics that describe the greatest variation in the data and reduce dimensionality (Ding and He, 2004; Vanderplas, 2016; scikit-learn developers, 2023d). This approach responded to the high degrees of collinearity among some sets of proxy metric columns. Selection of the final number of principal components was based on "marginal explained variance" (MEV) for each additional PC using the Rule N-criterion (Lipscomb, 1998; Preisendorfer, Zwiers and Barnett, 1981; Gerard, 2018). MEV for each PC in the SDR metrics dataset was compared to mean MEV values from 10,000 PCAs of random data matrices with the same dimensions (e.g., 500 rows and 107 columns). PCs were retained for analysis as long as their MEV outperformed the random dataset (Figure 3.4).

The GMM requires that the user specify the number (*k*) of clusters for the algorithm to identify. GMM clustering runs on the PC dataset were performed for 3-10 clusters using 100,000 initial random centroid initiations for each *k* (scikit-learn developers, 2023e). Optimum number of clusters was determined using the Bayesian Information Criterion (BIC), a selection criterion that penalizes model overfitting (Schwarz, 1978; Vanderplas, 2016). The fitted GMM model with the lowest BIC was selected as the final setting identification model exported along with the PCA rules into a single Python "pickle" file (Python Software Foundation, 2023; scikit-learn developers, 2023c). Pickling is used to package Python objects including trained models. This allows the model and its underlying setting identification rules to be read into other code such as that of the spatial data tool used for data aggregation and estuary unit classification.

The outcome from the spatial data analysis combines aggregated SDR proxy metrics and the trained GMM model to identify coastal setting classification for an estuary unit delineated from a user-defined outlet line. The aggregated SDR proxy metric data for a delineated estuary unit is first run through the imported PCA transformation. The PCs are then input into the fitted GMM model that uses fitting rules to assess which of the *k* defined clusters (setting classifications) the estuary unit belongs. The tool is designed to output the cluster identity in the aggregated proxy metric attribute table for each estuary unit under consideration.

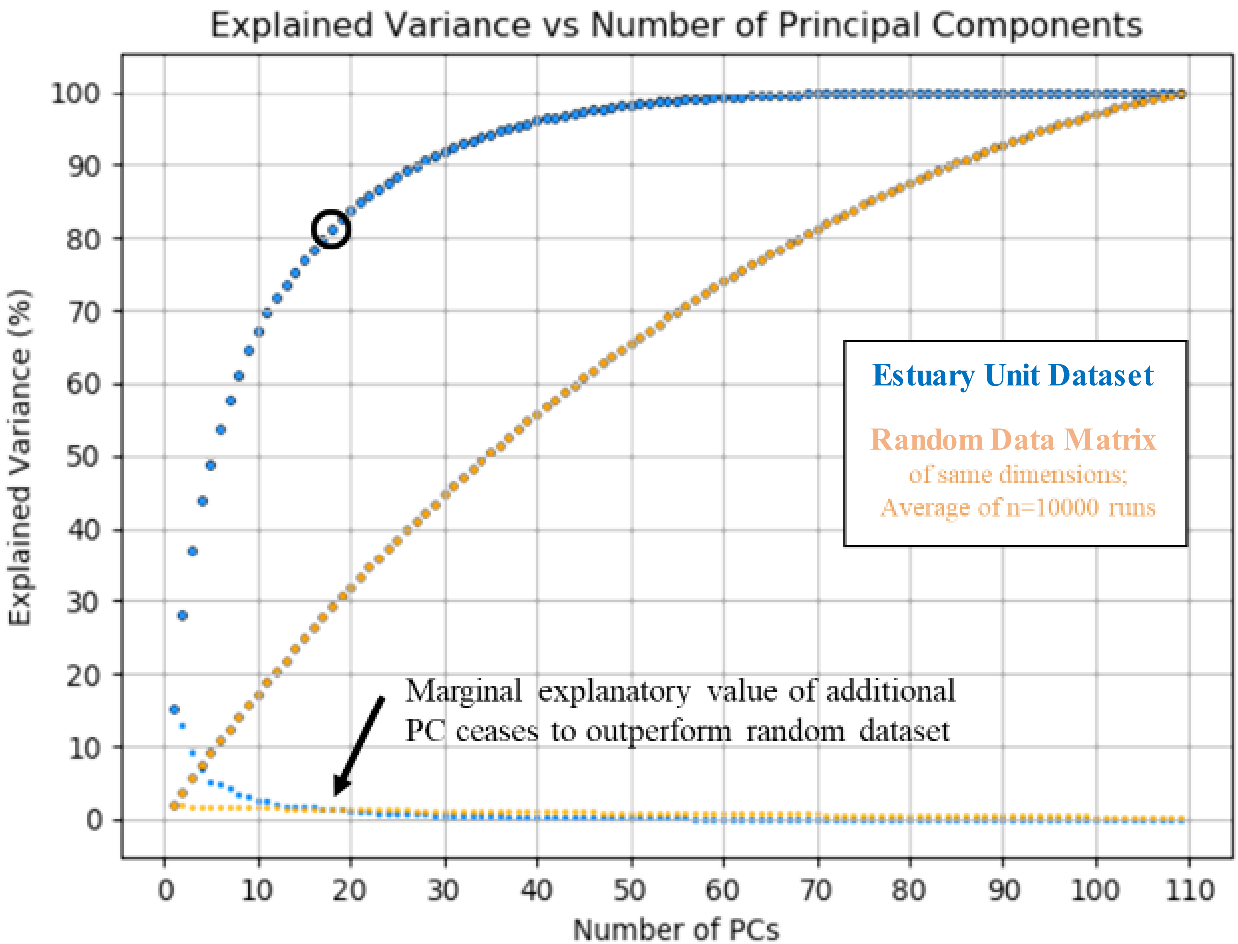

Figure 3.4. Plot describing selection of principal components (PCs) for analysis. An appropriate number of PCs for a dataset can be determined by assessing the marginal explanatory value of additional PCs for the dataset against a random data matrix with the same dimensions.

3.3. Results

Research results start from decisions with stakeholders to develop an expert system focused on proxy spatial data uses and estuary unit delineations to identify coastal settings related to bacteria pollution problems. They include the expert system creation and implementation as a interactive GIS tool and methodology to identify coastal settings. The final result is a thematic map of coastal Maine identifying CPRUs assumed to be relevant to pollution vulnerability.

3.3.1. Expert System Framework

The four primary requirements identified for the coastal Maine setting diagnostic system are listed below. These constitute research decisions that developed from consideration of the stakeholder interests and research objectives.

1. *Flexibility* – the system must be able to identify coupled land-sea settings based on any desired embayment outlet line.
2. *Comprehensiveness* – in recognition of the interconnected role of SDR processes in coastal pollution, the system must incorporate proxy metrics for the entire “estuary unit” consisting of contributing landscape areas, including margin watershed areas, as well as the embayment itself.
3. *Adaptability* – the system design must be able to incorporate new or updated data layers or clustering rules without modifying the core software framework.
4. *Accessibility* – the implementation of the expert system as a decision support tool should not require special technical knowledge from users to operate or to interpret output, and the tool should be well-documented with a user guide. This final requirement was not strictly necessary for the research itself but ensures that the tool remains useful beyond the initial life of the project by allowing other coastal researchers and resource managers to use it without needing to become experts in its use.

The tool was developed in two stages corresponding with its two primary functions, spatial data aggregation and setting identification. The initial spatial data aggregator tool was used to delineate a set of estuary units using outlet lines representing embayments and sub-embayments at multiple scales. Each estuary unit is then parameterized with the SDR proxy metrics determined from the research summarized by Van Dam (Chapter 2). Parameterized estuaries were used to develop setting identification rules using unsupervised machine learning methods (section 3.3.2). Setting identification rules were then integrated into the expert system that aggregates SDR spatial data and uses those data to identify the coastal setting type for an estuary unit above a user-defined outlet. A user guidance manual written for the expert system tool covering technical details is provided in APPENDIX D.

3.3.2. Setting Identification

Unsupervised Gaussian mixture model clustering of 500 delineated estuary units based on all columns of SDR proxy metrics (18 PCs explaining 79.6% of variance) resulted in the identification of seven coastal setting types for the Maine coastal domain, labeled A-G in order of decreasing membership count (Table 3.1). There is considerable variation in cluster size, with 87% of estuary units assigned to the largest three clusters and more than 50% to the single largest cluster, while the smallest three clusters each contained 2% or fewer of the delineated units.

Clustering was performed using principal components of SDR proxies rather than raw SDR data. However, histograms of original proxy metric values within each cluster can be used to identify characteristic attributes (

Figure 3.5) and describe the coastal settings. Clusters E-G had a membership of ten or fewer sites and tended to show the clearest distinctions from other sites. Almost all delineated units containing point source outfalls were assigned to one of those three clusters. Cluster E is also notable for containing sites with very large terrestrial drainage areas with a 25^{th} percentile drainage area value of over 125 km^2, well above the maximum non-outlier value of all other clusters but C. The relatively large sizes

Table 3.1. Attributes of seven coastal setting types identified using a GMM clustering approach.

Cluster Designation	**Site Count**	**Percent Total**	**Notable Attributes**
A	251	50.2%	Wide distributions of values for many proxy metrics relative to other clusters; "median conditions"
B	106	21.2%	Wide distributions of values for many proxy metrics relative to other clusters; high fraction in margins, rural; high fraction well-drained soils
C	78	15.6%	Large drainage area to estuary area ratios; low fractions margins
D	37	7.4%	Large drainage area inundated by highest high tide; high natural and engineered drainage densities in margins; average estuary circularity <10% with very low estuary openness; highest percentage conserved lands and storage in margins and high high tide areas; very poorly drained soils
E	10	2.0%	Very large landscape drainage areas, estuary areas and volumes; presence of point source outfalls (CSOs, PDES-Os); average estuary circularity <10% with very low estuary openness; large range of mean embayment depth (1 - 16 m, avg. ~7 m); large maximum elevation and slope
F	9	1.8%	100% Margins; high engineered drainage densities; highest average estuary circularity; very high fraction developed (mean >60%)
G	9	1.8%	Large range of mean embayment depth (2 - 16 m, avg. ~7 m); presence of point source outfalls (OBDs) and PDES facilities

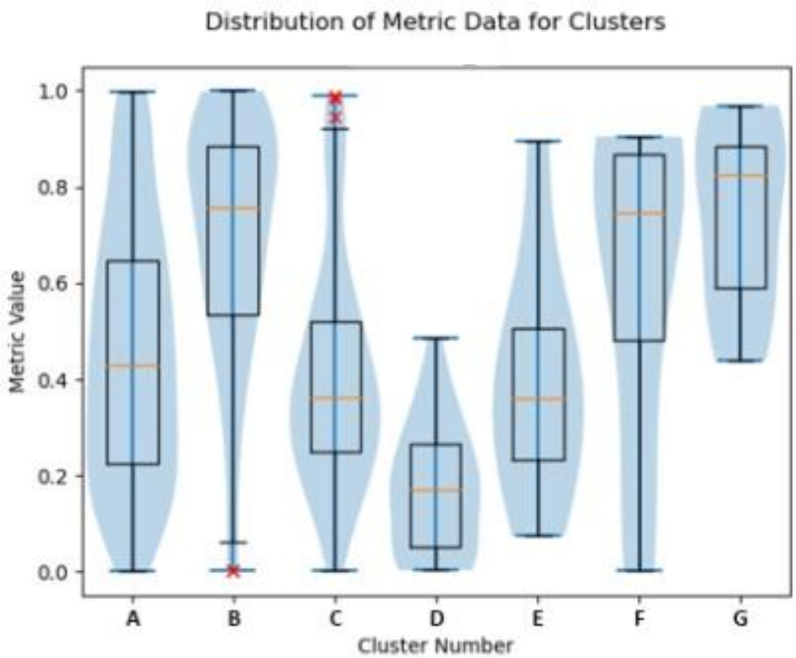

Figure 3.5. Example violin plot showing distribution of proxy metric values (here, for the fraction of margin watershed area with well-drained soils) for delineated estuary units within each of seven setting types identified by GMM clustering using source, delivery, and residence time proxy metric data.

present a confounding variable when looking at count-based metrics such as population or structure count. Cluster E is a clear standout in this regard despite having population and structure densities in line with other clusters. Cluster F, which does feature markedly higher population and structure densities, contains small urban estuary units consisting only of MWAs, most notably Back Cove in Portland.

The larger clusters predictably tend to show wider ranges of values for most proxy metrics. Nevertheless, distinctions among even the largest two clearly emerge from the data. No clearly urban cluster defined by population density, structure density, and fraction developed areas emerges among the four largest clusters. The smallest of these clusters, D, does have slightly higher values for these urban metrics than the other three. Cluster D also clearly contains the highest median fraction of conserved lands overall and within margin watershed areas of all seven clusters. The outcome suggests the presence of suburban parks and similar lands that would not be present in more rural areas.

Of the four large clusters, the highest fraction of MWAs are associated with Cluster B with a median value of ~0.7. Clusters C and D both contain comparatively small fractions of land as MWAs. Cluster A, the largest cluster, does not clearly stand out in any category but along with Cluster B features high median percentage rural lands and low population density. Clusters A and B are separated mostly clearly by differences in soil drainage, with Cluster B featuring a higher percentage of well drained soils, whereas Cluster A contains on average more soil storage depth.

GMM clustering was also carried out with subsets of the full suite of proxy metrics to test the effect of proxy metric inclusion in the cluster analysis on cluster membership assignment. Comparison with clusters derived by dropping point source columns (CSOs, OBDs, PDES-Fs, PDES-Os) is shown in Figure 3.6. The general trend of seven identified setting types persists with the truncated list of attributes related to water treatment infrastructure. GMM cluster analysis using five different proxy metric subsets also yielded seven settings. These included analyses of estuary unit sub-components, MWAs and embayment morphometry. The unequal cluster sizes also persisted but analysis of membership pairings between clusters from the two runs indicates that while some cluster assignments remain stable, others experience considerable rearrangement.

Hierarchical agglomerative clustering (HAC) of the 500 estuary units was also performed using the full set of SDR metrics to investigate the effect of clustering algorithm choice on setting identification. HAC is an unsupervised machine learning algorithm that iteratively finds pairs of points with the smallest Euclidian distance between them and merges them into increasingly large clusters, creating a hierarchical tree of cluster memberships (Nielsen, 2016). Examination of the resulting dendrogram indicated a strong model preference for two main clusters of unequal size plus one small cluster of just two estuary units, effectively suggesting that coastal Maine can be divided into just two setting “types.”

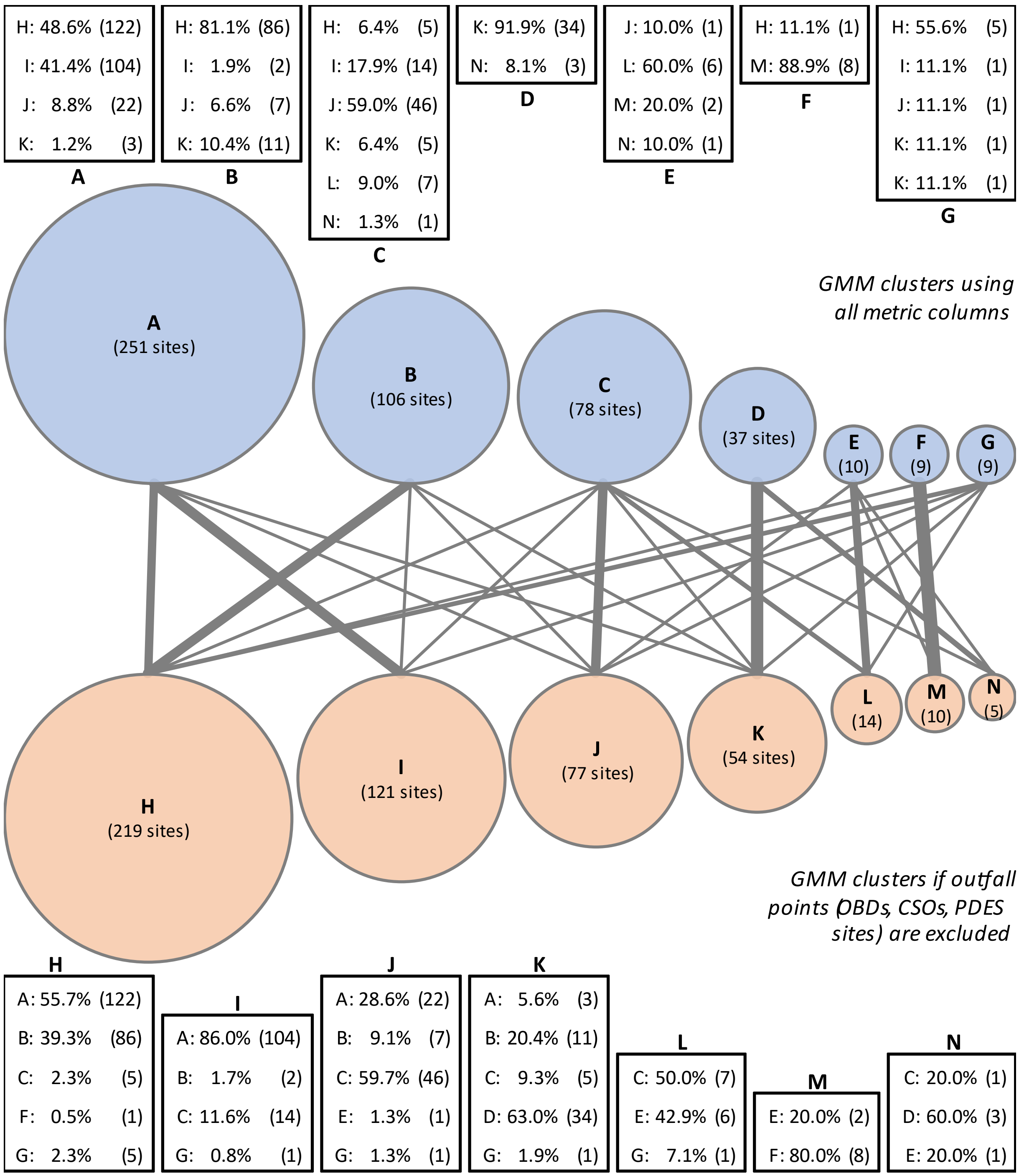

Figure 3.6. Illustration of shared membership among GMM clusters obtained using all source, delivery, and residence time metrics for 500 delineated estuary units (blue, top) and those obtained by dropping pollution elimination outfall points from analysis (orange, bottom). Strength of shared membership (as percentage of total sites in clusters) is indicated by thickness of connecting line between clusters.

3.4. Discussion

3.4.1. Assessment of Expert System Tool

The expert system approach developed for this research provides a novel and flexible diagnostic framework for site-specific coastal settings identification based on proxies for bacteria SDR processes in the Gulf of Maine study domain. The framework largely met the four primary design requirements set out for its design.

The primary requirement for flexibility in the line placement that drove the design of the expert system was met. The GIS tool is able to accept one or more lines for any coastal plan-form concavity as an input file or drawn interactively by the user at the time of tool run, delineate estuary unit(s), and return a polygon layer with aggregated SDR proxy metric data describing the space as well as calculated CPRU setting type designation. This flexibility allows the setting identification framework to be applied to sites without pre-defining embayment ontologies, although it also introduces potential for misleading results if line placement conflicts with actual estuary circulation conditions.

The requirement for comprehensiveness was also largely met. The analytical framework incorporates data for the entire contributing area comprising nontidal watersheds, margin watershed areas, and embayment areas above outlets in its identification of settings, providing a fuller understanding of sites within the coastal study domain. Some geographic areas of the coast could not be accurately represented within the tool due to data issues (Figure C.1). These were primarily areas downstream of the six large interior watersheds omitted from the analysis. An interconnected system of bays at the intersection of multiple tidal rivers results in unclear or temporally variable flow patterns incompatible with the unidirectional down-bathymetry flow routing employed by the tool in a portion of the study domain.

Minor tradeoffs arose in the requirement for adaptability, primarily in deference to processing time. For full modularity, any source data layer subject to change over time (for example, decennial census data or NLCD land cover, which is updated every five years) would be able to be swapped directly from the tool's geodatabase. However, repeated overlay operations with multiple data layers is computationally costly at the time of tool operation compared to accessing pre-parameterized watershed polygons and performing simple aggregation calculations with those data, supporting the choice to use the latter design. The inconvenience of occasionally re-parameterizing watershed polygons is particularly minor in light of the fact that any changes to input data already inescapably requires a full re-aggregation and re-clustering of the 500 embayments used to define the setting identification rules with the updated data. The setting identification rules themselves are among the most modular components of the tool since they are read from a separate file rather than being integrated directly into tool code. Direct incorporation of residence time rasters using modeled estuary circulation data has been completed in the Frenchman Bay portion of the study domain (Alahmed, Ross and Smith, 2022). Observations from this indicate that other data linking flow dynamics to coastal landforms represented in spatial datasets can be incorporated into the tool and analysis if it becomes available for the whole coast (Figure 3.7).

The final requirement for user accessibility is also judged to have been met. From a user standpoint, the only information necessary to retrieve spatial data information describing a setting is a drawn estuary outlet pour line. Beta versions of the Estuary Builder tool for both ArcMap 10 and ArcGIS Pro 3 were released at a workshop in spring 2023 to positive feedback from researchers and from shellfish managers, who indicate intention to use the expert system as a decision support tool for better understanding site conditions in the absence of bacteria monitoring data.

As a coastal spatial data aggregator and setting identification tool, the expert system represented by the Estuary Builder tool is not limited to bacteria pollution-specific research. Since its completion, the tool has been used for other coastal research in the Gulf of Maine study domain, including to identify and

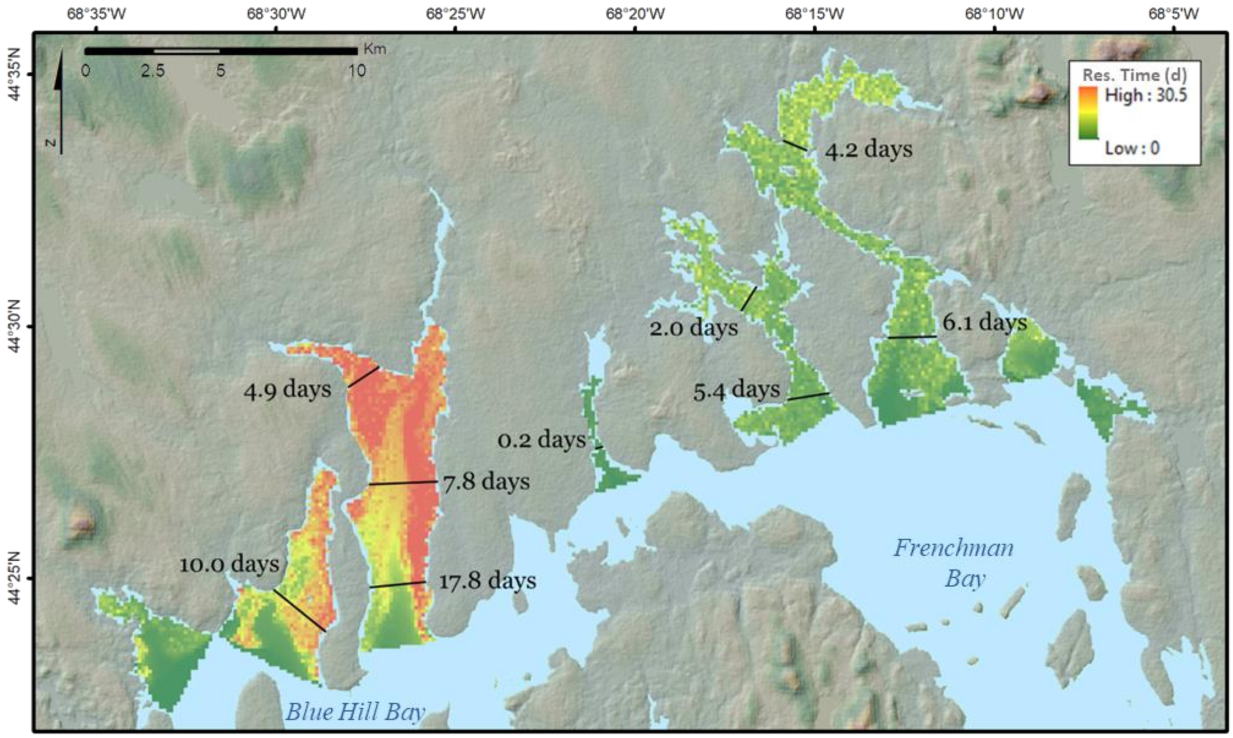

Figure 3.7. Example of incorporation of residence time rasters into the expert system tool using residence time maps for eight tidal estuaries surrounding Blue Hill and Frenchman Bays from Alahmed, Ross, and Smith (2022). For each embayment, pixel colors represent the evacuation time before water in that location reaches the lower embayment limits at Blue Hill or Frenchman Bay. Residence time for several sub-embayments defined by example outlet lines (black) is calculated here by comparing mean pixel values inboard of the outlet with the minimum value along the line.

delineate stream and river watersheds draining to large embayments for calibration of water models (Braun et al., 2021), estimation of other pollutant loads (Goodwin et al., 2021; Casella et al., 2023), and as freshwater inputs for large embayment hydrodynamic models (Alahmed, Ross and Smith, 2022; Bailey et al., 2023). Due to the modularity of the setting identification rules within the expert system framework, new rules featuring different unsupervised clustering methods or using supervised machine learning algorithms with specific data targets can also be incorporated into the tool, allowing researchers to adapt the framework to additional purposes.

3.4.2. Coastal Setting Identification

Unsupervised GMM and hierarchical agglomerative clustering suggest that the coastal Maine study domain can be divided into a small set of core setting types based on SDR proxy metrics. The assignment of a majority of sites to either three (GMM) or two (HAC) clusters seems surprising, but can be explained through a deeper examination of the approach driving the outcome.

Figure 3.8 shows a three-dimensional scatter plot with the first three principal components calculated from the full set of SDR proxy metrics for the 500 delineated estuary units. GMM cluster assignments are indicated by point colors. This plot is not a full visualization of the multi-dimensional structure of the data since it depicts only three of 18 dimensions considered in the principal component space. However, it shows that a clear majority of sites fall into one large grouping rather than occupying separate clusters across the first three axes of principal components selected to represent the greatest distinctions among the data.

This lack of clear separation among clusters can also explain the sensitivity of membership groupings to inclusion or omission of different metrics in the analysis. Because there is so much inter-group proximity even with non-collinear axes, a change in the included metrics that shifts points only slightly in Euclidean space may cause major realignment of cluster memberships. This can be observed in the plot space (Figure 3.6). This has implications to the selection of proxy metrics when building a setting identification framework, as inclusion of metrics unrelated to the targeted process or outcome may result in the identification of settings with less reliable relevance to the target outcome. The relevance of SDR proxy metrics used here to identify settings relevant to bacteria pollution is explored in Chapter 4.

The results of the unsupervised cluster analyses illustrate that setting identification can be accomplished using the expert system framework. They also suggest that most sites are more alike than different across the Gulf of Maine study domain. While this may be a challenging conclusion from a

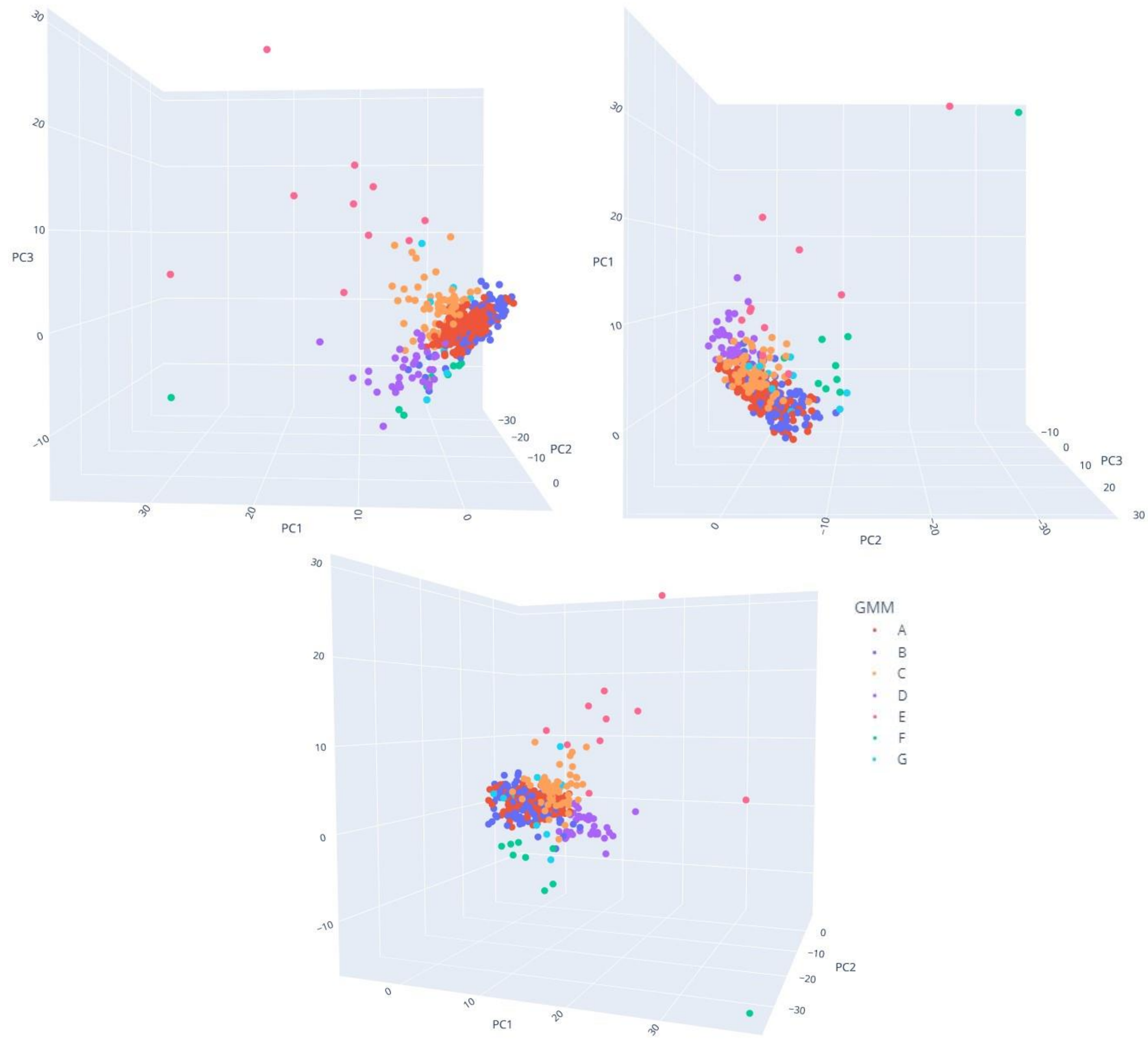

Figure 3.8. Multiple perspectives of a scatter plot of the first three principal components describing source, delivery, and residence time proxy metrics for the 500 delineated estuary units. Colors correspond to assigned Gaussian mixture model (GMM) cluster.

management perspective, it provides a useful set of observations about how SDR attributes collectively manifest along the coast to define settings. The increase identified settings from five LPRUs (Smith et al., 2016) to seven CPRUs does provide coastal managers with expanded opportunities to attempt to tailor strategies in accordance with certain emergent site attributes, particularly for the smallest clusters with stronger characteristic SDR proxy metric signals.

Comparisons can be made between CPRU results indicating coastal settings based on SDR attributes and the four coastal Maine regional compartments defined by Kelley (1987) based on coastal morphometry. A map of CPRU distribution along the coast (Figure 3.9) illustrates a distinct difference in the Southwest (SW) compartment compared to the South Central (SC), North Central (NC), and Northeast (NE) compartments. Stacked bar charts showing relations between CPRU cluster membership and compartment position (Figure 3.10) confirm a clear affinity between CPRU cluster D and the SW compartment. Cluster D contains estuary units with the highest proportions of tidal wetlands among the seven CPRUs, and the SW compartment was defined in part by a high proportion of marsh along its coastline. Closer investigation of the individual Cluster D sites within the compartment confirms that these estuary units correspond to the sites of the extensive salt marshes of the southern Maine coast, including the state's largest salt marsh in Scarborough adjacent to the border between the SW and SC compartments. The SW notably also contains only one estuary unit from Cluster A, fewer than any CPRU other than the ten-member Cluster E.

No similar affinities occur between other CPRUs and compartments. Within the other three compartments, CPRU counts conform to the overall distribution of estuary unit memberships within CPRUs, with most estuary units in Cluster A, a lesser number in Cluster B, and so forth. This outcome for most of the state's coastline corroborates Kelley's (1987) observation that regional compartment setting does not dictate conditions at individual sites and supports the classification and management of estuary

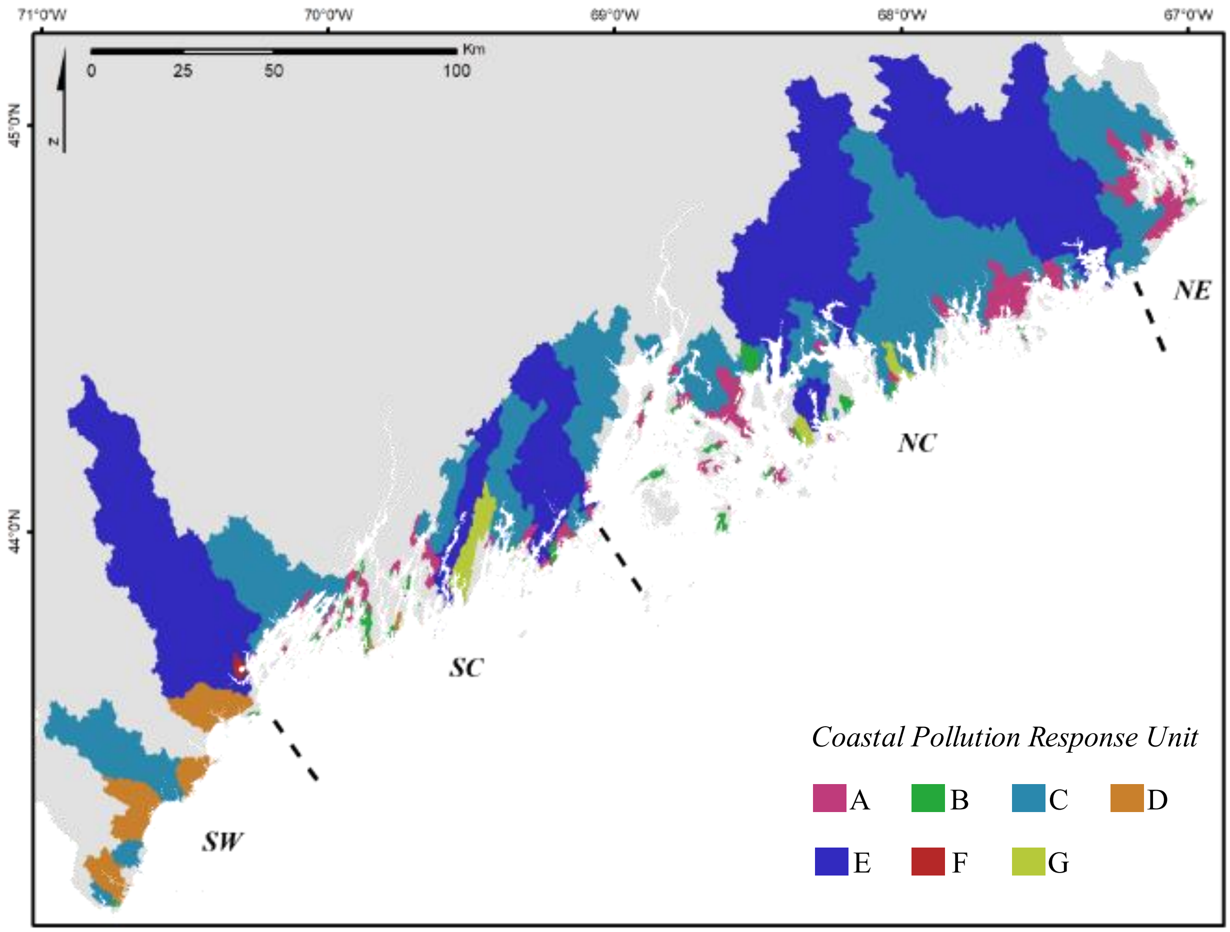

Figure 3.9. Map of identified Coastal Pollution Response Units (CPRUs) for the Maine coast. Dashed lines separate the four coastal compartments of Kelley (1987).

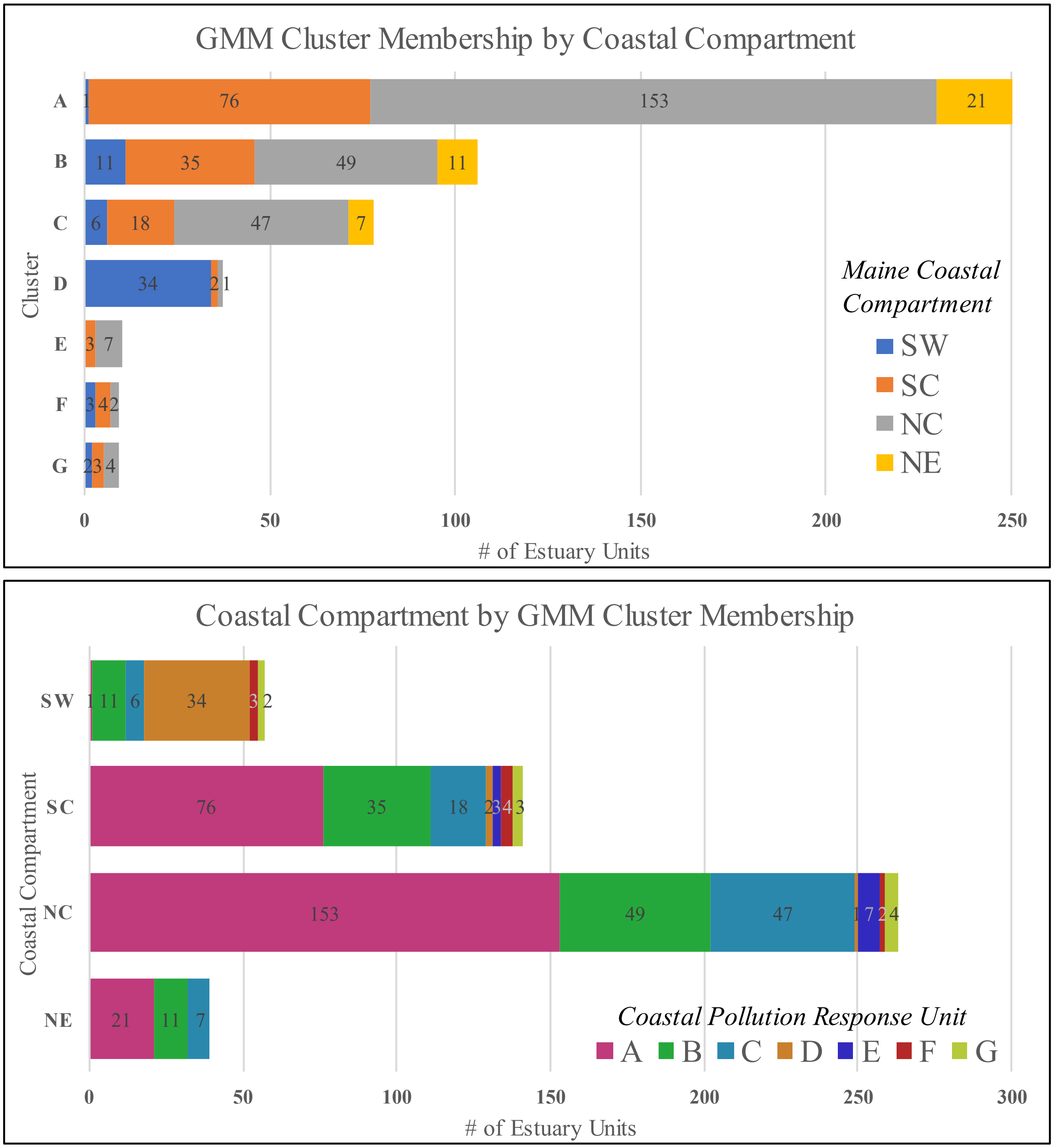

Figure 3.10. Stacked bar charts showing membership breakdowns between CPRUs and the coastal Maine regional compartments of Kelley (1987). Numbers within each bar section are estuary unit counts.

units based on combinations of local site attributes, while also inviting particular focus on marsh-related SDR processes in the SW compartment.

3.4.3. Remaining Data Gaps

The CPRU setting analysis performed here filled coastal spatial data gaps identified by Smith et al.'s (2016) LPRU analysis of nontidal watersheds by incorporating margin watershed areas and embayment areas into the identification framework. Other data gaps for the coastal Maine problem domain noted by Smith et al. (2016) remain, and additional gaps were identified. These include omission of data from the six largest watersheds in Maine and no consideration of stochastic events such as truck spills and infrastructure failures. The latter gap highlights the important point about the setting identification system that any implementation of the system presents a snapshot of conditions at a moment in time defined by its data sources. This points to the importance of updates to land use, demographic, and civil infrastructure data represented in DEMs.

Another important data gap is the absence of spatial data describing dominant in-estuary water circulation patterns across the study domain. This gap necessitates the use of bathymetry to route flow unidirectionally out of embayments and prevented the tool from being used in the Hockomock Bay and lower Sheepscot River region of the study domain (Figure C.1d). This complicates evaluations of estuarine water interaction and mixing within individual embayments, introducing potential error in delineations of the contributing watershed area that affects an embayment or sub-embayment of interest. The effect of changing pour line placement to address this issue is explored more fully in Chapter 4.

3.5. Conclusions

This chapter addressed the development of a diagnostic strategy for identification of coastal settings based on proxy metrics for bacteria SDR processes. Ontological challenges in establishing rules for estuary outlet placements resulted in the need and opportunity to develop a flexible framework for delineating estuary units from any outlet line and identifying their coastal setting type based on aggregated SDR proxy metric data. The successful implementation of this framework as a map-based expert system tool incorporating unsupervised GMM clustering rules created multiple scientific advances. The expert system facilitated delineation and setting identification for 500 estuary units along the length of the coastal Maine study domain, allowing for comparison of local settings based on SDR attributes against regional settings identified in previous research and setting up subsequent investigation of relations between setting and bacteria vulnerability. It also provides a platform for testing the outcomes from different assemblages of SDR metrics and clustering algorithms on cluster membership assignment. The research also identified multiple data gaps that future researchers can address, most notably inadequacy of existing spatial data layers to account for complex in-estuary tidal water circulation patterns across the Gulf of Maine study domain.

CHAPTER 4

COASTAL SETTING VULNERABILITY ASSESSMENT

4.1. Introduction

Information, data, and management decision tools are necessary to address contemporary coastal pollution problems in a sustainability solutions framework (Kates et al., 2001). Vulnerability analyses in sustainability solutions should strive to consider linkages between human and biophysical environmental conditions operating in coupled human environmental systems (Turner, et al. 2003). Coastal resource management thereby requires knowledge of diverse settings shaped by varied geomorphic processes and experienced interventions from human activities over the past two centuries. Related coastal management decision support tools should ideally be designed with an underlying scientific basis built on knowledge of land-sea connections and clear organization strategies to adapt the knowledge into management actions (Taylor, 2018a). These considerations were fundamental to the research presented here with a focus on land-sea connections in Maine, coastal attributes influencing pollution problems, and the development of decision support tools to evaluate vulnerability of coastal settings to pollution problems. Outcomes from the research provide new perspectives on the capacity of spatial data, machine learning approaches, and monitoring data from government agencies to predict the vulnerability of coastal settings to water pollution contamination events.

4.1.1. Background

Prior research on land-sea connections influencing coastal conditions had a goal of informing coastal resource managers of estuary and near shore water quality conditions. Extensive spatial data inventories have been developed to support modeling of watershed hydrology, estuary hydraulics, and biochemical processes in the Chesapeake Bay (Studds et al., 2012; Hood et al., 2021). The complexity of coastal areas influencing biogeochemical fluxes has been evaluated using a clustering and upscaling approach to highlight and compare coastal energy regimes and water residence times (Bartley, Buddemeier and Bennett, 2001). Geographical analyses of bacteria pollution affecting shellfish harvesting areas has been evaluated with respect to seasonal and climatic variables in the Mid-Atlantic region (Leight et al., 2016). Unsupervised cluster analyses based on embayment morphology have been used to identify validated estuary setting types in both the United States and Tasmania (Edgar et al., 2000; Engle et al., 2007).

The Maine coast study location has been the focus of previous work focused on land-sea connections affecting shellfish sanitation and harmful algae blooms (Smith et al., 2016; Gerard, 2018; Alahmed, Ross and Smith, 2022). Non-tidal stream and river watersheds have been used as the unit of analysis to identify “Landscape Pollution Response Units” (LPRUs) in Maine based on proxy spatial data metrics to represent coastal attributes influencing bacteria pollution (Smith et al., 2016; Gerard, 2018; Roy et al., 2018). Fecal coliform samples collected from Maine Department of Marine Resources (MEDMR) monitoring sites near the outlets of watersheds evaluated in the study show differences in median bacteria counts among the five LPRU settings identified in the analysis. Elevated bacteria counts were associated with the LPRU with the most extensive urban conditions (Figure 1.5) (Smith et al., 2016).

Beyond the association of bacteria counts with LPRU settings, the previous coastal Maine research brought attention to several information and data gaps of concern to stakeholders because of the potential influence on bacteria pollution in tidal waters. One prominent information gap was related to land areas immediately adjacent to estuary boundaries that do not fall within the non-tidal stream and river watersheds considered in the initial analyses. The Margin Watershed Areas (MWAs) are assumed to have considerable influence on bacteria pollution in tidal waters for several reasons. The land areas are in close proximity to tidal waters, have short runoff and subsurface flow pathways, coincide with shore bird populations that can defecate in intertidal areas, and they are locations with relatively high human populations that generate wastewater.

The coastal pollution vulnerability analysis that is the target of this research relates coastal settings defined by SDR proxy spatial information to outcomes from bacteria sampling in tidal waters over decadal time scales. Maine DMR maintains water quality testing programs for the purpose of managing shellfish harvest activities relative to sanitation goals (Maine Department of Marine Resources, 2023d). Bacteria data used in these analyses are from more than one thousand MEDMR water quality monitoring stations along the length of the Maine coast. The MEDMR Bureau of Public Health conducts water sampling for fecal coliform bacteria levels at designated locations within shellfish growing areas in accordance with National Shellfish Sanitation Program (NSSP) standards (Maine Department of Marine Resources, 2023b). In addition to systematic random sampling at all stations, sampling may be conducted in conditionally open growing areas during “adverse pollution conditions” historically associated with elevated fecal coliform levels, such as after high-magnitude precipitation events (US Food and Drug Administration, 2019).

4.1.2. Vulnerability Research Objectives

The identification and parameterization of coastal Maine settings using bacteria source, delivery, and residence time (SDR) proxy spatial data metrics related to coastal bacteria pollution is described by Van Dam (Chapter 2). Coastal Maine setting domain and attributes derived from the setting evaluations are referred to as Coastal Pollution Response Units (CPRUs) by Van Dam (Chapter 3) and include non-tidal watersheds, MWAs, and estuary metrics not considered in previous studies of Maine’s coast. Bacteria data used for the analysis is comprised of fecal coliform bacteria sampling results collected and archived by the MEDMR (see e.g., Maine Department of Marine Resources 2019). Relations of the sampling outcomes to coastal settings is performed using estuary pour line delineations to identify the domain of pollution sample outcomes quantified by statistical summaries of sample results. The associations are then used as a basis to identify and compare SDR attributes and coastal Maine settings in terms of vulnerability to bacteria contamination events.

The advancements pursued by this research relates to the assembled information and approaches used for analyses. A more is a more extensive diagnostic collection of proxy spatial data variables representing SDR attributes driving pollution problems has been considered. Machine learning approaches have been selected to improve accommodation of data outliers and clarify the relative influence of individual attributes and SDR pollution culprit categories. A spatial data tool has also been developed to provide more capacity to perform hypothesis testing exercises related to SDR attributes and the grain scale of setting evaluations. Four primary research questions frame the research tasks related to determining associations between coastal attributes, settings, and bacterial pollution vulnerability:

1. Can static proxy spatial data metrics describing SDR drivers of bacterial contamination problems be strategically assembled and related to bacteria sampling data to assess and compare coastal pollution vulnerability?
2. Does equal weighting of SDR pollution culprit categories dilute the capacity to compare coastal bacteria pollution vulnerability compared to evaluations limited to source and estuary attributes?
3. Does the proximity of SDR pollution-related factors in margin watershed areas have a detectable influence in coastal pollution vulnerability analyses?
4. What are the implications of estuary outlet pour line placement on the coastal pollution vulnerability analysis outcomes?

4.2. Methods

Settings represented as whole estuary units, margin watershed areas (MWAs), and estuarine areas defined by SDR attributes were evaluated for associations with bacteria sampling outcomes. Six permutations of coastal setting identification outcomes are considered using different subsets of SDR proxy spatial data metrics and CPRU subunits. Setting identification outcomes are related to MEDMR bacteria data statistics with an added sensitivity analysis to highlight the implications of estuary pour line selection on setting identification. Proxy metric relationships to bacteria vulnerability are also examined using linear regressions and supervised decision tree regression analysis. The relative influence of margin watershed areas on bacteria vulnerability is examined through assessment of results of these analyses.

4.2.1. Bacteria Data

Bacteria levels for each delineated estuary unit must be estimated from sampling data before analysis of coastal settings bacteria vulnerability can occur. DMR fecal coliform count data for monitoring sites are publicly available through the Maine GeoLibrary. Data are presented as yearly aggregate summaries of sample count, geometric mean, geometric standard deviation, maximum, and 90th percentile (P90) bacteria counts, geolocated to the latitude and longitude of each sampling location. Data for calendar years 2016-2018 were selected for analysis to correspond to the general time period in which many of the spatial data layers behind the SDR proxy metrics were gathered.

An estuary unit is intended to represent a contiguous space where SDR processes throughout the contributing area collectively act to affect water quality within the embayment waters. As such, bacteria levels for these analyses are based on all DMR monitoring stations within each unit. Bacteria samples were associated with estuary units by performing a spatial join of delineated estuary unit polygons against sample location points using geographic information system (GIS) software (Figure 4.1). Of the 500 delineated estuary units delineated for analysis, 480 had usable bacteria records. Selected records were exported as a spreadsheet with rows containing an estuary ID number, DMR monitoring station number, year, and bacteria summary data for each joined station point.

Geometric mean (GM) of fecal coliform count (/100 ml water) was selected as the summary statistic for the vulnerability analyses. The NSSP uses GM as a measure of central tendency in the bacteria data (US Food and Drug Administration, 2019). It has the advantages of being less prone to bias from a small number of high-magnitude events than P90. This makes it useful for analysis of long-term vulnerability. It is also easily recoverable from the aggregated summary data available. The geometric mean of a set of n values is the n^{th} root of their product. The total product for that point's bacteria subset is then its GM to the n^{th} power with GM and number of samples known for a monitoring station point. Overall GM for the set of all monitoring stations within an estuary unit is calculated by finding the product of all subset products and taking its N^{h} root, where N is the total number of bacteria samples taken across all sampling sites and years. Total GM across monitoring stations was calculated in this way for all 480 estuary units.

A histogram of GM bacterial counts for the 480 estuary units is right-skewed with a median of 3.7, mean of 4.21, and maximum of 27.7 counts/100 ml (Figure 4.2). For some statistical purposes this would invite an evaluation of outliers and removal of high values from the dataset before proceeding with further analyses. However, it is less likely that they are outliers in the sense of erroneous data because these values are geometric means of many samples over multiple years and not individual sample data points. As a result, data for all 480 estuary units are retained here for analysis.

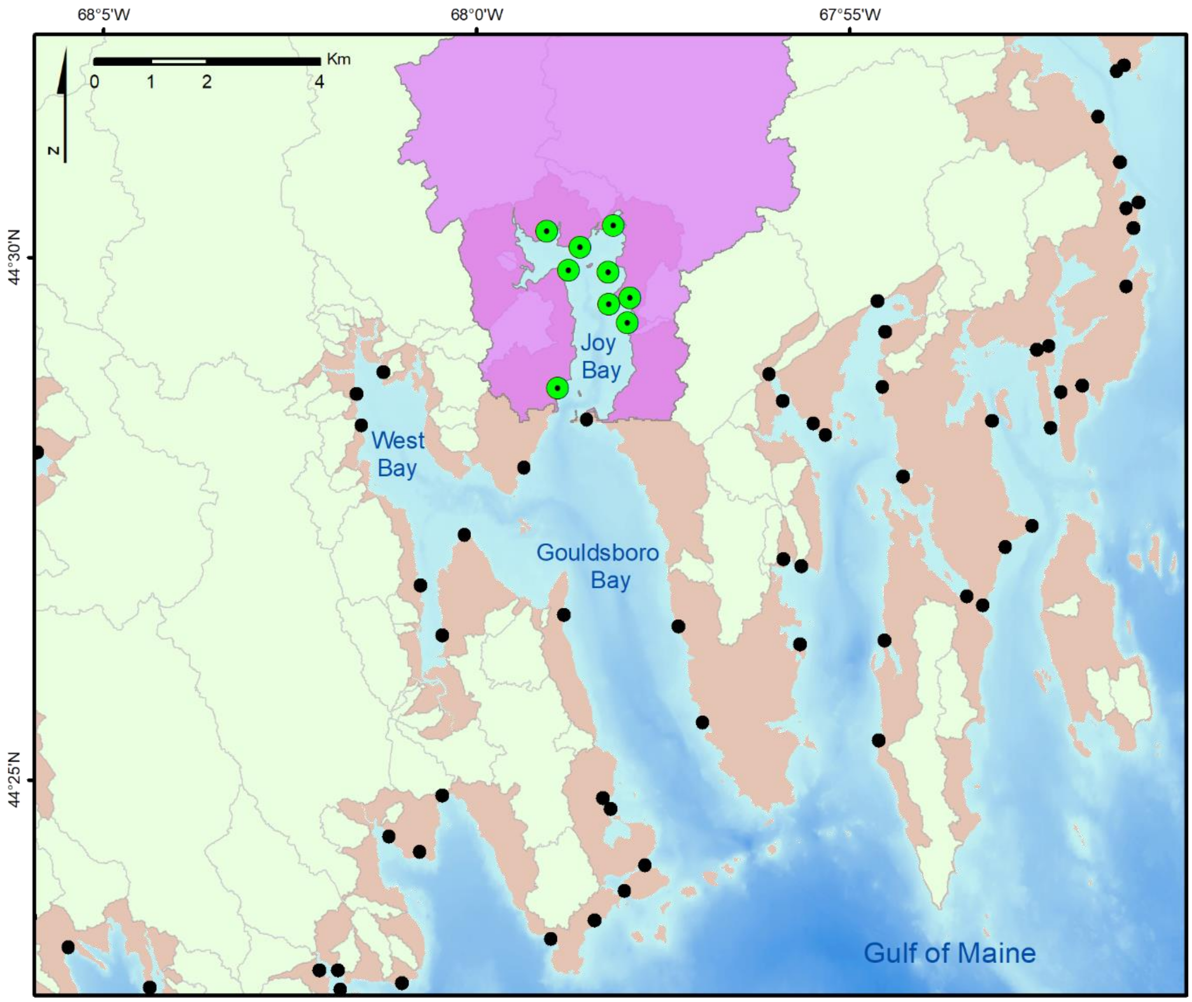

Figure 4.1. Map showing selection of bacteria monitoring stations for a delineated estuary unit. Maine Dept. of Marine Resources (MEDMR) bacteria monitoring sites along the coast are shown as black dots, with nontidal watersheds in yellow and margin watershed areas in brown. For the delineated estuary unit encompassing Joy Bay (purple), bacteria counts are aggregated from all monitoring stations within the estuary unit (highlighted green). Gouldsboro Bay, Gouldsboro and Steuben.

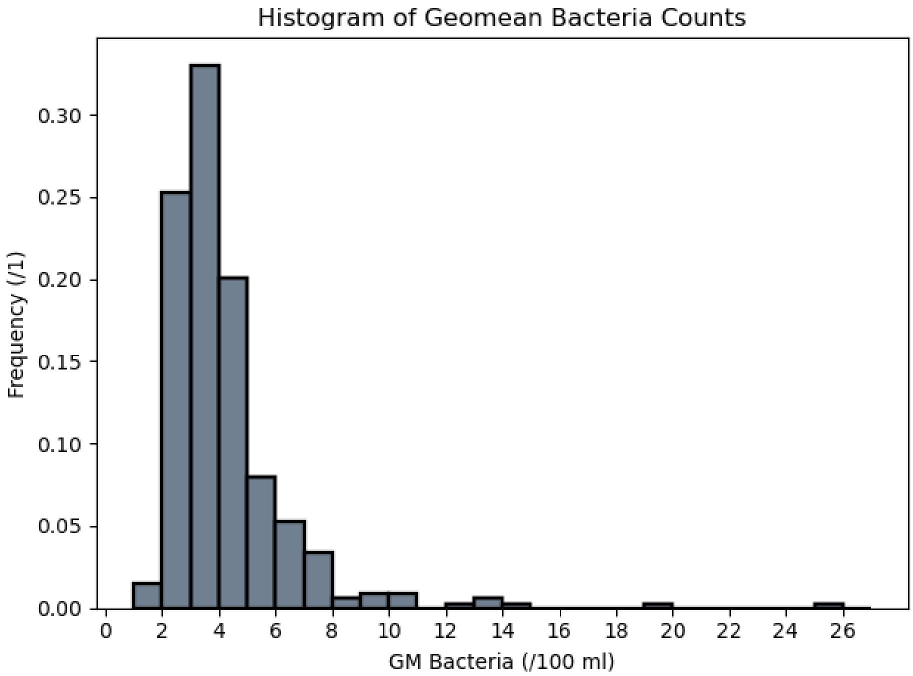

Figure 4.2. Histogram of geometric mean bacteria counts for 480 delineated estuary units included in the analysis.

4.2.2. Bacteria Vulnerability by Cluster

The first research question addressed in this chapter relates to whether identified coastal setting types, based on proxies for SDR processes, display differences in bacteria pollution vulnerability. Van Dam (Chapter 3) described the identification of different setting types for the Gulf of Maine problem domain using Gaussian mixture modeling (GMM) unsupervised clustering of 500 estuary units. The primary set of seven Coastal Pollution Response Units (CPRUs) identified for the research and incorporated into the expert system GIS tool is based on GMM clustering using all SDR proxy metrics, labeled as clusters A-G in Chapter 3 (Figure 3.6). "Signatures" of bacteria pollution vulnerability for each CPRU are assessed as distributions of geometric mean bacteria counts among delineated estuary units within the setting type and presented as box and whisker plots (Smith et al., 2016).

Five additional sets of coastal setting types based on clustering with subsets of the total SDR proxy metrics are also presented. Proxy metrics included in each are given in Table E.1, with the original CPRU analysis listed as (i). The first additional set (ii) omits proxy metric columns related specifically to margin watershed areas, such that the GMM cluster analysis to identify settings incorporates only proxy metrics related to the entire terrestrial contributing watershed area and embayment area morphometry. The next (iii) includes only margin-specific and embayment morphometry proxy metrics. The third (iv) is based only on the nine proxy metrics that relate specifically to estuary morphometry. These first three additional sets of clusters were selected to assess the extent to which coastal settings defined by subsets of SDR proxy metrics focused on different physical spaces within estuary units, including margin areas, are predictive of bacteria pollution vulnerability.

A second assessment is of the role of equal weighting of SDR proxy metrics in bacteria vulnerability prediction. This assessment incorporates set (i), which weights all metrics equally, set (iv), which by focusing on embayment morphometry includes only residence time metrics, and the final two additional permutations, sets (v) and (vi). Set (v) incorporates only proxy metrics for bacteria sources. Set (vi) incorporates only source proxy metrics specifically within MWAs. Assessments are completed through visual comparison of box and whisker plots of bacteria vulnerability "signatures."

4.2.3. Proxy Metric Linear Regression

There is interest in investigating the influence of individual coastal attributes on bacteria sample counts, partly to explain the associations of bacteria problems with specific CPRU settings. The evaluation of individual attributes involves the examination of linear regression relations between bacteria counts and individual proxy metrics (Yan and Su, 2009). Linear regression fits a straight line, defined as $y = mx + b$, through a set of points by calculating a residual distance between the actual y value of each point and predicted y value along the line for the same x, then minimizing the sum of the squares of all residuals (Kenney, 1939). Linear regressions are not expected to indicate that any one metric exerts a controlling effect on bacteria levels in coastal waters in the Gulf of Maine problem domain (Smith et al., 2016). However, the approach provides an easily interpretable method of assessing both direction and strength of correlation between proxy metrics and bacteria.

Regressions were performed in Python using the Scikit-learn statistical module for the 480 delineated estuary units with aggregated bacteria data (scikit-learn developers, 2023a). A best fit line for the response variable of GM bacteria counts is predicted using the explanatory variable of a single proxy metric column for each regression. Regression fit is assessed using the coefficient of determination R^2, to assess the proportion of variance in the response variable that is explained by the independent variable. An R^2 value of 1 indicates perfect fit and $R^2 = 0$ indicates zero correlation (Chicco, Warrens and Jurman, 2021).

4.2.4. Proxy Metric CART Analysis

A more sophisticated method for assessing the relative explanatory power of different proxy metrics on bacteria vulnerability in estuaries is through a classification and regression tree (CART) analysis (Breiman, 1984). CART is a type of supervised machine learning approach that constructs rules to iteratively split a dataset in order to predict a response variable, creating a tree of greater-than / less-than decisions. Hierarchical decision tree approaches have been used in estuarine setting identification, bacteria dynamics research, and other ecological modeling in diverse settings to better capture the complex non-linear interactions among sets of variables than parametric regression approaches (De'ath and Fabricius, 2000; Louis et al., 2003; Smeti et al., 2009; Aertsen et al., 2010; França and Cabral, 2015; Krishna et al., 2021).

CART regression analysis was performed in the Python (version 3.8) computer language using the Scikit-learn (version 1.0.2) statistical module for the 480 delineated estuary units with aggregated bacteria data (scikit-learn developers, 2023b). The target response variable for the CART regression analysis was the geometric mean bacteria counts. The independent variables for this multivariate analysis were the set of all proxy metric columns, such that the model could choose to split a branch on any one at a time. The total dataset of 480 estuary units was split into a training dataset (*X_train, y_train*) with which to fit the decision tree regressor and a test dataset (*X_test, y_test*) with which to assess the performance of the fitted model. A new random subset of 30% (144 estuary units) was retained as the test dataset for each run sequence.

The CART regressor is trained to predict response variables *y_train* from independent variables *X_train* for each model run. A maximum tree depth *d* may be set to limit how many times the tree splits to prevent model overfitting. For this analysis, *d* was allowed to vary between 1 and 20 for each random training subset. Performance of the model is assessed by passing the test independent variables *X_test* to predict test response variable *y_test*. Regression model fit for the test dataset is assessed using the coefficient of determination R^2. The resulting tree can be visualized to identify independent variables on which splits occur, with higher splits indicating variables with greater explanatory power for variation in the dataset related to prediction of the response variable.

4.2.5. Margin Watershed Influence on Vulnerability

Evaluation of the influence of land areas immediately adjacent to estuaries on bacteria pollution vulnerability is approached through comparison of outcomes of the analyses described above for MWA-specific proxy metrics and combinations of metrics describing other spatial components of the estuary units. The suite of analyses have provided highlighted attention to the comparison between outcomes for CPRU permutation *ii* for the fully estuary units and permutation *iii* that includes only MWA and estuary morphometry metrics. For simple linear regressions, comparisons are made between the MWA-specific attributes and their estuary unit equivalents. These include the pair of soil drainage scores

SOILDRAINS_M and SOILDRAINS and the fit of the regression for the for “fraction margins” attribute. The test with the CART analysis is focused on the relative prominence of margin-specific versus other metrics among the splits chosen by the model as the most significant for explaining variance in the dataset. A comparison of setting identifications produced here with the setting identification outcome of Smith et al. (2016) is also considered because of the omission of MWAs in that analysis.

4.2.6. Outlet Line Placement Case Study

An important consideration is the sensitivity of proxy metric values and cluster assignment to line placement because the expert system tool for coastal setting identification allows the delineation of estuary units from any user-defined outlet line. The implications of the pour line placement decisions is investigated through a case study of a bay with relatively simple plan form and multiple plausible embayment and sub-embayment outlet delineations.

Baileys Mistake is a colorfully named embayment in Washington County, along the rural northeastern section of Maine’s coastline popularly referred to as the Bold Coast. The roughly rectangular bay is bordered on land by the town of Lubec to the north and east and the unorganized territory of East Central Washington to the west (Figure 4.3a). Its mouth at a constriction between two headlands is slightly protected by small islands along Baileys Ledge, and is otherwise open to the wider Gulf of Maine to the south-southeast. The embayment has a water surface area of 1.74 km^2 and terrestrial watershed area of 8.66 km^2, of which 33.0% is in MWAs, at this outer limit (Figure 4.3b). There is a second constriction approximately 500 meters into the embayment and two semicircular lobes or sub- embayments positioned farther landward. The western sub-embayment receives freshwater discharge from two nontidal watersheds separated by a roughly 500 m long beach where DMR maintained a sampling site during the 2016-2018 period of record under consideration for this analysis.

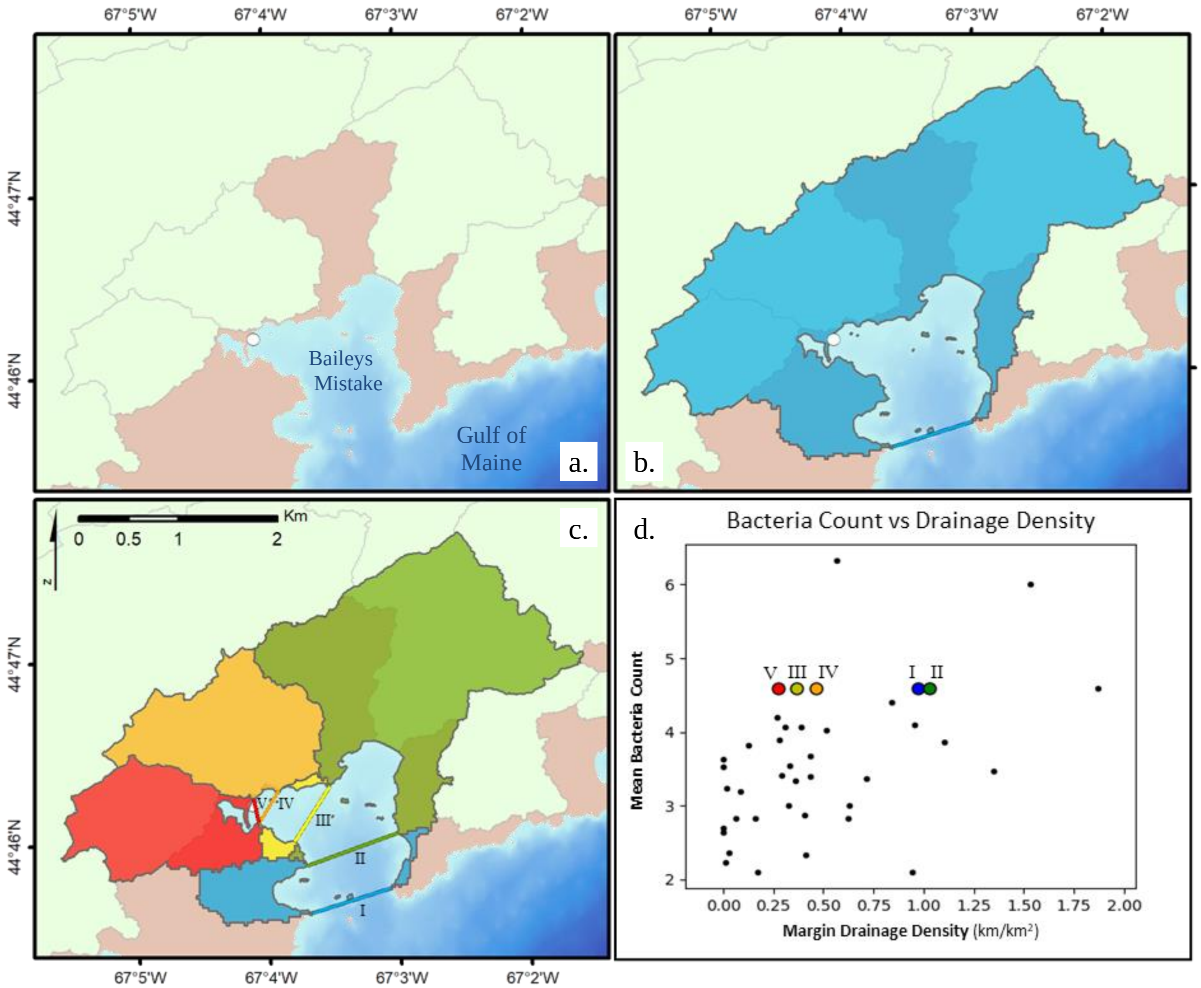

Figure 4.3. Baileys Mistake case study. a) Map of Baileys Mistake, an embayment in Washington County in the northeastern section of Maine's coast that receives surface water runoff from nontidal watersheds (light green) and margin watershed areas (MWAs) (brown). The Maine Department of Marine Resources (DMR) maintains a bacteria monitoring site (white circle) between two watershed outlets in the northwest of the embayment. b) An estuary unit delineated from an outlet line at the outer extents of the embayment has an estuary area of 1.74 km^2 and terrestrial watershed area of 8.66 km2, 33.0% of which is MWAs. c) A series of overlapping watersheds (labeled I-V from outer to inner) represent potential estuary unit delineations for Baileys Mistake to attempt to capture the coupled land-sea dynamics affecting bacteria counts at the single monitoring site. Note that unit I encompasses the area of units II-V as shown in (b), unit II encompasses III-V, etc. d) A plot of aggregated MWA drainage density for each delineated estuary unit vs mean bacteria count from the DMR site monitoring data illustrates how changing outlet line location affects calculated proxy metric values and confounds the relationships between proxy metrics and bacteria data. Black points represent a selection of other delineated embayments in the northeastern region of the Maine coast.

Hydrodynamic patterns and extent of mixing within the embayment are unknown due to the unavailability of spatial data pertaining to currents in the embayment. Estuary units were delineated for five different embayment or sub-embayment outlet lines to assess the effect of outlet line placement on calculated proxy metric values and setting identification output from the expert system. Each delineation represents a plausible separation of the space to isolate land and water that influences bacteria conditions at the sampling site. Selected outlet lines were at the embayment outer extent, the first constriction, across the mouth of the northwest sub-embayment, offshore of the beach to capture discharge from both non-tidal watersheds, and at the sample site capturing discharge from the innermost nontidal watershed (labeled I-V in Figure 4.3c).

4.3. Results

4.3.1. Cluster Associations with Bacteria

Setting clusters based on SDR culprit categories show relatively little variation in their associations with bacteria count “signatures” (Figure 4.4). This holds true for the original CPRUs based on all SDR proxy metrics (i) as well as the additional permutations associated with different spatial components of the estuary unit and bacteria culprit categories. Among large clusters (clusters with relatively high membership counts, such as Clusters A-D of the original CPRUs) median GM bacteria counts per 100 ml generally fall between 3 and 4, with no large cluster median above 5. Clusters (iii) identified from only MWA and embayment morphometry proxy metric columns do not show substantial differences in their associations with the bacteria signatures. All of the six clusters containing bacteria sampling stations had median GM bacteria counts within 1 per 100 ml of each other. Clusters (ii) that excluded margin-specific proxy metrics showed greater variability than the margin-focused clusters. Clusters developed solely from source attributes (v, vi) show little difference in their associations with the bacteria count signatures used in the initial evaluation compared to CPRUs based on equal weighting of

SDR (i). Clustering of source attributes did notably produce more identified setting types than any of the other permutations tested.

Results suggest that setting types defined by unsupervised clustering based on SDR attributes are poor predictors of bacteria contamination vulnerability because of the similar bacteria count signatures associated with CPRUs. However, the consistency in the bacteria signatures among the tested permutations do not support a conclusion that the different collections of attributes considered in the full estuary units, MWAs, or estuary water bodies have no influence on bacteria contamination vulnerability. The initial outcome with full estuary units and all SDR attribute categories indicate that there is a mismatch between metric values important to bacteria pollution levels and the variations in metric values the Gaussian mixture models found important for differentiating clusters. Concern with dilution of the influence from source attributes, MWAs, and estuary conditions within the CPRUs is addressed in the analyses through completed permutations. The similarity of bacteria signatures is consistent among all sets of attributes and subdomains considered. This implies that the limited capacity of the approach to distinguish vulnerable coastal settings by the SDR attributes is constrained by the bacteria sampling data.

For all permutations of proxy metric combinations representing different physical spaces or SDR culprit categories within estuary units, the least variation in bacteria vulnerability signatures among clusters is associated with the largest clusters. Van Dam (Chapter 3) shows that GMM clustering of estuary units based on SDR proxy metrics results in identified settings with uneven member counts. Large clusters tend to exhibit greater ranges of values for individual proxy metrics and more “average” conditions overall. Small clusters (some with fewer than ten members) were generally defined by notable deviations from the others in one or more metrics. The smallest clusters exhibit more unique bacteria signatures across all four GMM sets, and the three clusters with the highest median bacteria counts (ii-5,

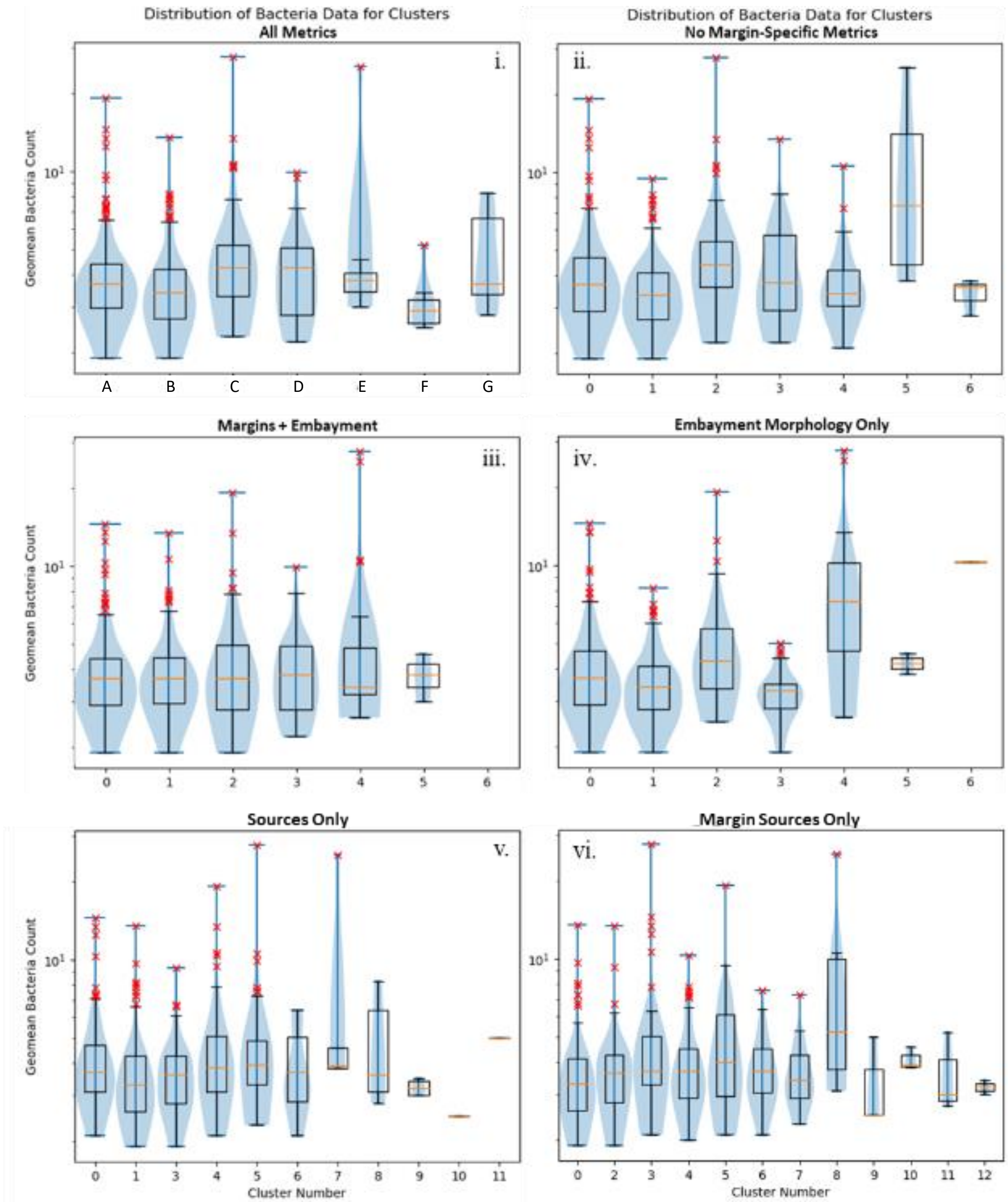

Figure 4.4. Box and whisker plots of bacteria count distribution for four sets of seven coastal setting types based on Gaussian Mixture Model (GMM) clustering of different subsets of source, delivery, and residence time (SDR) proxy metrics. Boxes mark 25th and 75th data percentiles, orange lines 50th percentiles, and red Xes outliers. Underlain violin plot width corresponds to relative membership count for clusters. Clusters are numbered in in order of descending member count for each set of settings. The smallest cluster for the "Margins + Embayment" set of settings (iii) did not have any member sites containing bacteria monitoring stations and thus does not appear.

iv-4, and the one-member iv-6) are all in this group. The outcomes provide useful knowledge for coastal managers because of the greater clarity of the associations related to the smaller clusters, although it is less helpful for the locations included in the large clusters.

4.3.2. Proxy Metric Linear Regression

Simple linear regressions of geometric mean fecal coliform counts against SDR proxy metric values revealed little strength of correlation between bacteria values and individual metrics (Figure 4.5, Table E.2). No metric had a coefficient of determination greater than 0.16. The metrics with the highest correlation with bacteria counts were combined sewer overflow counts and density in overall estuary units, within MWAs, and in areas inundated during highest high tide. These point source attributes related to wastewater had positive regression coefficients (slopes), indicating that higher values for these metrics within an estuary unit are correlated with higher bacteria vulnerability. Regressions for population count and addressable structure count also had relatively high R^2 for the dataset (0.14) but with correlation coefficients (regression line slopes) of zero. Fraction margin had a negative coefficient, indicating that watersheds with a high fraction of MWAs show less vulnerability to bacteria pollution, although with an R^2 value of only 0.04.

Coefficient directions for metrics with $R^2 > 0$ generally conform to assumptions based on knowledge of source, delivery, and residence time processes. Nonpoint source attributes related to increased bacteria generation, such as fraction of the landscape in developed conditions and fraction farmed, were positively correlated with bacteria counts. Attributes expected to associate with relatively lower bacteria counts, such as fraction rural and fraction under conservation status, had negative coefficients. Along with CSO count and density, point source metrics of pollution discharge elimination system facilities (PDES-F) and outfalls (PDES-O) both had positive correlations with bacteria counts. Increases in the capacity for surface runoff to transport bacteria through runoff generation captured

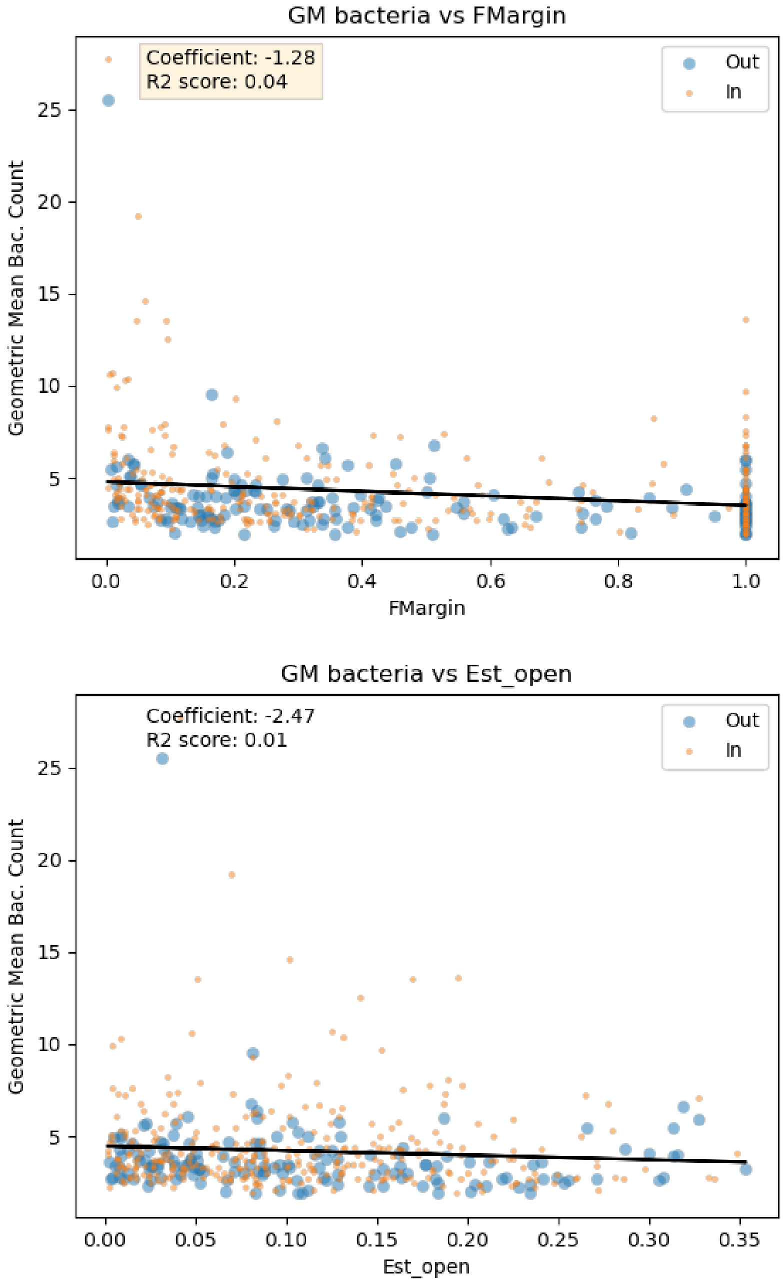

Figure 4.5. Examples of simple linear regression plots of bacteria counts against proxy metrics. Regressions show weak correlations between individual proxies and bacteria levels in estuary units. “Out” and “In” here refer to estuary units delineated for outer embayments whose mouths open directly to Gulf of Maine waters vs inner sub-embayments whose mouths open into larger outer embayments (Figure E.1) and are included for visual reference. Regressions were performed on the unified set of 480 estuary units for each metric.

using soil drainage estimates was positively correlated with bacteria counts. Soil conditions associated with lower rates of runoff generation were negatively correlated. All proxy metric columns for geomorphically-derived and engineered drainage densities with $R^2>0$ were slightly positively correlated with bacteria vulnerability (coeff. < 1), corresponding to more efficient transport of bacteria from land areas into tidal waters. Among residence time proxies, mean estuary depth, estuary openness, and estuary circularity were negatively correlated with bacteria vulnerability; regressions for other metrics were flat or had R^2 of zero.

Some proxy metrics produced counterintuitive results. The fraction of land cover representing surface water storage was positively correlated with bacteria counts despite the assumption that storage would correlate with less efficient transport and delivery of contaminated to tidal waters. Average watershed surface slope was negatively correlated with bacteria vulnerability. Soil storage depth indicating the capacity of the soil to store water before generating runoff during a precipitation event was positively correlated with bacteria vulnerability in contradiction to the trend of other soil drainage metrics. This particular result may be an artefact of poor overall soil depth data in the USDA Soil Survey Geographic Database (SSURGO). Incorrect soil storage values would also affect the Qmin_m3 metric representing the volume of runoff produced during a two-inch precipitation event for an estuary unit after filling soil storage and the ratio of Qmin to estuary volume. Both are calculated using soil storage depth for an estuary unit. The metrics and Qmax_m3 and Qmax_Ev_r calculated only from 2" rainfall without subtracting soil storage had coefficients of 0.00.

4.3.3. Proxy Metric CART Analysis

CART regression analysis to predict bacteria counts from all proxy metric columns was run with maximum tree depths ranging from 1 to 20 splits for 10,000 permutations of the training and test datasets, producing a best fit tree with a maximum depth of just two splits and R^2 of 0.63 (Figure 4.6). This model identified a top-level split on the density of addressable structures in the MWA, separating the training dataset into a group of 318 and a group of 18. Second-level splits were on the ratio of drainage area to estuary area and the fraction of moderately drained soils in the MWA, respectively. The largest of the four training data subsets created by the regression analysis contained 235 (of a total of 336) estuary units, meaning that the best-fit model of 200,000 model initiations resulted in 69.9% of the dataset being assigned a single bacteria count value (GM = 3.67). This appears at first to be an unreasonable result, but revisiting the histogram of bacteria counts for estuary units (Figure 4.2) shows that it is in line with the general data distribution.

Two additional sets of 200,000 model initiations were subsequently run. The first achieved an R^2 value of 0.64 with a maximum tree depth of 6 splits. The second set achieved an R^2 of 0.66 with maximum tree depth of 4 splits. The splits for the first four levels of each analysis are given in Table 4.1. All three best-fit models produced through the CART analyses have different maximum tree depths and arrangement of splits. Several commonalities in proxy metrics selected by the models to explain data variation emerged. The density of addressable structures within margin watershed areas, a proxy for bacteria sources related to human population during the summer tourist season, was selected as a first or second level split by all three models; overall watershed structure density also appeared as a fourth level split in two models. Poorly drained soils in the overall estuary units and in the highest high tide zone, also emerged as common splits as well as multiple variations of drainage density. The ratio of maximum potential surface water runoff volume to estuary volume during a two-inch precipitation event, a proxy metric for freshwater flushing in embayments, was identified by two models as the top-level split. Interestingly, this metric had not emerged as an important explanatory variable for bacteria in simple linear regressions with a coefficient of 0.00 and R^2 of 0.01.

4.3.4. Margin Watershed Influence on Vulnerability

The examination of outcomes from the three analyses revealed mixed results about the extent to which margin watersheds drive variation in estuarine bacteria pollution vulnerability measured as the geometric mean bacteria count in the estuary. Assessment of bacteria distributions by setting types indicated that unsupervised GMM clustering based only on margin and embayment metrics produced clusters that were least sensitive to bacteria. The fraction margin watershed area within an estuary unit explained only 4% of variance in bacteria levels, although this is relatively high explanatory value among a set of metrics where only 25 (of 107) had R^2 of 0.04 or better. Overall, the thirty-two margin specific proxy metrics had a slightly lower R^2 than their corresponding whole-watershed metrics (means of 0.022 and 0.026, respectively). For CART regressions, margin-specific metrics featured in two of the three splits for the first model. The SDR attributes of MWAs or areas inundated by highest high tide (which may be part of either margin or nontidal watershed areas) featured in eight of thirteen splits for the second model and seven of thirteen for the third model, approximately in line with their proportional representation among the whole set of proxy metrics.

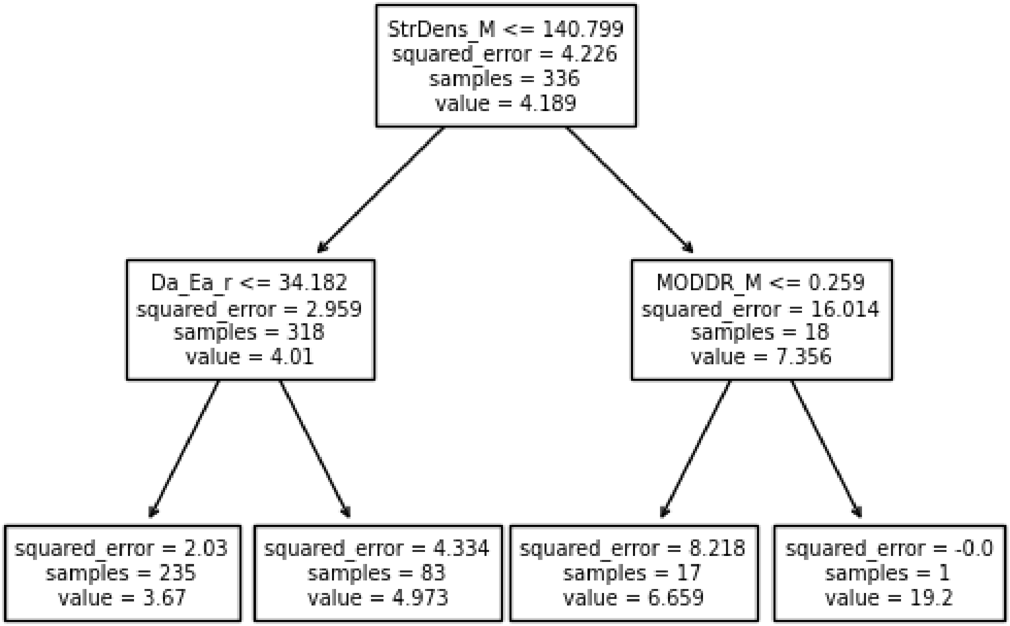

Figure 4.6. Decision tree produced by initial CART regression analysis. This tree had a maximum split depth of 2, producing four subsets from the original training dataset with a maximum subset population of 235 estuary units, and had an R^2 of 0.63 when evaluated against the test dataset.

Table 4.1. Top four decision tree levels for second and third versions of CART regression analysis. Metrics that appear as splits in both analyses are shown in bold.

Decision Tree Level	**Second Analysis**	**Third Analysis**
Top	**Qmax_Ev_r**	**Qmax_Ev_r**
Second	FDEVELO_HH; **StrDens_M**	Zmax; **StrDens_M**
Third	DD_05_HH; FRURAL_HH; **DD_05_Rds**; **POORDRAIN**	DD_2_M; **POORDR_HH**; DD_3_HH; FHighHigh
Fourth	**POORDR_HH** (x2); StrDens_HH; SSTORME_HH; DD_2_Rds; **StructDens**	Bear_Out; FMargin; Est_Circ; **DD_05_Rds**; **StructDens**; **POORDRAIN**; Pop_Dns_M

4.3.5. Outlet Line Placement Case Study

The Baileys Mistake case study reveals that proxy metric values for estuary units delineated to represent a single embayment can vary widely depending on outlet line placement (Table E.3). From the five outlet lines delineated at Baileys Mistake, forty proxy metric columns had percent changes of at least 100% between their lowest and highest values, twenty-five over 200%, thirteen over 500%, and five over 1,000%. The largest changes were among metrics related to estuary size, with the greatest change for any single metric an over 80,000% (800-fold) increase in embayment volume from the innermost (V) to outermost (I) delineated estuary unit. These variations also affect CPRU assignment because identification of coastal setting type for an estuary unit is based on SDR proxy metric values. The estuary units delineated from the five outlet lines were assigned to four different CPRU clusters, with only the two innermost lines assigned to the same cluster.

The pour line case study is simplified by having only one DMR bacteria monitoring station within the embayment. The bacteria count signature for all five potential outlet lines remains constant and only proxy metric values change. Figure 4.3d illustrates an approximately threefold difference in MWA drainage density between the innermost and outermost delineated estuary unit for Baileys Mistake in a plot of bacteria counts vs drainage density. Changing outlet line location would also affect the subset of monitoring stations within the delineated estuary unit and would thus result in movement along both axes of the plot in a more complex situation with multiple bacteria monitoring stations distributed throughout the embayment (Figure 4.1). This would further complicate efforts to assess relations between setting and bacteria pollution vulnerability.

4.4. Discussion

The research presented in this chapter used multiple approaches to evaluate relations between coastal settings defined by bacteria pollution SDR attributes and MEDMR bacteria sample data.

Outcomes from the analyses suggest the pursued approach has limited capacity to distinguish the vulnerability of Maine's varied coastal settings to bacteria contamination problems. CPRU bacteria distributions and linear regressions of bacteria counts against SDR proxy spatial data metrics show surprisingly limited correlations between SDR conditions and bacteria counts. Four permutations evaluating setting associations with bacteria contamination were pursued using all SDR attributes to describe either estuary units, non-tidal watersheds, MWAs, or estuary water bodies. Two additional permutations were examined to evaluate associations between estuary units and MWAs described by only runoff and bacteria source attributes with MEDMR bacteria sample counts. The consistently similar bacteria count signatures among settings considered in each permutation provide evidence that the limited predictive power may be related to bacteria sample data and not the settings described by SDR attributes or estuary subdomains. Exploratory CART regression analyses provided more promising results by indicating some similarities in the groupings of proxy spatial data metrics that can explain variations in bacteria levels.

The investigation of the associations between coastal settings and MEDMR bacteria count signatures produced several important outcomes. The relatively small difference in bacteria count signatures among tested combinations of coastal setting domains and attributes places heightened attention towards the potential weaknesses in bacteria sampling data when used for vulnerability analyses. The approaches used to scrutinize the influence of the coastal setting domains and attribution choices demonstrate a procedure through which data gaps might be closed to enhance the vulnerability predictions. Spatial information and data analysis tools created for the research operations provide support for decision-making by investigators. The flexibility of the created platform for coastal setting identification can also offer utility to coastal managers by adding geographic information data and capacity for setting determinations based on factors governing water pollution problems.

The accommodation of research and management interests fulfills the central research ambition to design an expert system to identify coastal settings and land-sea connections influencing pollution problems affecting shellfish harvesting areas. Development of the expert system designed to assess the vulnerability of varied coastal settings to bacteria pollution required that practical and theoretical research questions be addressed, one of which is the feasibility of using existing bacteria sampling data for the vulnerability analyses. The four related research questions framing this component of the research can now be revisited.

<u>Question 1:</u> Can static proxy spatial data metrics describing SDR drivers of bacterial contamination problems be strategically assembled and related to bacteria sampling data to assess and compare coastal pollution vulnerability?

Limitations in research outcomes related to this question can be explained by several information gaps constraining the capacity of the approach to consider dynamics and processes linked to bacteria pollution to distinguish the vulnerability of varied settings to pollution problems in coastal estuaries.

1. *Estuary hydrodynamics*. There were difficulties in the analysis with the placement of outlet lines to delineate estuary units that act as contiguous land-sea connection domains. The analysis is sensitive to pour line placement because it influences both CPRU determination and bacteria sampling outcomes. Detailed hydrodynamic modeling has revealed prevailing circulation patterns in a small number of coastal Maine bays (Alahmed, Ross and Smith, 2022; Bailey et al., 2023), but such data do not currently exist for the majority of the coast. The ability of proxy spatial data metrics to describe coastal morphology has limits related to spatial and temporal dynamics at varied scales.
2. *Spatial and temporal variation in runoff production*. Runoff production in a landscape is a product of precipitation input and infiltration losses governed by soil types and impervious

surface (Dunne and Leopold, 1978; Gupta, 2017; Bierman and Montgomery, 2020). The analytical approach lumped proxy spatial data within watersheds and estuary units despite awareness that the factors related to runoff production vary spatially and temporarily. Spatial data can represent large differences in runoff based on land cover. The detection of more subtle differences across a domain and over time can be much less certain. Spring freeze/thaw and snow pack dynamics create added complexity, including spring freshet runoff.

3. *Bacteria reproduction and die-off dynamics*. Fecal coliforms and other culprit bacteria are living organisms which feed, reproduce, and die in aquatic environments with dynamics dependent on local environmental conditions, affecting the magnitude and length of pollution problems. These conditions include nutrient availability, salinity, temperature, ultraviolet radiation, and predation, which vary in space and time (Jones, 2011; Leight et al., 2016; Rothenheber and Jones, 2018; Korajkic et al., 2019).
4. *Stochastic events*. Stochastic bacteria pollution events constitute random occurrences that introduce bacteria to the system, particularly in large quantities. Examples include pipe breaks (Sutton, Sczesny and Cone, 2023), sewage truck spills (Blakely, 2018; Walsh, 2022), and animal waste spills (PCD Live Briefs, 2023). Inclusion of these events was outside of the scope of the research presented here, but their potential importance has been suggested for the problem domain (Smith et al., 2016; Roy et al., 2018).
5. *Large river and ocean influences*. Large rivers were identified as a data gap by Smith et al. (2016), who noted that during high magnitude discharge events their outflow plume may “drape” the coast and carry transported constituents along shorelines and into adjacent embayments (Cole et al., 2020; Du et al., 2020). Figures Figure 4.7 and Figure 4.8 illustrate the difference in extent between turbidity plumes from the Union River, the second-largest (1,380 km^2) nontidal watershed included in the current analysis, and the 15,200 km^2 Kennebec River, one of six very large watersheds excluded from analysis. Large river plumes and other landward movement of

contaminants from ocean or gulf waters into estuaries causes a potential disconnect between pollution levels and estuary unit SDR conditions, confounding efforts to assess bacteria pollution vulnerability based on static spatial data representations of setting conditions.

6. *Bias in bacteria sampling datasets*. The MEDMR water quality sampling regimens are dictated by management goals related to NSSP standards rather than random sampling of the full Maine coastal domain. Sampling is focused on areas known to have occasional pollution issues where conditional shellfish flat closures are more likely to be needed to maximize management efficacy of limited sampling capacity. Areas with historically low risk of adverse pollution events receive less targeted sampling. Much of the tidal shoreline in “prohibited” harvesting status due to known persistent bacteria problems or proximity to certain points sources is rarely or never tested, including areas adjacent to the urbanized landscape of the state’s largest city, Portland, and other developed areas with high seasonal populations such as Old Orchard Beach, Boothbay Harbor, Rockland, Camden, and Belfast (Maine Department of Marine Resources, 2022), limiting the ability of analyses to identify proxy metrics that correlate with persistently high bacteria levels.

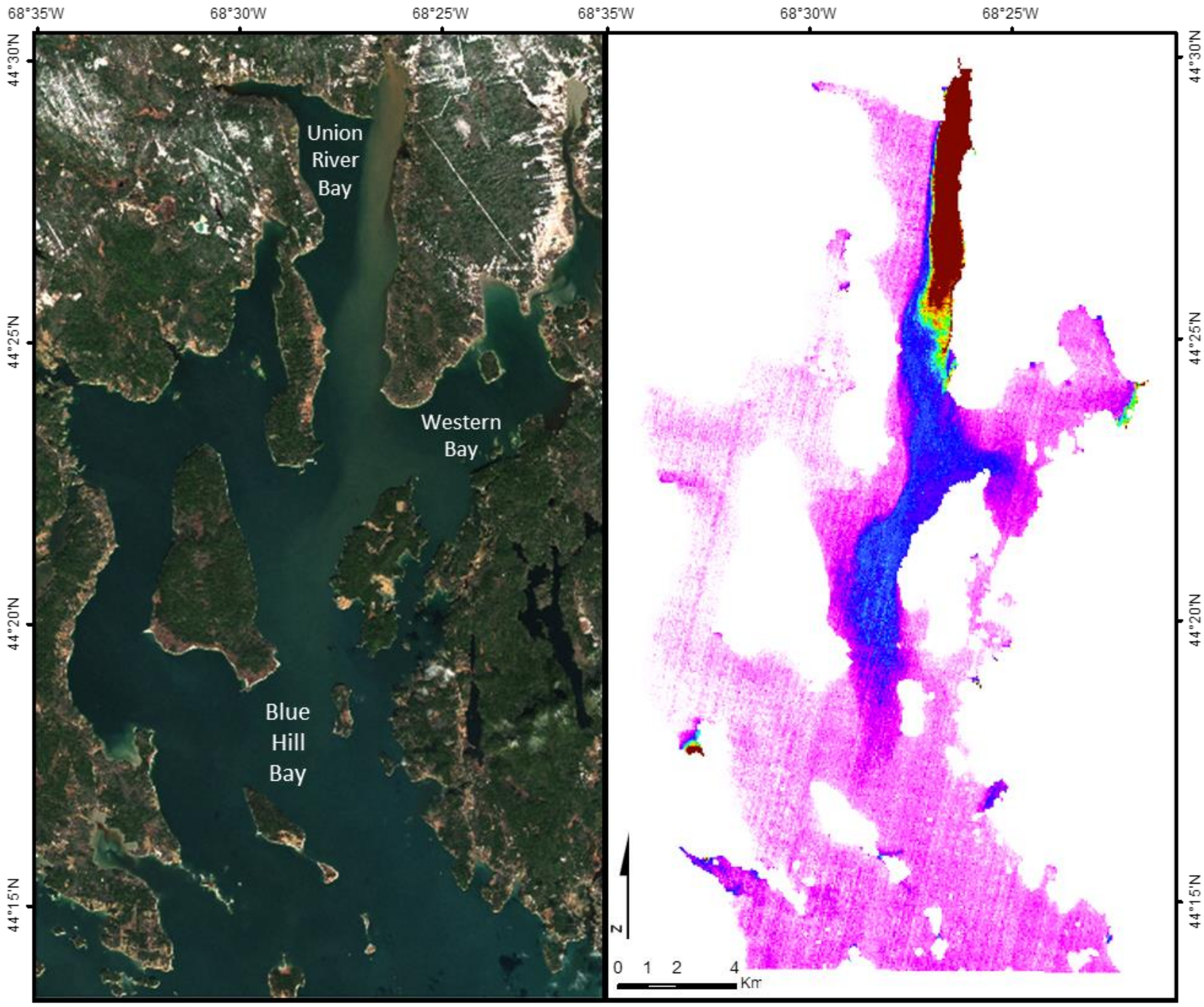

Figure 4.7. Sediment plume from Union River (Ellsworth, ME). The plume from the Union River, the second-largest nontidal watershed included in analysis for this research, spilling through Union River Bay into Western Bay and greater Blue Hill Bay after large December rain event. (left) Satellite image from Copernicus Sentinel-2A. (right) Image of turbidity levels post-processed by author from Sentinel-2A data.

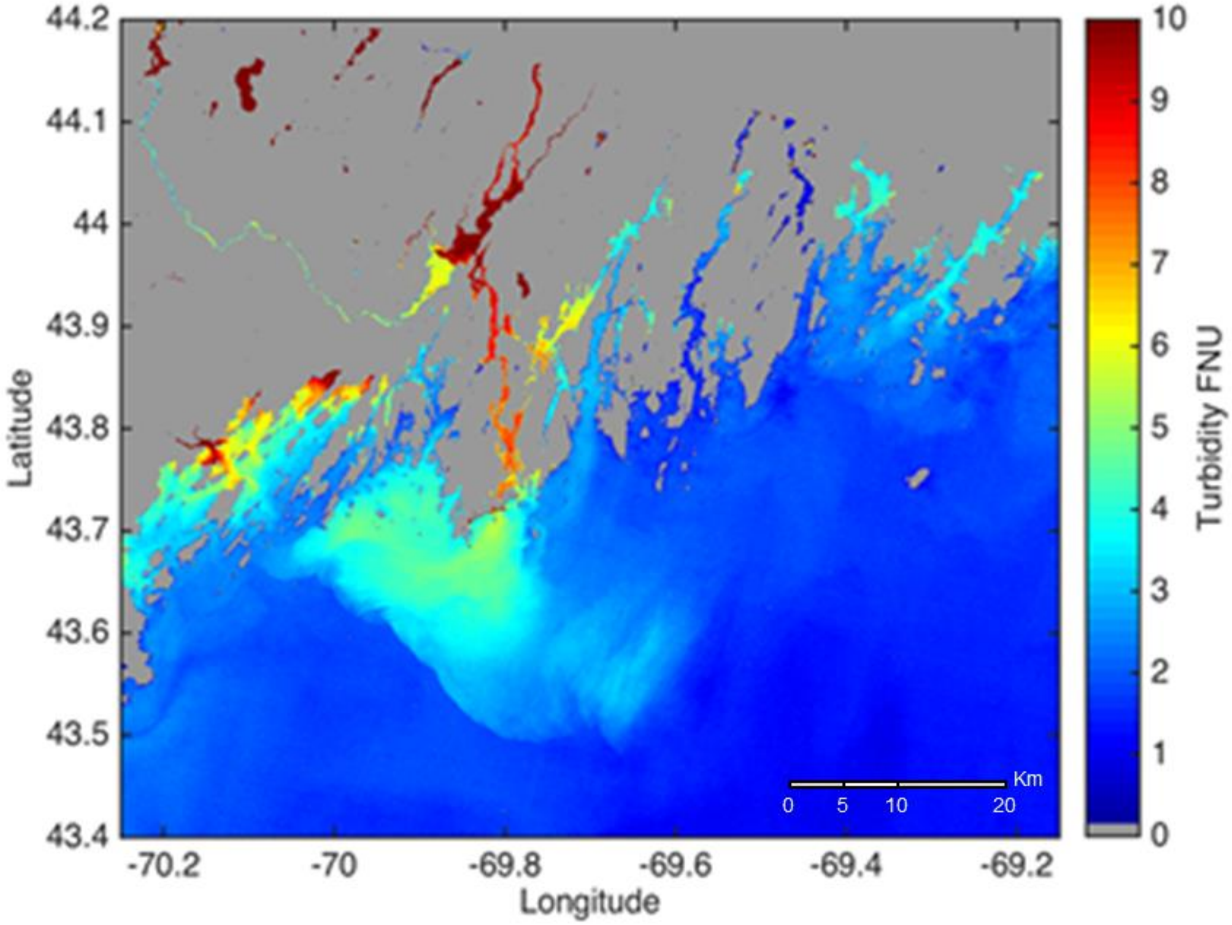

Figure 4.8. Turbidity plume from the Kennebec River draping over eastern Casco Bay. The Kennebec is one of six very large Maine rivers excluded from analysis (adapted from Ross and Smith, 2020; University of Maine Satellite Oceanography Data Lab).

Question 2: Does equal weighting of SDR pollution culprit categories dilute the capacity to compare coastal bacteria pollution vulnerability compared to evaluations limited to source and estuary attributes?

Comparison of CPRUs derived from equal weighting of all SDR pollution culprit categories against clusters derived from only source or residence time categories did not indicate major differences in bacteria count distribution patterns. This indicates that equal weighting of SDR culprit categories for coastal setting identification did not diminish the capacity of clustering outcomes to predict bacteria pollution vulnerability. Relatively poor predictive capacity among clusters derived from all permutations of proxy metrics must therefore be a result of a different data gap, such as issues with bacteria sampling data biases. A difference that did arise from the use of source-only proxy metrics was the expansion of the number of coastal setting types selected by the GMM cluster analysis, from seven clusters for CPRUs with equal SDR weighting to twelve for source-only metrics and thirteen for source metrics specifically within MWAs. This may be useful from a management standpoint by allowing finer discrimination of source culprits within clusters. This warrants further exploration of the possibility of incorporating separate clustering of source, delivery, and residence time proxy metrics into the decision support tool rather than calculating a single setting type based on equal SDR weighting. This would be in line with the general strategies for coastal setting classification used in decision support tools such as that of (Hume et al., 2007), which reports estuary site classification as a combination of hydrodynamic class, geological category, and land cover category rather than as a single combined category.

Question 3: Does the proximity of SDR pollution-related factors in margin watershed areas have a detectable influence in coastal pollution vulnerability analyses?

The evaluation of margin watershed influence on coastal bacteria pollution vulnerability analyses indicates that, based on the statistical methods used here, margin watershed areas do not appear to play a dominant role in pollution vulnerability. This is an unexpected result that should be interpreted with caution. Connections of coastal land areas to tidal waters have been recognized as being important to estuary habitat and water quality over the past century, inspiring the US Coastal Zone Management Act in 1972 and the establishment of coastal zone management programs and regulations at state levels (Anderson, 1975; Davis, 1987; Maine Coastal Program, 2023; NOAA Office for Coastal Management, no date). Physical attributes of the MWAs such as close proximity to tidal waters, susceptibility of low lying areas to flooding, and even seabird bacteria sources in tidal marsh environments suggest that these areas should play an important role in bacteria vulnerability. These unexpected outcomes may be explained by the confounding factors related to deficiencies in the bacteria sampling data in coastal areas and capacity of the reliance on spatial data to describe dynamic processes driving estuary vulnerability. Future work should focus on filling these data gaps to test these results before major changes in coastal management strategies are considered.

<u>Question 4:</u> What are the implications of estuary outlet pour line placement on the coastal pollution vulnerability analysis outcomes?

The question of the effect of changing outlet line pour placement for delineating estuary units involved an investigation of the modifiable areal unit problem (MAUP) for the coastal Gulf of Maine problem domain. MAUP is a problem in spatial statistics where changing the size, shape, or orientation of the spatial units used for aggregation or sampling of underlying data changes the results of the sampling output, potentially even the direction of trends observed from sampled data (Openshaw, 1984; Dungan et al., 2002; Buzzelli, 2020). This is a concern for analyses using spatial units delineated by the expert system framework for setting identification developed for this research, which allows users to make delineations based on any drawn line. A defense against MAUP is the use of spatial meaningful units (Buzzelli, 2020). However, without accurate estuary flow pattern data it is difficult to assess *a priori* to what extent an estuary unit delineated from a given outlet line does act as a meaningful unit for SDR processes and selection of bacteria monitoring stations.

The Baileys Mistake case study clarifies some of the outcomes from the vulnerability analyses relating coastal settings to bacteria sampling data by showing the effect of outlet line placement on calculated metric values. Line placement has an effect both on aggregated proxy metric values and on the selection of bacteria monitoring stations that affects calculated bacteria levels. All areas may not interact equally to collectively drive bacteria conditions in delineated estuaries that are too large or have complex branching structures. Conversely, small estuaries and sub-embayments with high openness values may be subject to tidal exchange influences from an encompassing embayment or the ocean that cannot be accounted for in the analysis.

4.5. Conclusions

The analytical approach and expert system derived from this research advance the capacity to consider and evaluate spatial information relative to available pollution sampling data. Outcomes are not intended to serve as a forecast model over short time scales such as a single coastal storm event. The approach provides a framework for the identification and comparison of settings based on proxy metric spatial data to support both research and management decision-making. The system meets the demands for decision support tools to guide watershed and coastal land-sea connection research investigations related to water pollution. It also addresses coastal management needs in its capacity to support decision-making related to deployment of monitoring resources, development of regulations for seafood harvesting, and guidance for comprehensive planning of coastal areas.

Conclusions about bacteria pollution vulnerability for different setting types in the Gulf of Maine problem domain are ultimately constrained by the available spatial information and bacteria sampling data. The absence of spatial information that describes complex circulation patterns within estuaries was identified as one of the factors limiting comparisons of the coastal settings. Associations of bacteria sample signatures with outcomes from multiple permutations of coastal setting attributes and estuary unit subdomains indicate that equal weighting of pollution SDR culprit categories are not the primary reason the system does not reveal substantially varied vulnerability by location along the Maine coast. The problem lies in the inadequacies of the bacteria sampling dataset designed for shellfish sanitation management operations and not vulnerability research purposes. Suggestions for addressing the bacteria sampling deficiencies to accommodate coastal bacteria pollution vulnerability assessments are provided.

CHAPTER 5

COASTAL POLLUTION VULNERABILITY RESEARCH OUTCOMES

5.1. Introduction

Understanding linkages between human and environmental systems and conditions is at the heart of sustainability science to meet societal and ecosystem needs (Turner et al., 2003). The vulnerability of coupled human–environment systems is a central consideration in the communication between science and decision-making necessary for modern coastal resources management. Research in sustainability science by (Alonso Roldán et al., 2019) draws attention to boundaries between social and ecological systems operating at the land-sea interface that produce deficiencies in collaborative research, science communication, and governance. The authors point out that the generation of scientific knowledge is often not approached integrally across land-sea boundaries. Failure to do so can place constraints on the successes of approaches based on ecosystem services and interconnected whole systems to support decision-making and governance related to coastal resources management.

Determining vulnerabilities of systems to environmental and human changes and identifying where the vulnerabilities occur is central to the sustainability solutions that are the inspiration for this research. Several problems related to bacteria pollution vulnerability are posed by the social-ecological system at the Gulf of Maine land-sea boundary. The original research framed around the scientific basis for decision-making with a focus on shellfish harvesting area closures in Maine was inspired by a coastal sustainability problem related to shellfishing business and communities. Complex land-sea connections and human influence on modern conditions create the biophysical problem of illness-causing fecal-borne pathogens in harvested shellfish. Coastal resource management agencies working in complex social-ecological systems at the land-sea interface face policy problems related to competing needs among shellfish harvesters, consumers, and other community members. Addressing these problems can be complicated by boundaries separating decision-making processes involving academic research groups and staff implementing policies within terrestrial or marine environments. A knowledge and information problem is at the center of the other three problems, which provided an inspiration for the focus of this research on applications of spatial data proxies for solutions to coastal pollution problems.

Ambition to address the problems led to the overarching question driving the research: "How can an expert system domain, metrics, and analysis approach be framed to identify coastal settings and land-sea connections influencing pollution problems affecting shellfish harvesting areas?" It was an added challenge to recognize and address the boundaries in the social-ecological systems at the land-sea boundary in support of an approach to coastal management that incorporates ecosystem services and connected whole systems. Strategic objectives of the research were framed by four research questions listed below:

1. What are the appropriate proxy metrics related to source, delivery, and residence time (SDR) that capture the processes leading to bacteria pollution problems in estuaries?
2. How can coupled land-sea connection settings be compared using spatial data information to evaluate their relative vulnerability to bacteria pollution?
3. What role do the coastal margin watershed areas play in bacteria pollution vulnerability (i.e., does proximity of certain pollution-related factors to the estuary matter)?
4. What are the implications of the estuary outlet pour line placement on the outcome from the expert system to identify coastal vulnerability in varied settings?

The research presented in this dissertation approached land-sea connections related to coastal bacteria from a decision science framework that sought to link data with stakeholder needs and policy objectives. Chapter 1 identified goals and objectives related to identification of SDR proxy metrics, identification and comparison of coastal settings, evaluation of margin watershed area roles in pollution vulnerability, and evaluation of implications of pour line placement on expert system outcomes. Chapter 2 addressed the identification and parameterization of spaces within a domain using SDR attributes related to pollution problems. Chapter 3 addressed the development of an expert system framework for identification and comparison of coastal setting types ("coastal pollution response units," CPRUs) based on SDR metrics in the parameterized problem domain. Chapter 4 summarized the testing of expert system outcomes and the coupling of analysis outcomes with Maine Dept. of Marine Resources (MEDMR) bacteria data in relation to bacteria pollution vulnerability.

5.2. Research Outcomes

5.2.1. Expert System Development

A key outcome of the research was the development of an expert system for identifying coastal setting types. This was not an original objective during project conceptualization, but instead emerged as an important research goal through collaboration with stakeholders. Achieving this outcome required the accomplishment of several functional objectives related to the development of the necessary spatial data needed to facilitate setting identification and pollution vulnerability analyses in the coastal Maine problem domain (Figure 5.1). Data development presented in this dissertation provides a roadmap for researchers seeking to delineate and parameterize spatial domains with proxy metrics related to runoff-borne constituents.

Development of the expert system provided advancements of two different types. The first narrowly relate to the system's implementation as interactive geospatial tool (Figure 5.2). This tool had direct uses as a delineator and data aggregator for researchers and as a decision support tool for coastal managers. More broadly, the research also represents scientific advancements in the form of the set of decisions that were made regarding the parameterization of the space, organization of data, choice of analytical approaches, and formatting of output. The tool also provides a scientific platform for

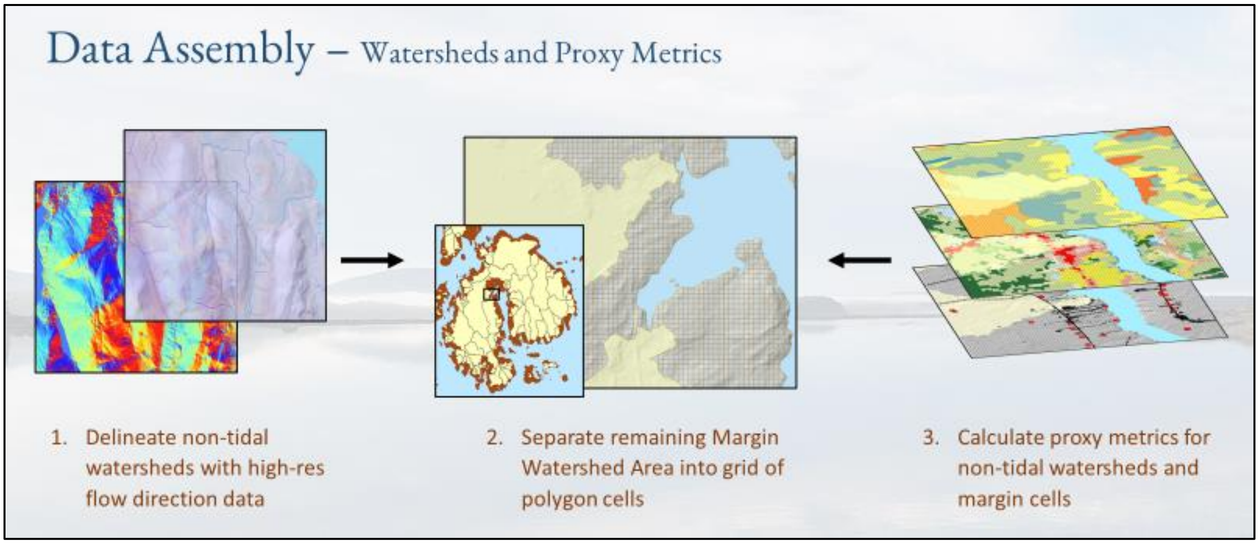

Figure 5.1 Illustration of major data assembly steps supporting development of the expert system for coastal setting identification.

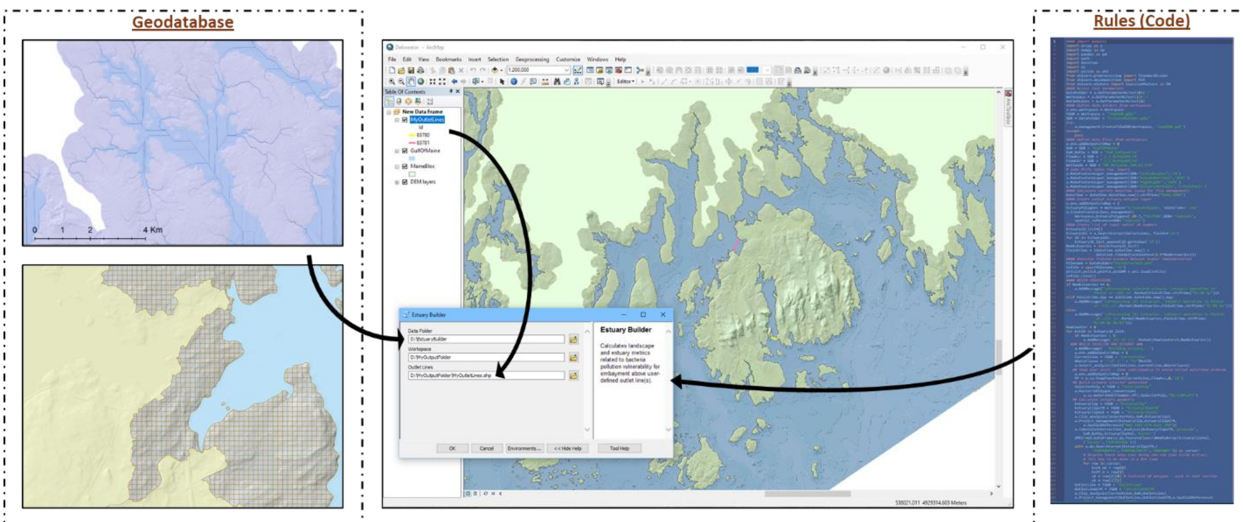

Figure 5.2 Desktop interface for the expert system for coastal setting identification. The implementation of the expert system as a GIS tool relies on a geodatabase containing spatial data (left) and expert rules for assembling and interpreting data in the form of computer code (right).

assessment of additional setting identification approaches and testing of research questions such as the implications of flexible estuary outlet line placement on aggregated proxy metric values and resulting setting identification.

5.2.2. Coastal Setting Identification

Results from experimental analyses conducted with the expert system customized for coastal bacteria pollution targets produced a new specialized map of the Maine coast (Figure 5.3). This is the first thematic map for coastal Maine with a focus on bacteria contamination vulnerability that incorporates the entire coastal subdomains of nontidal watersheds, margin watershed areas, and estuarine waters. It builds on many different forms of thematic and reference maps that have been produced for these coastal areas, including the geomorphologically-based coastal compartments of Kelley (1987) and the SDR categorizations of Gerard (2018) and Smith et al. (2016).

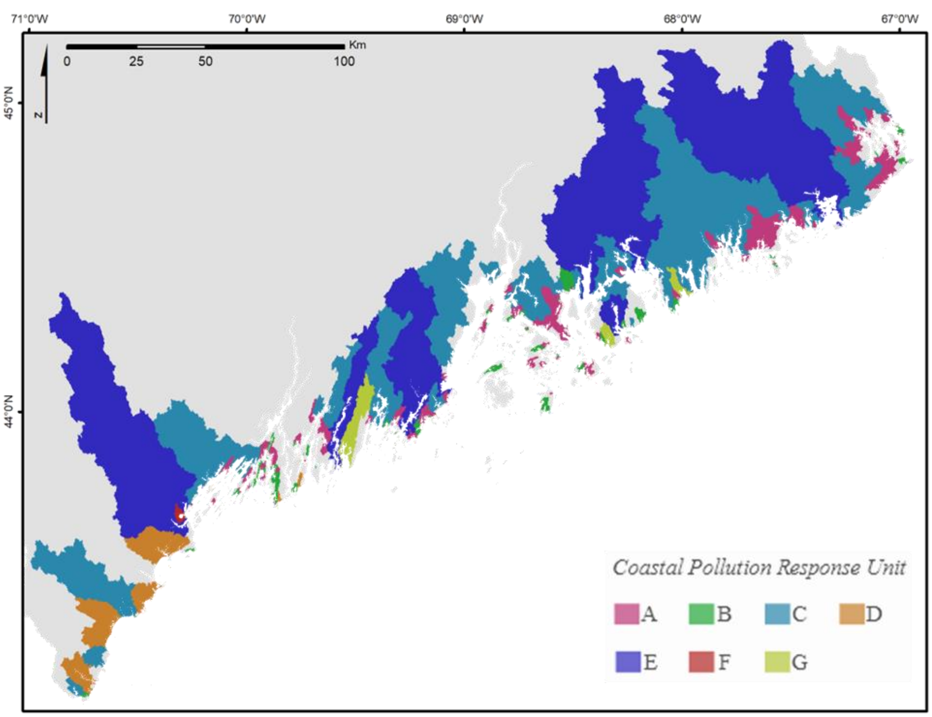

Figure 5.3 Map of Maine's coastline with polygons identifying CPRUs related to estuary unit subdomains and attributes describing pollution source, delivery, and residence time of bacteria pollution.

5.2.3. Pollution Vulnerability

Experimental evaluations of the association of identified coastal settings and SDR attributes with bacteria data signatures highlighted approaches to compare collections of attributes linked to pollution problems and estuary unit subdomains. The investigation of different permutations of attributes and subdomains demonstrated methods for evaluation of how individual SDR attributes or varied collections of attributes describing source, delivery, and residence time influence coastal pollution. Attention towards whole land-sea connection domains or exclusive focus on subdomains such as land areas adjacent to tidal waters or estuary water bodies allowed for comparisons of different coastal locations. Examination of a suite of permutations indicate that standard monitoring protocols used for shellfish sanitation and management are inadequate for bacteria pollution vulnerability analyses.

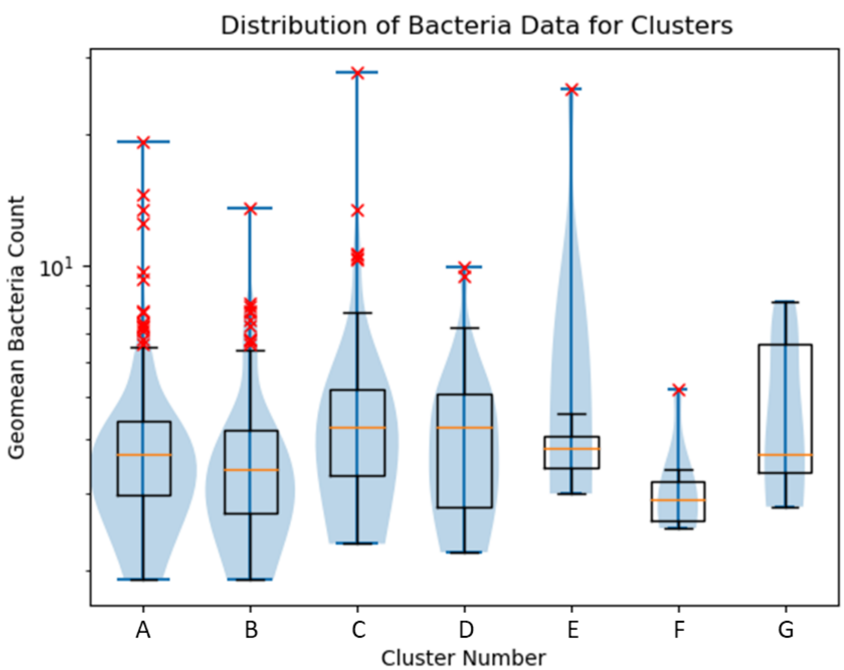

Figure 5.4 Example violin plot with box and whiskers showing bacteria count signatures derived from sampling stations in identified coastal settings.

5.3. Research Conclusions

The outcomes of the research presented in this dissertation advance the relations between coastal attributes influencing pollution problems, spatial information systems, and bacteria sampling data collected by a coastal resource management agency. The work creates the most comprehensive coastal spatial dataset related to land-sea connections and coastal pollution problems for Maine. The evaluation of the associations of settings with bacteria data provides a framework for distinguishing locations based on spatial data proxies with relevance to pollution problems. Aside from tangible products, this research contributed several intellectual advancements. These include the establishment of problem domains, the expansion of the amount and type of land area included in the analysis, and research decisions related to the use of spatial data proxy information to address problems. The research tested individual attributes and the cumulative influence of collections of attributes related to SDR processes. It required the development of an intellectual framework for putting together the components of the solutions to pollution problems related to complex land-sea connections. Finally, it identified potential applications and critically assessed shortcomings and limitations of the research outcomes.

The expert system created as part of this research investigation provides a flexible framework to evaluate land-sea connections and identify coastal settings based on drivers of bacteria pollution problems. It also provided a platform for stakeholder access to spatial data and determinations of coastal setting characteristics to support decisions related to monitoring and land use planning. The platform offers a consistent mechanism for communication of information and results from statistical evaluations of coastal conditions related to pollution problems. The observations derived from use of the platform have addressed the four problems inherently connected to social-ecological systems at the land-sea interface by focusing on knowledge and information. In particular, the sustainability problem is addressed through attention to multiple conditions influenced by humans and biophysical processes with advancements directed towards ecosystem services and whole connected land-sea systems in coastal areas. The manner in which coupled coastal land-sea system produce pollution problems and their sensitivity to pollution problems through scenario evaluations has been demonstrated, conforming to suggestions by Turner, et al. (2003). Components of the coupled system have also been examined to test the influence of equal weighting of pollution culprit categories first conceptualized by Smith et al. (2016). The created expert system establishes a new linkage between information, knowledge, and decision-making for coastal resource managers.

REFERENCES

Aertsen, W. *et al.* (2010) 'Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests', *Ecological Modelling*, 221(8), pp. 1119–1130. Available at: https://doi.org/10.1016/j.ecolmodel.2010.01.007.

Alahmed, S., Ross, L. and Smith, S.M.C. (2022) 'Coastal Hydrodynamics and Timescales in Meso-Macrotidal Estuaries in the Gulf of Maine: a Model Study', *Estuaries and Coasts* [Preprint]. Available at: https://doi.org/10.1007/s12237-022-01067-9.

Alonso Roldán, V. *et al.* (2019) 'Are we seeing the whole picture in land-sea systems? Opportunities and challenges for operationalizing the ES concept', *Ecosystem Services*, 38, p. 100966. Available at: https://doi.org/10.1016/j.ecoser.2019.100966.

Anderson, H. (1975) *Delineation of Coastal Zone Management Units Using Ecological and Legal Criteria: The Case of Atlantic Canada*. University of Rhode Island. Available at: https://doi.org/10.23860/thesis-anderson-henry-1975.

Athearn, K. (2008) 'Economic Impact of Maine's Shellfish Industry'. University of Maine at Machias. Available at: https://digitalcommons.library.umaine.edu/maine_env_organizations/173.

Bailey, T. *et al.* (2023) 'Linking Coastal Circulation to Harmful Algal Blooms with Measurements & Modeling'. *Maine Sustainability and Water Conference*, Augusta, ME.

Baker, C.M. *et al.* (2022) 'From Climate Change to Pandemics: Decision Science Can Help Scientists Have Impact', *Frontiers in Ecology and Evolution*, 10, p. 792749. Available at: https://doi.org/10.3389/fevo.2022.792749.

Barnes, M.D. *et al.* (2019) 'Decision analysis to support wastewater management in coral reef priority area', *Marine Pollution Bulletin*, 148, pp. 16–29. Available at: https://doi.org/10.1016/j.marpolbul.2019.07.045.

Bartley, J.D., Buddemeier, R.W. and Bennett, D.A. (2001) 'Coastline complexity: a parameter for functional classification of coastal environments', *Journal of Sea Research*, p. 11.

Beven, K. and Freer, J. (2001) 'Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology', *Journal of Hydrology* [Preprint].

Bierman, P.R. and Montgomery, D.R. (2020) *Key concepts in geomorphology*. Second edition. Austin ; Boston ; New York ; Plymouth: Macmillan Learning.

Blakely, N. (2018) *Sewer-cleaning truck spills 1,200 gallons of sewage into storm retention pond*, *Battle Creek Enquirer*. Available at: https://www.battlecreekenquirer.com/story/news/local/2018/04/04/sewer-cleaning-truck-spills-1-200-gallons-sewage-into-storm-retention-pond/487263002/ (Accessed: 24 July 2023).

Borns, H.W. *et al.* (2004) ‘The deglaciation of Maine, U.S.A.’, in *Developments in Quaternary Sciences*. Elsevier, pp. 89–109. Available at: https://doi.org/10.1016/S1571-0866(04)80190-8.

Braun, J. *et al.* (2021) ‘National Water Model Application in Maine Coastal Water Quality Management’. *Maine Sustainability and Water Conference,* Augusta, ME.

Breiman, L. (ed.) (1984) *Classification and Regression Trees*. Belmont, Calif: Wadsworth International Group (The Wadsworth statistics/probability series).

Brown, C.J. *et al.* (2019) ‘A guide to modelling priorities for managing land-based impacts on coastal ecosystems’, *Journal of Applied Ecology*. Edited by J. Blanchard, 56(5), pp. 1106–1116. Available at: https://doi.org/10.1111/1365-2664.13331.

Brown, C.M. *et al.* (2017) ‘A High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal Bacteria in a Lake Superior Estuary’, *Environmental Science & Technology*, 51(15), pp. 8263–8271. Available at: https://doi.org/10.1021/acs.est.7b01353.

Buitinck, L. *et al.* (2013) ‘API design for machine learning software: experiences from the scikit-learn project’. arXiv. Available at: http://arxiv.org/abs/1309.0238 (Accessed: 16 July 2023).

Buto, S.G. and Anderson, R.D. (2020) *NHDPlus High Resolution (NHDPlus HR)—A Hydrography Framework for the Nation*. Fact Sheet Fact Sheet 2020–3033. U.S. Geological Survey. Available at: https://doi.org/10.3133/fs20203033.

Buzzelli, M. (2020) ‘Modifiable Areal Unit Problem’, in *International Encyclopedia of Human Geography*. Elsevier, pp. 169–173. Available at: https://doi.org/10.1016/B978-0-08-102295-5.10406-8.

Campos, C.J.A., Kelly, L.T. and Banks, J.C. (2023) ‘Using a weight of evidence approach to identify sources of microbiological contamination in a shellfish-growing area with “Restricted” classification’, *Environmental Monitoring and Assessment*, 195(4), p. 529. Available at: https://doi.org/10.1007/s10661-023-11041-y.

Canfield, K.N., Mulvaney, K. and Chatelain, C.D. (2022) ‘Using researcher and stakeholder perspectives to develop promising practices to improve stakeholder engagement in the solutions-driven research process’, *Socio-Ecological Practice Research*, 4(3), pp. 189–203. Available at: https://doi.org/10.1007/s42532-022-00119-5.

Casella, A. *et al.* (2023) ‘The Road to Load Estimations for Frenchman Bay Water Quality Assessments’. *Maine Sustainability and Water Conference*, Augusta, ME.

Chen, X. and Nordhaus, W.D. (2011) ‘Using luminosity data as a proxy for economic statistics’, *Proceedings of the National Academy of Sciences*, 108(21), pp. 8589–8594. Available at: https://doi.org/10.1073/pnas.1017031108.

Chicco, D., Warrens, M.J. and Jurman, G. (2021) ‘The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation’, *PeerJ Computer Science*, 7, p. e623. Available at: https://doi.org/10.7717/peerj-cs.623.

Clark, S. *et al.* (2019) ‘Pseudo-nitzschia bloom dynamics in the Gulf of Maine: 2012–2016’, *Harmful Algae*, 88, p. 101656. Available at: https://doi.org/10.1016/j.hal.2019.101656.

Cole, K.L. *et al.* (2020) ‘River plume source-front connectivity’, *Ocean Modelling*, 150, p. 101571. Available at: https://doi.org/10.1016/j.ocemod.2020.101571.

Cox, E.P. (1927) ‘A Method of Assigning Numerical and Percentage Values to the Degree of Roundness of Sand Grains’, *Journal of Paleontology*, 1(3), pp. 179–183. Available at: https://www.jstor.org/stable/1298056 (Accessed: 5 July 2023).

Cronin, H., Smith, S. and Fisher, N. (2022) ‘Starting Fresh! Gathering New Information on Freshwater Flow Pouring into Maine’s Estuaries’. *Maine Sustainability and Water Conference*, Augusta, ME.

Davis, C.A. (1987) ‘A strategy to save the Chesapeake shoreline’, *Journal of Soil and Water Conservation*, 42(2).

De’ath, G. and Fabricius, K.E. (2000) ‘Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis’, *Ecology*, 81(11), pp. 3178–3192. Available at: https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2.

Delevaux, J.M.S. *et al.* (2018) ‘A linked land-sea modeling framework to inform ridge-to-reef management in high oceanic islands’, *PLOS ONE*. Edited by C.A. Chen, 13(3), p. e0193230. Available at: https://doi.org/10.1371/journal.pone.0193230.

Demattê, J.A.M. *et al.* (2020) ‘Bare Earth’s Surface Spectra as a Proxy for Soil Resource Monitoring’, *Scientific Reports*, 10(1), p. 4461. Available at: https://doi.org/10.1038/s41598-020-61408-1.

Dheenan, P.S. *et al.* (2016) ‘Geographic information systems and multivariate analysis to evaluate fecal bacterial pollution in coastal waters of Andaman, India’, *Environmental Pollution*, 214, pp. 45–53. Available at: https://doi.org/10.1016/j.envpol.2016.03.065.

Ding, C. and He, X. (2004) ‘K-means clustering via principal component analysis’, in *Twenty-first international conference on Machine learning - ICML ’04*. *Twenty-first international conference*, Banff, Alberta, Canada: ACM Press, p. 29. Available at: https://doi.org/10.1145/1015330.1015408.

Drucker, D. and Williams, S.A. (2003) ‘When Does Water Find the Shortest Path Downhill? The Geometry of Steepest Descent Curves’, *The American Mathematical Monthly*, 110(10), pp. 869–885. Available at: https://doi.org/10.1080/00029890.2003.11920028.

Du, J. *et al.* (2020) ‘Massive pollutants released to Galveston Bay during Hurricane Harvey: Understanding their retention and pathway using Lagrangian numerical simulations’, *Science of The Total Environment*, 704, p. 135364. Available at: https://doi.org/10.1016/j.scitotenv.2019.135364.

Dungan, J.L. *et al.* (2002) ‘A balanced view of scale in spatial statistical analysis’, *Ecography*, 25(5), pp. 626–640. Available at: https://doi.org/10.1034/j.1600-0587.2002.250510.x.

Dunne, T. and Leopold, L.B. (1978) *Water in environmental planning*. San Francisco: W. H. Freeman.

Dyer, K.R. (1973) *Estuaries: a physical introduction*. London, New York: John Wiley & Sons Ltd.

Edgar, G.J. *et al.* (2000) ‘The conservation significance of estuaries: a classification of Tasmanian estuaries using ecological, physical and demographic attributes as a case study’, *Biological Conservation*, p. 15.

Ellis, K.L. and Waterman, M. (eds) (1998) *The Maine Clam Handbook*. Orono, Maine: Maine/New Hampshire Sea Grant College Program. Available at: http://www.bremenmaine.org/wp-content/uploads/2014/09/98-Clam-Handbook.pdf (Accessed: 21 May 2020).

Engle, V.D. *et al.* (2007) ‘A Classification of U.S. Estuaries Based on Physical and Hydrologic Attributes’, *Environmental Monitoring and Assessment*, 129(1–3), pp. 397–412. Available at: https://doi.org/10.1007/s10661-006-9372-9.

Esri (2016) *How To: Create a watershed model using the Hydrology toolset*, *Esri Technical Support*. Available at: https://support.esri.com/en/technical-article/000012346 (Accessed: 6 May 2020).

Esri (2021a) *A quick tour of ArcPy—ArcMap | Documentation*, *ArcMap Documentation*. Available at: https://desktop.arcgis.com/en/arcmap/latest/analyze/arcpy/a-quick-tour-of-arcpy.htm (Accessed: 16 July 2023).

Esri (2021b) *An overview of the Zonal toolset*, *ArcMap Documentation*. Available at: https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/an-overview-of-the-zonal-tools.htm (Accessed: 16 July 2023).

Esri (2022) *A quick tour of creating tools with Python*, *ArcMap Documentation*. Available at: https://desktop.arcgis.com/en/arcmap/latest/analyze/creating-tools/a-quick-tour-of-creating-tools-in-python.htm (Accessed: 16 July 2023).

Evans, K.S. *et al.* (2016) ‘Measuring the impact of pollution closures on commercial shellfish harvest: The case of soft-shell clams in Machias Bay, Maine’, *Ocean & Coastal Management*, 130, pp. 196–204. Available at: https://doi.org/10.1016/j.ocecoaman.2016.06.005.

França, S. and Cabral, H.N. (2015) ‘Predicting fish species richness in estuaries: Which modelling technique to use?’, *Environmental Modelling & Software*, 66, pp. 17–26. Available at: https://doi.org/10.1016/j.envsoft.2014.12.010.

Fredston-Hermann, A. *et al.* (2016) ‘Where Does River Runoff Matter for Coastal Marine Conservation?’, *Frontiers in Marine Science*, 3. Available at: https://doi.org/10.3389/fmars.2016.00273.

Geary, P. and Lucas, S. (2019) ‘Contamination of estuaries from failing septic tank systems: difficulties in scaling up from monitored individual systems to cumulative impact’, *Environmental Science and Pollution Research*, 26(3), pp. 2132–2144. Available at: https://doi.org/10.1007/s11356-018-1364-0.

Gerard, B.R. (2018) *Stream Dynamics in the Headwaters of Post-Glacial Watershed Systems*. University of Maine. Available at: https://digitalcommons.library.umaine.edu/etd/2948.

Goodwin, K. *et al.* (2021) ‘Predicting freshwater inflows to Frenchman Bay, Maine’. *Maine Sustainability and Water Conference*, Augusta, ME.

Gupta, R.S. (2017) *Hydrology and hydraulic systems*. 4th ed. Long Grove, Ill: Waveland Press.

Hillyer, G.V. *et al.* (2022) 'Using a Stakeholder-Engaged Approach to Understand and Address Bacterial Transport on Soft-Shell Clam Flats', *Estuaries and Coasts*, 45(3), pp. 691–706. Available at: https://doi.org/10.1007/s12237-021-00997-0.

Hood, R.R. *et al.* (2021) 'The Chesapeake Bay program modeling system: Overview and recommendations for future development', *Ecological Modelling*, 456, p. 109635. Available at: https://doi.org/10.1016/j.ecolmodel.2021.109635.

Hume, T.M. *et al.* (2007) 'A controlling factor approach to estuary classification', *Ocean & Coastal Management*, 50(11–12), pp. 905–929. Available at: https://doi.org/10.1016/j.ocecoaman.2007.05.009.

Jackson, C.T. (1837) *The Geology of the State of Maine* (3 vol). Augusta, ME: Smith & Robinson.

Jackson, P. (1986) 'Introduction to expert systems'. Available at: https://www.osti.gov/biblio/5675197.

Jennings, S., McCormack, C. and Sheane, R. (2020) *Performance metrics for key sustainability issues*. 3Keel.

Jenson, S.K. (1985) 'Automated derivation of hydrologic basin characteristics from digital elevation model data', in *Auto-Carto VII: Proceedings of the digital representations of spatial knowledge*. *Auto-Carto VII: Digital Representations of Spatial Knowledge*, Washington D.C.: Cartography and Geographic Information Society, pp. 301–310. Available at: https://pubs.er.usgs.gov/publication/70236860.

Jenson, S.K. and Domingue, J.O. (1988) 'Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis', *Photogrammetric Engineering & Remote Sensing*, 54(11), pp. 1593–1600.

Jones, S. (2011) *Microbial Pathogens and Biotoxins: State of the Gulf of Maine Report*. 443, p. 24.

Kates, R.W. *et al.* (2001) 'Sustainability Science', *Science*, 292(5517).

Kelley, J.T. (1987) 'An Inventory of Coastal Environments and Classification of Maine's Glaciated Shoreline', in *Glaciated Coasts*. Elsevier, pp. 151–176. Available at: https://doi.org/10.1016/B978-0-12-257870-0.50010-6.

Kelley, J.T., Belknap, D.F. and Claesson, S. (2010) 'Drowned coastal deposits with associated archaeological remains from a sea-level "slowstand": Northwestern Gulf of Maine, USA', *Geology*, 38(8), pp. 695–698. Available at: https://doi.org/10.1130/G31002.1.

Kenney, J.F. (1939) *Mathematics of Statistics*. Third Printing. New York: D. Van Nostrand Company, Inc.

Korajkic, A. *et al.* (2019) 'Persistence and Decay of Fecal Microbiota in Aquatic Habitats', *Microbiology and Molecular Biology Reviews*, 83(4), pp. e00005-19. Available at: https://doi.org/10.1128/MMBR.00005-19.

Krishna, K. *et al.* (2021) 'Hydrological regulation of Vibrio dynamics in a tropical monsoonal estuary: a classification and regression tree approach', *Environmental Science and Pollution Research*, 28(1), pp. 724–737. Available at: https://doi.org/10.1007/s11356-020-10486-9.

Leight, A.K. *et al.* (2016) ‘Climate relationships to fecal bacterial densities in Maryland shellfish harvest waters’, *Water Research*, 89, pp. 270–281. Available at: https://doi.org/10.1016/j.watres.2015.11.055.

Li, J. and Wong, D.W.S. (2010) ‘Effects of DEM sources on hydrologic applications’, *Computers, Environment and Urban Systems*, 34(3), pp. 251–261. Available at: https://doi.org/10.1016/j.compenvurbsys.2009.11.002.

Lichter, M., Zviely, D. and Klein, M. (2010) ‘Morphological patterns of southeastern Mediterranean river mouths: The topographic setting of the beach as a forcing factor’, *Geomorphology*, 123(1–2), pp. 1–12. Available at: https://doi.org/10.1016/j.geomorph.2010.05.007.

Lidberg, W. *et al.* (2017) ‘Evaluating preprocessing methods of digital elevation models for hydrological modelling’, *Hydrological Processes*, 31(26), pp. 4660–4668. Available at: https://doi.org/10.1002/hyp.11385.

Lipscomb, S.W. (1998) *Hydrologic classification and estimation of basin and hydrologic characteristics of subbasins in central Idaho*. Report 1604. Available at: https://doi.org/10.3133/pp1604.

Lischeid, G. *et al.* (2017) ‘Forensic hydrology: what function tells about structure in complex settings’, *Environmental Earth Sciences*, 76(1), p. 40. Available at: https://doi.org/10.1007/s12665-016-6351-5.

Louis, V.R. *et al.* (2003) ‘Predictability of Vibrio cholerae in Chesapeake Bay’, *Applied and Environmental Microbiology*, 69(5), pp. 2773–2785. Available at: https://doi.org/10.1128/AEM.69.5.2773-2785.2003.

Maina, J. *et al.* (2012) ‘Linking coral river runoff proxies with climate variability, hydrology and land-use in Madagascar catchments’, *Marine Pollution Bulletin*, 64(10), pp. 2047–2059. Available at: https://doi.org/10.1016/j.marpolbul.2012.06.027.

Maine Coastal Program (2023) *Maine Guide to Federal Consistency Review*. 5th Edition-Update 5, revision 2. Augusta, ME: Maine Department of Marine Resources.

Maine Department of Environmental Protection (2018) ‘DEP Issue Profile: Overboard Discharges (OBDs)’. Available at: https://www.maine.gov/dep/water/wd/OBD/index.html.

Maine Department of Marine Resources (2019a) ‘MaineDMR Public Health - 2018 P90 Scores’. Maine GeoLibrary. Available at: https://maine.hub.arcgis.com/datasets/cc9de0dc5fe54a9aba4fdde566d05125 (Accessed: 22 July 2023).

Maine Department of Marine Resources (2019b) *Most Recent Maine Commercial Landings, Commericial Fishing Landings Data*. Available at: https://www.maine.gov/dmr/commercial-fishing/landings/index.html (Accessed: 4 May 2020).

Maine Department of Marine Resources (2021) *Shellfish Water Quality Classifications and Aquaculture Leases Map*. Available at: https://maine.maps.arcgis.com/apps/webappviewer/index.html?id=3b3450693fe54bf497004b153e536de8 (Accessed: 12 July 2021).

Maine Department of Marine Resources (2022) *Interactive Shellfish Closure Map*. Available at: https://www.maine.gov/dmr/fisheries/shellfish/shellfish-closures-and-aquaculture-leases-map (Accessed: 24 July 2023).

Maine Department of Marine Resources (2023a) *About the Maine Coastal Program, Maine Coastal Program*. Available at: https://www.maine.gov/dmr/programs/maine-coastal-program/about-the-maine-coastal-program (Accessed: 1 July 2023).

Maine Department of Marine Resources (2023b) *Growing Area Program: Water Quality*. Available at: https://www.maine.gov/dmr/fisheries/shellfish/shellfish-growing-area-classification (Accessed: 19 July 2023).

Maine Department of Marine Resources (2023c) *Maine Shellfish Closures & Monitoring Data (Webmap)*. Available at: https://maine.maps.arcgis.com/apps/webappviewer/index.html?id=3b3450693fe54bf497004b153e536de8 (Accessed: 6 August 2023).

Maine Department of Marine Resources (2023d) *Shellfish Sanitation and Management*. Available at: https://www.maine.gov/dmr/fisheries/shellfish (Accessed: 22 July 2023).

Mallin, M.A. *et al.* (2001) 'Demographic, landscape, and meteorological factors controlling the microbial pollution of coastal waters', in J.W. Porter (ed.) *The Ecology and Etiology of Newly Emerging Marine Diseases*. Dordrecht: Springer Netherlands, pp. 185–193. Available at: https://doi.org/10.1007/978-94-017-3284-0_17.

Marcot, B.G. *et al.* (2012) 'Recent advances in applying decision science to managing national forests', *Forest Ecology and Management*, 285, pp. 123–132. Available at: https://doi.org/10.1016/j.foreco.2012.08.024.

Martin, J. *et al.* (2023) 'Decision science as a framework for combining geomorphological and ecological modeling for the management of coastal systems', *Ecology and Society*, 28(1), p. art50. Available at: https://doi.org/10.5751/ES-13696-280150.

Maxwell, K., Hubbell, B. and Eisenhauer, E. (2019) 'Institutional insights on integrating social and environmental science for solutions-driven research', *Environmental Science & Policy*, 101, pp. 97–105. Available at: https://doi.org/10.1016/j.envsci.2019.08.003.

McGreavy, B. (2015) 'When water quality is the easy problem: An untold story of sustainability and human well-being'. *Senator George J. Mitchell Center Seminar Series*, Orono Maine, 23 November. Available at: https://vimeo.com/147586955 (Accessed: 4 May 2020).

Merriam-Webster.com (2023) *setting*, *Merriam-Webster*. Available at: https://www.merriam-webster.com (Accessed: 3 August 2023).

Montgomery, D.R., Grant, G.E. and Sullivan, K. (1995) 'WATERSHED ANALYSIS AS A FRAMEWORK FOR IMPLEMENTING ECOSYSTEM MANAGEMENT', *Journal of the American Water Resources Association*, 31(3), pp. 369–386. Available at: https://doi.org/10.1111/j.1752-1688.1995.tb04026.x.

National Marine Fisheries Service (2016) *Fisheries of the United States, 2015*. NOAA Current Fishery Statistics No. 2018. Silver Spring, MD: U.S. Department of Commerce. Available at: https://www.st.nmfs.noaa.gov/Assets/commercial/fus/fus15/documents/FUS2015.pdf (Accessed: 4 May 2020).

National Science Foundation (2015) *Maine's Sustainability Science Initiative*. NSF EPSCoR Award Abstract # 0904155. Available at: https://www.nsf.gov/awardsearch/showAward?AWD_ID=0904155.

National Science Foundation (2016) *Collaborative Research: Strengthening the scientific basis for decision-making: Advancing sustainability science and knowledge-action capacities in coupled coastal systems*. NSF EPSCoR Award Abstract # 1330691. Available at: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1330691.

Nielsen, F. (2016) 'Hierarchical Clustering', in Nielsen, F., *Introduction to HPC with MPI for Data Science*. Cham: Springer International Publishing (Undergraduate Topics in Computer Science), pp. 195–211. Available at: https://doi.org/10.1007/978-3-319-21903-5_8.

NOAA Office for Coastal Management (no date) *The National Coastal Zone Management Program*. Available at: https://coast.noaa.gov/czm/ (Accessed: 1 July 2023).

O'Callaghan, J.F. and Mark, D.M. (1984) 'The Extraction of Drainage Networks from Digital Elevation Data', *Computer Vision, Graphics, and Image Processing*, 28(3), pp. 323–344. Available at: https://doi.org/10.1016/S0734-189X(84)80011-0.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem*. Norwich [Norfolk]: Geo Books (Concepts and techniques in modern geography, No. 38).

PCD Live Briefs (2023) *30K gallons: Hog-waste spill prompts state investigation, Port City Daily*. Available at: https://portcitydaily.com/local-news/2023/01/29/30k-gallons-hog-waste-spill-prompts-state-investigation/ (Accessed: 24 July 2023).

Peckham, S.D. (1998) 'Efficient Extraction of River Networks and Hydrologic Measurements from Digital Elevation Data', in Barndorff-Nielsen, O. E. et al., *Stochastic Methods In Hydrology*. WORLD SCIENTIFIC (Advanced Series on Statistical Science and Applied Probability), pp. 173–203. Available at: https://doi.org/10.1142/9789812839725_0006.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine Learning in Python', *J. Mach. Learn. Res.*, 12(null), pp. 2825–2830.

Polsby, D. and Popper, R. (1991) 'The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering', *Yale Law & Policy Review*, 9. Available at: http://hdl.handle.net/20.500.13051/17448.

Poppenga, S.K. *et al.* (2014) 'Hydrologic enforcement of lidar DEMs: U.S. Geological Survey Fact Sheet 2014–3051'. USGS (Fact Sheet). Available at: https://dx.doi.org/10.3133/fs20143051.

Poppenga, S.K., Gesch, D.B. and Worstell, B.B. (2013) 'Hydrography Change Detection: The Usefulness of Surface Channels Derived From LiDAR DEMs for Updating Mapped Hydrography1', *JAWRA Journal* *of the American Water Resources Association*, 49(2), pp. 371–389. Available at: https://doi.org/10.1111/jawr.12027.

Preisendorfer, R.W., Zwiers, F.W. and Barnett, T.P. (1981) *Foundations of principal component selection rules*. University of Calif. (SIO reference series). Available at: https://books.google.com/books?id=2gnsHAAACAAJ.

Python Software Foundation (2023) *pickle — Python object serialization*, *Python documentation*. Available at: https://docs.python.org/3/library/pickle.html (Accessed: 14 July 2023).

Ralston, D.K. and Stacey, M.T. (2007) 'Tidal and meteorological forcing of sediment transport in tributary mudflat channels', *Continental Shelf Research*, p. 18.

Raquet, M., Williams, M. and Kucken, D. (2008) 'Bacteria and Water Quality Impacts', in *Supplemental Guide to North Carolina's Basinwide Planning*. Second Revision. Raleigh, NC: North Carolina Department of Environmental Quality.

Reay, W.G. (2004) 'Septic Tank Impacts on Ground Water Quality and Nearshore Sediment Nutrient Flux', *Ground Water*, 42(7), pp. 1079–1089. Available at: https://doi.org/10.1111/j.1745-6584.2004.tb02645.x.

Riley, M.S. (2022) 'Maine Combined Sewer Overflow 2021 Status Report'. Maine Department of Environmental Protection.

Ross, L. *et al.* (2021) 'Tidal and subtidal transport in short, tidally-driven estuaries with low rates of freshwater input', *Continental Shelf Research*, 224, p. 104453. Available at: https://doi.org/10.1016/j.csr.2021.104453.

Ross, L. and Smith, S. (2020) 'The Land-Sea Conjunction Junction…what's the function? Connecting coastal places, people, and science'. *Sen. George J. Mitchell Center Sustainability Talks*, Orono, ME, 19 October.

Rothenheber, D. and Jones, S. (2018) 'Enterococcal Concentrations in a Coastal Ecosystem Are a Function of Fecal Source Input, Environmental Conditions, and Environmental Sources', *Applied and Environmental Microbiology*. Edited by E.G. Dudley, 84(17), pp. e01038-18. Available at: https://doi.org/10.1128/AEM.01038-18.

Roy, S. *et al.* (2016) 'A runoff-based vulnerability analysis to examine and communicate bacteria pollution dynamics along the Gulf of Maine coast'. *Geological Society of America Annual Meeting*, Denver, CO.

Roy, S.G. *et al.* (2018) 'Resilient Coastal Communities: using interdisciplinary research and stakeholder engagement to face coastal pollution challenges in New England'. *AGU Fall Meeting 2018*, Washington, D.C., 14 December. Available at: https://agu.confex.com/agu/fm18/meetingapp.cgi/Paper/439145.

Roydhouse, J.K. *et al.* (2022) 'The use of proxies and proxy-reported measures: a report of the international society for quality of life research (ISOQOL) proxy task force', *Quality of Life Research*, 31(2), pp. 317–327. Available at: https://doi.org/10.1007/s11136-021-02937-8.

Royer-Gaspard, P., Andréassian, V. and Thirel, G. (2021) ‘Technical note: PMR – a proxy metric to assess hydrological model robustness in a changing climate’, *Hydrology and Earth System Sciences*, 25(11), pp. 5703–5716. Available at: https://doi.org/10.5194/hess-25-5703-2021.

Schilling, O.S. *et al.* (2014) ‘Using tree ring data as a proxy for transpiration to reduce predictive uncertainty of a model simulating groundwater–surface water–vegetation interactions’, *Journal of Hydrology*, 519, pp. 2258–2271. Available at: https://doi.org/10.1016/j.jhydrol.2014.08.063.

Schnitker, D. (1974) ‘Postglacial Emergence of the Gulf of Maine’, *Geological Society of America Bulletin*, 85, pp. 491–494.

Schulz, K. *et al.* (2020) ‘The Fate of Mud Nourishment in Response to Short-Term Wind Forcing’, *Estuaries and Coasts* [Preprint]. Available at: https://doi.org/10.1007/s12237-020-00767-4.

Schwarz, G. (1978) ‘Estimating the Dimension of a Model’, *The Annals of statistics*, 6(2), pp. 461–464. Available at: https://doi.org/10.1214/aos/1176344136.

scikit-learn developers (2023a) *1.1. Linear Models, scikit-learn User Guide*. Available at: https://scikit-learn/stable/modules/linear_model.html (Accessed: 21 July 2023).

scikit-learn developers (2023b) *1.10. Decision Trees, scikit-learn User Guide*. Available at: https://scikit-learn/stable/modules/tree.html (Accessed: 21 July 2023).

scikit-learn developers (2023c) *Model persistence, scikit-learn User Guide*. Available at: https://scikit-learn/stable/model_persistence.html (Accessed: 16 July 2023).

scikit-learn developers (2023d) *sklearn.decomposition.PCA, scikit-learn User Guide*. Available at: https://scikit-learn/stable/modules/generated/sklearn.decomposition.PCA.html (Accessed: 16 July 2023).

scikit-learn developers (2023e) *sklearn.mixture.GaussianMixture, scikit-learn User Guide*. Available at: https://scikit-learn/stable/modules/generated/sklearn.mixture.GaussianMixture.html (Accessed: 16 July 2023).

Seaber, P.R., Kapinos, F.P. and Knapp, G.L. (1987) *Hydrologic Unit Maps*. Water Supply Paper 2294. Denver, CO: US Geological Survey. Available at: https://pubs.usgs.gov/wsp/wsp2294/pdf/wsp_2294.pdf (Accessed: 9 August 2023).

Shahidul Islam, Md. and Tanaka, M. (2004) ‘Impacts of pollution on coastal and marine ecosystems including coastal and marine fisheries and approach for management: a review and synthesis’, *Marine Pollution Bulletin*, 48(7–8), pp. 624–649. Available at: https://doi.org/10.1016/j.marpolbul.2003.12.004.

Shu-Hsien Liao (2005) ‘Expert system methodologies and applications—a decade review from 1995 to 2004’, *Expert Systems with Applications*, 28(1), pp. 93–103. Available at: https://doi.org/10.1016/j.eswa.2004.08.003.

Simenstad, C.A. *et al.* (2011) *Columbia River Estuary Ecosystem Classification — Concept and Application*. U.S. Geological Survey Open-File Report 2011–1228.

Sims, M. and Kaczor, K. (2017) *Summary Report of Enhanced Monitoring and Pollution Source Tracking Efforts in Goosefare Brook, Maine, 2012-2016*. 138, p. 40.

Smeti, E.M. *et al.* (2009) 'Treated water quality assurance and description of distribution networks by multivariate chemometrics', *Water Research*, 43(18), pp. 4676–4684. Available at: https://doi.org/10.1016/j.watres.2009.07.023.

Smith, B. and Mark, D.M. (2003) 'Do Mountains Exist? Towards an Ontology of Landforms', *Environment and Planning B: Planning and Design*, 30(3), pp. 411–427. Available at: https://doi.org/10.1068/b12821.

Smith, R. *et al.* (2022) 'Decision Science Can Help Address the Challenges of Long-Term Planning in the Colorado River Basin', *JAWRA Journal of the American Water Resources Association*, 58(5), pp. 735–745. Available at: https://doi.org/10.1111/1752-1688.12985.

Smith, S. *et al.* (2022) 'Land-sea Connections of Coastal Places, People, and Science in Maine'. *Maine Sustainability and Water Conference*, Augusta, ME.

Smith, S.M. *et al.* (2015) 'Downeast Drainage - Examining and Communicating the Dynamics of Bacteria Pollution Events in the Gulf of Maine (EP21C-0927)'. *AGU Fall Meeting 2015*, San Francisco, CA, 15 December. Available at: https://agu.confex.com/agu/fm15/meetingapp.cgi/Paper/86521.

Smith, S.M. *et al.* (2016) 'Downeast Drainage: Examining and Communicating Pollution Problems Along the Coast of Maine'. *Shellfish Advisory Council Meeting*, Ellsworth, Maine, 29 November.

Smith, S.M. *et al.* (2019) 'Connecting and Projecting the Dynamics of Coastal Bacteria Pollution Problems', in *AGU Fall Meeting 2019*. AGU.

Smith, S.M.C. (2016) 'INVITED - Keynote Presentation - Downeast drainage: sources, delivery and residence time of runoff and pollutants along the Gulf of Maine coast'. *Geological Society of Maine Fall Meeting*, Augusta, ME, 21 November.

Sowah, R. *et al.* (2014) 'Evaluating the influence of septic systems and watershed characteristics on stream faecal pollution in suburban watersheds in Georgia, USA', *Journal of Applied Microbiology*, 117(5), pp. 1500–1512. Available at: https://doi.org/10.1111/jam.12614.

Staley, Z.R. and Edge, T.A. (2016) 'Comparative microbial source tracking methods for identification of fecal contamination sources at Sunnyside Beach in the Toronto region area of concern', *Journal of Water and Health*, 14(5), pp. 839–850. Available at: https://doi.org/10.2166/wh.2016.296.

Stirpe, C.R. *et al.* (2021) 'The Mg/Ca proxy for temperature: A Uvigerina core-top study in the Southwest Pacific', *Geochimica et Cosmochimica Acta*, 309, pp. 299–312. Available at: https://doi.org/10.1016/j.gca.2021.06.015.

Studds, C.E. *et al.* (2012) 'Land Cover and Rainfall Interact to Shape Waterbird Community Composition', *PLoS ONE*. Edited by S. Thrush, 7(4), p. e35969. Available at: https://doi.org/10.1371/journal.pone.0035969.

Sutton, P.C. *et al.* (2009) ‘Paving the planet: impervious surface as proxy measure of the human ecological footprint’, *Progress in Physical Geography: Earth and Environment*, 33(4), pp. 510–527. Available at: https://doi.org/10.1177/0309133309346649.

Sutton, S., Sczesny, M. and Cone, A. (2023) *Spill stopped after millions of gallons of sewage went into Intracoastal in Boynton Beach, WPTV News Channel 5 West Palm*. Available at: https://www.wptv.com/news/region-s-palm-beach-county/boynton-beach/leak-causes-million-of-gallons-of-sewage-to-spill-into-intracoastal-waterway-in-boynton-beach (Accessed: 24 July 2023).

Tanner, B.R., Perfect, E. and Kelley, J.T. (2006) ‘Fractal Analysis of Maine’s Glaciated Shoreline Tests Established Coastal Classification Scheme’, *Journal of Coastal Research*, 225, pp. 1300–1304. Available at: https://doi.org/10.2112/05-0474R.1.

Tarboton, D.G. (1997) ‘A new method for the determination of flow directions and upslope areas in grid digital elevation models’, *Water Resources Research*, 33(2), pp. 309–319. Available at: https://doi.org/10.1029/96WR03137.

Taylor, P.H. (2018a) *Safe Beaches & Shellfish Project Decision Support Toolkit (DST) Development Plan*. Report for the New England Sustainability Consortium. Waterview Consulting, p. 35.

Taylor, P.H. (2018b) *Web Content for the Safe Beaches & Shellfish Project Decision Support Toolkit (DST)*. Report for the New England Sustainability Consortium. Waterview Consulting, p. 13.

Thomas, A.C. *et al.* (2010) ‘Interannual variability of shellfish toxicity in the Gulf of Maine: Time and space patterns and links to environmental variability’, *Harmful Algae*, 9(5), pp. 458–480. Available at: https://doi.org/10.1016/j.hal.2010.03.002.

Turner, B.L. *et al.* (2003) ‘A framework for vulnerability analysis in sustainability science’, *Proceedings of the National Academy of Sciences*, 100(14), pp. 8074–8079. Available at: https://doi.org/10.1073/pnas.1231335100.

Tyre, K.N. *et al.* (2023) ‘Widespread human waste pollution in surface waters observed throughout the urbanized, coastal communities of Lee County, Florida, USA’, *Science of The Total Environment*, 879, p. 162716. Available at: https://doi.org/10.1016/j.scitotenv.2023.162716.

Uchupi, E. and Bolmer, S.T. (2008) ‘Geologic evolution of the Gulf of Maine region’, *Earth-Science Reviews*, 91(1–4), pp. 27–76. Available at: https://doi.org/10.1016/j.earscirev.2008.09.002.

US EPA, O. (2013) *Summary of the Clean Water Act*. Available at: https://www.epa.gov/laws-regulations/summary-clean-water-act (Accessed: 3 July 2023).

US Food and Drug Administration (2019) *National Shellfish Sanitation Program (NSSP) Guide for the Control of Molluscan Shellfish: 2019 Revision*, p. 502. Available at: http://www.fda.gov/Food/GuidanceRegulation/FederalStateFoodPrograms/ucm2006754.htm.

Van Dam, B., Smith, S. and Beard, K. (2019) ‘Thinking at the Margin: Examining the role of estuary margin watersheds in coastal bacteria pollution vulnerability’. *Maine Sustainability & Water Conference*, Augusta, ME, 28 March.

Van Dam, B.E. (2022) *ArcMap Hydrology Toolset*, p. 17. Available at: digitalcommons.library.umaine.edu/student_work/17.

Van Dam, B.E. and Smith, S.M.C. (2023) 'Remote Detection of Upland Surface Water Storage Capacity in Deglaciated Terrain', *Manuscript submitted for publication.* [Preprint].

Van Den Berg, A.E., Jorgensen, A. and Wilson, E.R. (2014) 'Evaluating restoration in urban green spaces: Does setting type make a difference?', *Landscape and Urban Planning*, 127, pp. 173–181. Available at: https://doi.org/10.1016/j.landurbplan.2014.04.012.

Van Niekerk, L. *et al.* (2020) 'An Estuary Ecosystem Classification that encompasses biogeography and a high diversity of types in support of protection and management', *African Journal of Aquatic Science*, 45(1–2), pp. 199–216. Available at: https://doi.org/10.2989/16085914.2019.1685934.

Vanderplas, J.T. (2016) *Python data science handbook: essential tools for working with data*. First edition. Sebastopol, CA: O'Reilly Media, Inc.

Walsh, M. (2022) 'Chester County HazMat team tackles raw sewage leaking from 3,500-gallon tanker into creek', *Daily Local*, 15 March. Available at: https://www.dailylocal.com/2022/03/15/chester-county-hazmat-team-cleans-large-sewage-leak-from-tanker-into-creek/ (Accessed: 24 July 2023).

Wan, Y. *et al.* (2013) 'Modeling residence time with a three-dimensional hydrodynamic model: Linkage with chlorophyll a in a subtropical estuary', *Ecological Modelling*, 268, pp. 93–102. Available at: https://doi.org/10.1016/j.ecolmodel.2013.08.008.

Weaver, C.P. and Miller, C.A. (2019) 'A Framework for Climate Change-Related Research to Inform Environmental Protection', *Environmental Management*, 64(3), pp. 245–257. Available at: https://doi.org/10.1007/s00267-019-01189-0.

Wei Luo *et al.* (2009) 'Web Service for extracting Terrain Openness from DEM Data', in. IEEE, pp. 1–5. Available at: https://doi.org/10.1109/GEOINFORMATICS.2009.5293470.

Withers, P.J. *et al.* (2014) 'Do septic tank systems pose a hidden threat to water quality?', *Frontiers in Ecology and the Environment*, 12(2), pp. 123–130. Available at: https://doi.org/10.1890/130131.

Wood, G. (2021) 'Status of Licensed Discharges'. Maine Department of Environmental Protection. Available at: https://www.maine.gov/dep/publications/reports/index.html.

Woods, M.N. *et al.* (2018) 'Microplastic fiber uptake, ingestion, and egestion rates in the blue mussel (Mytilus edulis)', *Marine Pollution Bulletin*, 137, pp. 638–645. Available at: https://doi.org/10.1016/j.marpolbul.2018.10.061.

Yahara, T. (ed.) (2021) *Decision Science for Future Earth: Theory and Practice*. Singapore: Springer Singapore. Available at: https://doi.org/10.1007/978-981-15-8632-3.

Yan, X. and Su, X. (2009) *Linear regression analysis: theory and computing*. Singapore ; Hackensack, NJ: World Scientific.

Zhang, W. and Montgomery, D.R. (1994) ‘Digital elevation model grid size, landscape representation, and hydrologic simulations’, *Water Resources Research*, 30(4), pp. 1019–1028. Available at: https://doi.org/10.1029/93WR03553.

Zhao, S. *et al.* (2018) ‘Field-Based Evidence for Microplastic in Marine Aggregates and Mussels: Implications for Trophic Transfer’, *Environmental Science & Technology*, 52(19), pp. 11038–11048. Available at: https://doi.org/10.1021/acs.est.8b03467.

APPENDIX A

TECHNICAL MANUAL: HYDROLOGY GEOPROCESSING WORKFLOWS

The following pages are a reproduction of the *"WPES, How Do I…?" Quick Guide to ArcMap Hydrology Toolset* (Van Dam, 2022). This technical document was developed to fill a literature gap in described workflows and established methodology for performing geospatial operations important to accomplishing the research work described in this dissertation, particularly related to protocols for the identification of road-crossing culvert locations.

WPES, How Do I...?

Quick Guide to

ArcMap Hydrology Toolset

Prepared by Bea E. Van Dam

for

Watershed Process and Estuary Sustainability Research Group
University of Maine, Orono, ME
umaine.edu/watershedresearch

The "WPES, How Do I...?" series of instruction manuals were originally developed for internal use by the Watershed Process and Estuary Sustainability Research Group (WPES) at the University of Maine. They have been updated and published through DigitalCommons@UMaine to provide open access for academic use.

Funding support for production of this publication has been provided by the Maine Agricultural and Forest Experiment Station (MAFES) Project No. ME0ME022209 and by the Maine Water Resources Research Institute (WRRI) Project No. 2018ME331B and FY21 Project No. 5520980-5409118-20.

This document is formatted for print in booklet form on letter size paper.

Table of Contents

Introduction

This document provides guidance for performing hydrological tasks using the Hydrology Toolset in ArcMap 10.x (ArcGIS Desktop) and assumes general familiarity with the program. For general ArcMap guidance, refer to "WPES, How Do I...?" Quick Guide to ArcMap Basics.

Hydrology Toolset

The Hydrology Toolset in ArcMap contains tools for performing hydrologic tasks and analyses related to the movement of surface water runoff across a digital elevation model (DEM). These tools are found in the Spatial Analyst toolbox and require an active Spatial Analyst license. In-depth documentation from Esri for the Hydrology Toolset is also available at https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/an-overview-of-the-hydrology-tools.htm.

Basin

Uses flow direction raster to delineate all basins draining to the edge of a DEM. Does not allow for the specification of basin outlet locations (pour points). In most cases using the Watershed tool instead is preferable.

Fill

Eliminates all sinks (pits) from a DEM, producing a raster in which surface runoff from any cell can reach the edge of the DEM uninterrupted. This process is also known as hydro-conditioning. Max fill depth can be specified to avoid the elimination of actual closed (endorheic) basins. This is generally the first tool used in any hydrology workflow.

Flow Accumulation

Uses flow direction raster to calculate number of upstream cells contributing surface flow to each DEM cell (note that to calculate contributing area you must then multiply output values by DEM cell area). Method must match that used for the creation of the flow direction raster. Flow accumulation with DINF method takes considerably longer (factor of 10 or more) than with D8 method for substantially the same results.

Flow Direction

Calculates direction of surface runoff for each cell in a DEM based on slope to adjacent cells. Default type/method is D8 (all flow from a cell is routed in one of eight directions into single adjacent cell); this is the only method that will work when working with a DEM with “float” data type, which most LiDAR DEMs are.

Flow Distance

Uses DEM and stream channel rasters to calculate distance of overland flow from DEM cell into channel network. May be set to calculate either vertical drop or horizontal distance. Output raster cell values are lengths in map units.

Flow Length

Uses flow direction raster to calculate either downstream flow length from cell to edge of raster or longest upstream flow length from cell to watershed boundary. Output raster cell values are lengths in map units.

Sink

Uses flow direction raster to delineate raster layer of all sinks (pits) in DEM. Contiguous patches of pit cells are consolidated (e.g., a pond whose water surface in the DEM is below its outlet elevation would be considered a single sink). Sinks are numbered but no further information (such as depth) is calculated. In cases where depth/volume is important, using the pit fill detection workflow is preferable.

Snap Pour Point

Converts feature basin outlets (aka pour points) to raster to be used by Watershed tool. Each outlet should have a unique numeric identifier set in its attribute table before running this tool.

Stream Link

Converts stream network raster (where all stream cells are indicated with an integer value) to raster where each stream segment (reach) is assigned a unique identifier.

Stream Order

Uses stream network raster and flow direction raster to produce raster of stream order (Strahler or Shreve) for each stream segment.

Stream to Feature

Converts stream network raster to polylines with splits at each confluence (i.e., each reach is its own line segment). Ensure non-stream cells of raster are "Null" rather than "0" - the tool will otherwise create polyline features for all of these cells.

Watershed

Uses flow direction raster and snapped user-defined pour point(s) to produce raster of delineated watershed(s). Each pour point should have a unique numeric value, otherwise tool will combine resulting watersheds into single unit (this behavior may be desirable in uncommon cases where one basin does have two outlets, such as an estuary mouth that is split by an island).

Additional Tools in Workflows

For readers of print version of manual, tool documentation from Esri (blue links) can be found in the "Tool Reference" section at https://desktop.arcgis.com/en/arcmap/latest/tools.

Con (conditional)

Spatial Analyst Tools > Conditional
The Con tool evaluates a specified condition on a raster dataset. In the context of hydrology workflows, it is used to identify stream networks by evaluating whether flow accumulation is greater than a specified threshold, generally corresponding to an assumed area or length to stream initiation. This tool requires an active Spatial Analyst license.

Hillshade

Spatial Analyst Tools > Surface -or- *3D Analyst Tools > Raster Surface*
The Hillshade tool creates a shaded relief raster from a DEM, giving the appearance of a 3D map and making it easier to interpret the terrain and visually locate features such as stream channels. It is often desirable to give some transparency (15-25%) to resulting hillshade layer in Table of Contents in order to see colormap from underlying DEM layer. This tool requires an active Spatial Analyst or 3D Analyst license.

Mosaic to New Raster

Data Management Tools > Raster > Raster Dataset
The Mosaic to New Raster tool is used to merge multiple rasters to create a single new raster. A common use case for this tool is combining DEM tiles into a single file that covers a full area of interest.
The Mosaic tool in the same toolset can be used to merge multiple rasters into an existing raster; this saves storage space, but does not preserve a copy of the target raster original data.

Raster Conversion toolsets

Conversion Tools > From Raster
Convert rasters to features, with option to create simplified features. Rasters must be of integer data type.
Raster to Polygon – creates polygon for each raster zone (integer value). Used to convert watershed raster to watershed polygons.

Raster to Polyline – creates polyline for each raster zone (integer value). Used to convert stream network rasters to stream polylines.
Conversion Tools > To Raster
Polyline to Raster – creates raster from polyline features. Resulting raster cell value may be from any attribute column of the input features. Cell size can be directly set as a numeric value, or user may browse to an existing raster layer to select that raster's cell size. Processing extent settings (under "Environments…") may be updated to snap (align) raster cells to those of an existing raster.

Raster Math toolset

Spatial Analyst Tools > Math -or- *3D Analyst Tools > Raster Math*
Perform mathematical operations on raster datasets. These tools require an active Spatial Analyst or 3D Analyst license.
Int – converts raster data type to integer by truncating decimal (i.e., rounding down to nearest whole number).
Minus – subtract spatially coincident cell values of one raster from another; can also subtract specified numeric value from all cells.
Times – multiply raster cell values by those of another raster or a specified numeric value.

Zonal Statistics

Spatial Analyst Tools > Zonal
The Zonal Statistics tool is used to calculate a statistic (e.g. minimum elevation) from a raster within raster or feature zones. Output cells in the resulting raster dataset will reflect the statistic values for each overlaid feature, with null values for all other cells. Tool will fail if run while an editing session is active.
The Zonal Statistics as Table tool similarly calculates one or many statistics from a raster within zones, creating output in tabular form; the table may then be joined to a feature or raster attribute table.
Both of these tools require an active Spatial Analyst license.
An analogous tool in the 3D Analyst toolbox is Add Surface Information (*3D Analyst Tools > Functional Surface*), which adds a column in the feature's attribute table for each of multiple selected statistics; the features may then be converted to a raster separately. This tool requires an active 3D Analyst license.

Workflows

Tool names (referenced above) are indicated with **bold** text. Data layer names are indicated in *italics*.

Pit fill detection

1. Begin with bare earth DEM (*DEM*) and hillshade (*HS*) raster for area of interest.
2. **Fill** DEM.

   Input surface raster: *DEM*

   Output raster: *DEM_fill*

   Note: Unless working with a known closed basin, "Z limit" field should be left blank.
3. **Minus** - subtract original DEM elevations from filled DEM.

   Input raster 1: *DEM_fill*

   Input raster 2: *DEM*

   Output raster: *DEM_pits*

   Note: Output raster will have positive fill depth in pit cells and zero values in all other cells.
4. Examine resulting fill depth raster (*DEM_pits*) for presence of artificial fills.

   Note: Artificial fills occur where structures in the DEM (often roads) obscure real flow paths that pass below them (usually via culverts, as bridge structures spanning larger streams and rivers are generally removed during limited hydro-performed before public release of the bare earth DEM). The Fill tool then raises the surface level of the "sink" on the upslope side until it overtops the structure or finds another outlet (often along roadside drainage ditches).

   4.1. Add *DEM_pits* to map above *HS* and set Layer Properties.

      4.1.1. Check "Display Background Value: 0" as null color (Symbology tab).

      4.1.2. (Optional) Change color ramp for increased visibility (Symbology tab).

      4.1.3. (Optional) Set layer transparency to 50% (Display tab).

   4.2. Check for large or deep fill areas.

Note: These most noticeably occur where roads cut perpendicularly across valleys with relatively shallow bottom slopes, creating elongated fills reminiscent of reservoirs behind large river dams.

4.3. Scan along roads for intersections with channel or swale features visible in hillshade layer.

Note: Artificial pit fill features here may be small but should nonetheless be cleared with appropriate culverts. This is particularly important when operating as part of a larger workflow related to surface flow routing. It can be helpful to create ***flow accumulation*** *raster for filled DEM to help identify potential intersections.*

4.3.1. Where road is close to perpendicular to hillslope and channel feature continues on downslope side of road, this almost always indicates the presence of a road-crossing culvert.

4.3.2. Where road is close to parallel to hillslope and tail of fill points up road, this often indicates the presence of a driveway-crossing culvert along roadside.

4.4. Where possible, verify with Google Street View or similar.

Note: Even when stream or pipe/box are not directly visible (e.g., in thick roadside vegetation), common telltale signs of road-crossing culvert presence include raised line of cracks perpendicular to road surface where road base has settled on either side of culvert (alternately, dip where culvert pipe has partially collapsed), or paired drainage grates at nadir of dip in road.

5. Burn in culverts as necessary to remove artificial fills.

5.1. See hydro-enforcement workflow steps 4 – 6.

6. Repeat pit fill detection steps 2 – 5 as necessary until all artificial fills have been identified and cleared and only natural pit features remain.

Note: ***Fill*** *and* ***Minus*** *will be performed on each new iteration of hydro-enforced DEM (*hydroDEM*), but culvert burning should always be performed on original bare earth DEM (*DEM*).*

7. Final *DEM_pits* raster can be used to investigate statistics related to surface water detention storage:

7.1. Multiply raster value (fill depth) by cell area to calculate potential storage volume for each cell.

7.2. Overlay feature or raster layer (e.g., with **Zonal Statistics as Table** tool) to calculate storage statistics within zones of interest, such as by land cover or soil type.

Hydro-enforcement ("Burning in" culverts)

1. Begin with bare earth DEM (*DEM*) and hillshade (*HS*) raster for area of interest.
2. **If you already have line layer of delineated culverts with associated elevations**:
   2.1. Use **Polyline to Raster** to convert culverts to raster (*cv_raster*) with inlet elevation as the value field and the same cell size as *DEM*.
   2.2. Skip to step 6.

   If you have line layer of delineated culverts without elevations: Skip to step 5.

   If you have point layer showing culvert locations: skip to step 4.

   If you have none of these: continue to step 3.
3. Identify culvert locations.
   3.1. See pit fill detection workflow steps 2 – 4 for using presence of artificial fills to indicate likely culverts.
4. Create polyline feature for each identified culvert.
   4.1. Create new polyline shapefile or feature dataset (*Culverts*) in ArcCatalog.
   4.2. Make *Culverts* dataset editable.
   4.3. At each culvert location, draw polyline crossing artificial impediment.
   4.4. Save edits and stop editing session.
5. Use **Zonal Statistics** to create new raster with culvert cell elevations equal to minimum DEM elevations along culvert lines.

   Input zone data: *Culverts*

   Zone field: OBJECTID or other field with unique identifier for each culvert

   Input value raster: *DEM*

   Output raster: *cv_raster*

   Statistics type: MINIMUM
6. Use **Mosaic to New Raster** to create hydro-enforced raster with culverts burned in.

Input Rasters: *DEM; cv_raster*

Note: Rasters should be in this order; if swapped, Mosaic Operator field must be correspondingly changed.

Raster Dataset Name: *hydroDEM*

Pixel Type: 32_BIT_FLOAT

Number of Bands: 1

Mosaic Operator: LAST

Watershed delineation

1. Begin with hydro-enforced bare earth DEM (*hydroDEM*) and hillshade (*HS*) raster for area of interest.
   1.1. See hydro-enforcement workflow.
2. **Fill** DEM.

   Input surface raster: *hydroDEM*

   Output raster: *DEM_fill*

   Note: Unless working with a known closed basin, "Z limit" field should be left blank.
3. Run **Flow Direction** tool on DEM.

   Input surface raster: *DEM_fill*

   Output flow direction raster: *Flow_Dir*

   Flow direction type: D8
4. Run **Flow Accumulation** tool on flow direction raster.

   Input flow direction raster: *Flow_Dir*

   Output accumulation raster: *Flow_Acc*

   Flow direction type: D8

   4.1. Add *Flow_Acc* to map above *HS* and set Layer Properties.

   4.1.1. Set layer transparency to 75% (Display tab).

   4.1.2. (Optional) Increase flow line visibility by setting color ramp stretch type to "Minimum-Maximum" and lowering "High" value (Symbology tab).
5. Create watershed outlet feature(s).

   Note: These are most commonly point features, as flow accumulation tends to collapse to single-pixel wide lines. However, wide channels (such as at mouths of larger rivers and estuaries) often feature multiple parallel flow accumulation *lines moving downstream. In these cases, it is helpful to instead create an outlet line feature across the mouth.*

5.1. Create new shapefile or feature dataset (*Outlets*) in ArcCatalog.

5.2. Make *Outlets* dataset editable.

5.3. At each desired watershed outlet location, place point directly on flow accumulation line pixel.

Note: For outlet line features, outlet line should stretch from bank to bank and cross all flow accumulation lines.

5.4. Ensure each watershed outlet has a unique identifier in feature attribute table.

Note: In most cases, the OBJECTID field is sufficient. In rare cases where multiple outlet features correspond to a single watershed (e.g., two points or line segments at the mouth of a river split by an island in the middle), all features for that watershed must have the same identifier. Features may either be merged (available only for lines, not points), or a new attribute column may be added and populated accordingly.

5.5. Save edits and stop editing session.

6. **Snap Pour Point**(s) to outlet raster.

   Input feature pour point data: *Outlets*

   Pour point field: OBJECTID or other field with unique watershed identifier

   Input accumulation raster: *Flow_Acc*

   Output raster: *pp_raster*

7. Run **Watershed** tool.

   Input flow direction raster: *Flow_Dir*

   Input pour point data: *pp_raster*

   Note: Using pour point features rather than snapped pour point raster often produces anomalous results and should be avoided.

   Pour point field: Value

   Output raster: *wshed_raster*

8. Run **Raster to Polygon** tool to convert watershed raster to features.

   Input raster: *wshed_raster*

   Field: Value

   Output polygon features: *wshed_poly*

Note: Simplifying polygons (the default option) is generally recommended, with a caveat – any adjacent watershed polygons from the same wshed_raster *layer will be simplified in such a way that their shared boundaries will match node-for-node, but if another adjacent watershed is subsequently delineated and simplified in a separate iteration of this workflow, the resulting polygon may have small boundary mismatches. If delineating adjacent watersheds in multiple iterations, polygons should not be simplified during raster conversion process.*

Stream network delineation

1. Begin with hydro-enforced bare earth DEM (*hydroDEM*) raster for area of interest.
   1.1. See hydro-enforcement workflow.
2. **Fill** DEM.

   Input surface raster: *hydroDEM*

   Output raster: *DEM_fill*

   Note: Unless working with a known closed basin, "Z limit" field should be left blank.
3. Run **Flow Direction** tool on DEM.

   Input surface raster: *DEM_fill*

   Output flow direction raster: *Flow_Dir*

   Flow direction type: D8
4. **For channel initiation based on source area**:
   4.1. Run **Flow Accumulation** tool on flow direction raster.

   Input flow direction raster: *Flow_Dir*

   Output accumulation raster: *Flow_Acc*

   Flow direction type: D8

   4.2. Skip to step 5.

   For channel initiation based on overland flow length:

   4.3. Run **Flow Length** tool on flow direction raster.

   Input flow direction raster: *Flow_Dir*

   Output raster: *Flow_Len*

   Direction of measurement: UPSTREAM

   4.4. Continue to step 5.

5. Run **Con** tool on flow raster.

   Input conditional raster: [Flow raster created in step 4]

   Expression: VALUE >= [Desired channel initiation condition]

   Note: Replace "[Desired channel initiation condition]" with numeric value (no brackets). For Flow_Acc *raster, this value will be the number of raster cells corresponding to the area to channel initiation – for example, if area to initiation were assumed to be 0.1 km² (100,000 m²) and you are working with a 2 m resolution DEM (cell area 4 m²), the initiation condition would be 25,000 cells and the Expression field would be entered as VALUE >= 25000. For* Flow_Len *raster, value will be the overland flow length in map units – for example, if length to channel initiation is assumed to be 1 km and your map units are meters, the initiation condition would be 1000 m and the Expression field would be entered as VALUE >= 1000.*

   Input true constant value: 1

   Input false constant value: [Leave blank]

   Output raster: *Stream_Net*

6. (Optional) Use **Stream to Feature** tool to create stream polyline features from stream network raster.

Overland flow distance calculation

1. Begin with delineated stream network raster (*Stream_Net*), hydro-enforced bare earth DEM (*hydroDEM*), and flow direction raster (*Flow_Dir*).

   1.1. See stream network delineation workflow.

2. Run **Flow Distance** tool.

   Input stream raster: *Stream_Net*

   Input surface raster: *hydroDEM*

   Output raster: *Flow_Dist*

   Input flow direction raster: *Flow_Dir*

   Note: Use of flow direction raster is optional but does speed up processing time.

   Distance type: HORIZONTAL

   Statistics Type: MINIMUM

3. (Optional) Calculate average overland flow distance for catchment(s).

3.1. Begin with catchment raster (*wshed_raster)* or polygon features (*wshed_poly*).

3.1.1. See watershed delineation workflow.

3.2. Run **Zonal Statistics** tool on flow distance raster.

Input zone data: *wshed_raster* or *wshed_poly*

Zone field: [Unique watershed ID column]

Input value raster: *Flow_Dist*

Statistics type: MEAN

*Note: To simultaneously calculate other statistics such as maximum and standard deviation, instead use **Zonal Statistics as Table** tool.*

DINF flow routing on floating point raster

*Note: **NOT RECOMMENDED** - Flow Accumulation tool on DINF raster can split flow and produce output values that do not monotonically increase when moving downstream, resulting in complications such as loops and gaps when delineating stream network from flow accumulation raster.*

D-Infinity flow routing methods require an integer data type raster, while most high-resolution DEMs are floating point. This workflow provides a workaround without sacrificing needed vertical resolution.

1. Begin with floating point type DEM raster.
2. **Times** – multiply DEM values by a power of ten to retain desired level of precision.

   Note: Converting directly to integer type without multiplication results in elevations being truncated to the previous full meter, removing differentiation between cells and resulting in an unnatural step effect in the DEM. Multiplying by 10^1 retains original DEM precision to the decimeter level, 10^2 to the centimeter level, and so on.
3. **Int** – convert resulting raster to integer data type.
4. Integer type raster can then be used for DINF routing methods in **Flow Direction** and **Flow Accumulation** tools.

   *Note that if using **Flow Distance** tool to calculate vertical component of overland flow distance from integer DEM and DINF flow direction raster, results will be exaggerated by same multiplicative factor used in step 2.*

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER 2

Data Sources: The Maine GeoLibrary (https://geolibrary-maine.opendata.arcgis.com) was the primary state-level spatial data source for this research. Some of these data were originally obtained through its since-deprecated predecessor spatial data clearinghouse, the Maine Office of GIS (https://www.maine.gov/megis/catalog); GeoLibrary has been indicated as the source here as long as an equivalent file is currently hosted.

Table B.1 Descriptions and sources of elevation, boundary, and hydrology data used for coastal flow path delineation.

Data Layer	Description	Data Type	Source
2 m DEM	Two-meter cell length topographic digital elevation model of Maine coast.	Raster (2 m)	Maine GeoLibrary; Todd Metzler (pers. comm., May 2019)
Shoreline Boundaries (METWP24P)	1:24,000 scale town and township boundaries for the State of Maine, corresponding to shoreline (mean high water)	Polygon	Maine GeoLibrary
Highest High Tide	Highest astronomical high tide for the state of Maine	Polyline	Maine GeoLibrary
MaineDOT Bridges, Cross Culverts, Large Culverts	Point features for the locations of three road-crossing structure types along state roadways	Point	Maine GeoLibrary
NHD Flowlines	1:24,000 scale or larger "flow network consisting predominantly of stream/river and artificial path vector features"	Polyline	USGS https://www.usgs.gov/national-hydrography

Table B.2. Bacteria source, delivery, and residence time proxy metrics selected for analysis of Coastal Pollution Response Units. *Denotes metrics considered in Smith et al. (2016) analysis of Landscape Pollution Response Units.

Pollution Culprit Category	**Coastal Process**	**Proxy Metric**	**References**
Source	Point source direct discharge of waste into coastal waters under system failures	• Overboard Discharges (OBDs) • Combined Sewer Overflows (CSOs) • Pollutant discharge elimination system outfalls (PDES-O) • Pollutant discharge elimination system facilities (PDES-F)	(Maine Department of Environmental Protection, 2018; Wood, 2021; Riley, 2022)
	Nonpoint source generation by humans, pets (year-round and seasonal populations)	• Population count, density * • Structure count, density *	(Raquet, Williams and Kucken, 2008; Jones, 2011; Sims and Kaczor, 2017)
	Nonpoint source generation related to land use practices and animal populations	• Drainage area (terrestrial) * • Fractions developed, farmed, rural * • Fraction conserved • Fraction tidal wetlands	(Jones, 2011; Studds et al., 2012; Hood et al., 2021)
Delivery	Generation of surface water runoff during precipitation events	• Soil drainage score * • Fractions well, moderately, poorly drained soils * • Soil water storage capacity * • Est. runoff volume from 2” precipitation *	(Jones, 2011; Gupta, 2017)
	Efficiency of pollutant delivery from landscape to estuarine waters	• Drainage density (geomorphically-derived) * • Drainage density (engineered) * • Fraction surface storage * • Maximum elevation • Mean, maximum slope • Fraction margins	(Bierman and Montgomery, 2020)
Residence Time	Freshwater forcing of estuary water evacuation	• Estuary area, depth, volume • Drainage area / Estuary area ratio * • Runoff volume / estuary volume ratio	(Dyer, 1973; Wan et al., 2013; Smith et al., 2019)
	Controls on estuary circulation patterns and external forcings	• Estuary circularity * • Estuary openness • Estuary outlet bearing	(Dyer, 1973; Bartley, Buddemeier and Bennett, 2001; Ralston and Stacey, 2007; Schulz et al., 2020; Ross et al., 2021; Bailey et al., 2023)

Table B.3. Descriptions and sources of Source, Delivery, and Residence Time (SDR) proxy metric data.

Data Layer	Description	Data Type	Source
CSO	Location of combined sewer outfall pipes licensed under Maine Pollution Discharge Elimination System	Point	Maine DEP; Maine GeoLibrary
OBD	Location of overboard discharge outfalls licensed with Maine DEP's Overboard Discharge Program	Point	Maine DEP https://www.maine.gov/dep/gis/datamaps/
PDES-F	Location of permitted Pollution Discharge Elimination System facilities	Point	Maine DEP; Maine GeoLibrary
PDES-O	Location of permitted Pollution Discharge Elimination System outfalls	Point	Maine DEP; Maine GeoLibrary
NLCD	National Land Cover Database thematic classification of land cover data (2016)	Raster (30 m)	USGS https://www.mrlc.gov/data/nlcd-2016-land-cover-conus
Conserved Lands	Boundaries of properties protected under conservation easements	Polygon	Maine GeoLibrary
NWI	National Wetlands Inventory area classifications	Polygon	US Fish & Wildlife Service https://www.fws.gov/program/national-wetlands-inventory
SSURGO	Soil Survey Geographic Database soils data with drainage classes	Polygon	USDA https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo
E911 Roads	Official roads layer for state of Maine	Polyline	Maine GeoLibrary
E911 Addresses	Location of all addressable structures in Maine	Point	Maine GeoLibrary
Pop10	Population by census block (2010)	Polygon	US Census Bureau
Bathymetry	One-ninth arc-second cell length bathymetric-topographic digital elevation model of Maine coast, resampled to 1/3”	Raster (1/3”)	NOAA www.ncei.noaa.gov/maps/bathymetry

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER 3

Table C.1. Descriptions of source, delivery, and residence time metrics incorporated into expert system tool for coastal setting identification.

Category	Name	Metric Description	Units	Aggregation
Identifiers	Est_ID_	Unique ID number	n/a	-
Watershed size and shape	Da_km2	Watershed (landscape) drainage area	km^2	Total sum
	Da_km2_M	*…within Margins*	km^2	Total sum
	Da_km2_HH	*…within Highest High Tide area*	km^2	Total sum
	FMargin	Fraction of total watershed that is in Margin watershed areas	decimal	Da_km2_M/Da_km2
	FHighHigh	Fraction of total watershed that is inundated by highest high tide	decimal	Da_km2_HH/ DA_km2
	FHighHi_M	*…within Margins*	decimal	Da_km2_HH/ Da_km2_M
	Qmax_m3	Estimated total runoff discharge from watershed based on 2 inches of rainfall	m^3	2” * Da_km2
	Qmin_m3	Minimum runoff discharge based on 2” of rainfall and total soil storage capacity	m^3	Qmax_m3 - SSTORAGEME
	Zmax	Maximum elevation within watershed	m	Maximum
	EstA_km2	Estuary water surface area	km^2	Calculated from estuary polygon
	EstD_m	Mean estuary water depth	m	Calculated from bathymetry raster
	EstV_m3	Estuary water volume	m^3	EstA_km2 * EstD_m
Channel Network	DD_3	Natural drainage density, channel initiation at 0.3 km^2	1/km	Area-weighted mean
	DD_3_M	*…within Margins*	1/km	Area-weighted mean
	DD_3_HH	*…within Highest High Tide area*	1/km	Area-weighted mean
	DD_2	Natural drainage density, init. at 0.2 km^2	1/km	Area-weighted mean
	DD_2_M	*…within Margins*	1/km	Area-weighted mean
	DD_2_HH	*…within Highest High Tide area*	1/km	Area-weighted mean
	DD_05	Natural drainage density, init. at 0.05 km^2	1/km	Area-weighted mean
	DD_05_M	*…within Margins*	1/km	Area-weighted mean
	DD_05_HH	*…within Highest High Tide area*	1/km	Area-weighted mean
	DD_3_Rds	Natural and engineered (based on roads) drainage dens., channel initiation at 0.3 km^2	1/km	Area-weighted mean
	DD_3_R_M	*…within Margins*	1/km	Area-weighted mean
	DD_3_R_HH	*…within Highest High Tide area*	1/km	Area-weighted mean
	DD_2_Rds	Natural and engineered (based on roads) drainage dens., channel initiation at 0.2 km^2	1/km	Area-weighted mean
	DD_2_R_M	*…within Margins*	1/km	Area-weighted mean
	DD_2_R_HH	*…within Highest High Tide area*	1/km	Area-weighted mean
	DD_05_Rds	Natural and engineered (based on roads) drainage dens., channel initiation at 0.05 km^2	1/km	Area-weighted mean
	DD_05_R_M	*…within Margins*	1/km	Area-weighted mean
	DD_05_R_HH	*…within Highest High Tide area*	1/km	Area-weighted mean

Table C.1 continued.

Category	Name	Metric Description	Units	Aggregation
Land Cover	FDEVELOPED	Fraction of developed (urbanized, impervious) land in watershed	decimal	Area-weighted mean
	FDEVELO_M	…*within Margins*	decimal	Area-weighted mean
	FDEVELO_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	FFARM	Fraction of land that is used for non-forestry agriculture in watershed (livestock, crops, etc)	decimal	Area-weighted mean
	FFARM_M	…*within Margins*	decimal	Area-weighted mean
	FFARM_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	FRURAL	Fraction of land that is low development, forested, barren rock, shrubland in watershed	decimal	Area-weighted mean
	FRURAL_M	…*within Margins*	decimal	Area-weighted mean
	FRURAL_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	FSTORAGE	Fraction of land that is open water or marsh in watershed	decimal	Area-weighted mean
	FSTORAG_M	…*within Margins*	decimal	Area-weighted mean
	FSTORAG_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	FConserve	Fraction of land that is in permanently-secured non-ag conservation	decimal	Area-weighted mean
	FConser_M	…*within Margins*	decimal	Area-weighted mean
	FConser_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	FTidalWet	Fraction of estuary land/water that is any tidal wetlands	decimal	Area-weighted mean
Soil	SOILDRAINS	Soil drainage score based on the average soil drainage class by area in the watershed, normalized 0-1, where 1 is well drained.	decimal	Area-weighted mean
	SOILDRS_M	…*within Margins*	decimal	Area-weighted mean
	SOILDRS_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	GOODDRAIN	Fraction area has well to excessively well drained SSURGO drainage classification.	decimal	Area-weighted mean
	GOODDR_M	…*within Margins*	decimal	Area-weighted mean
	GOODDR_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	MODDRAIN	Fraction area has moderately well drained SSURGO drainage classification.	decimal	Area-weighted mean
	MODDR_M	…*within Margins*	decimal	Area-weighted mean
	MODDR_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	POORDRAIN	Fraction area has somewhat poorly to very poorly drained SSURGO drainage classification.	decimal	Area-weighted mean
	POORDR_M	…*within Margins*	decimal	Area-weighted mean
	POORDR_HH	…*within Highest High Tide area*	decimal	Area-weighted mean
	SSTORAGEME	Mean available soil water storage in the watershed, based on the available water capacity and soil thickness from SSURGO.	cm	Area-weighted mean
	SSTORME_M	…*within Margins*	cm	Area-weighted mean
	SSTORME_HH	…*within Highest High Tide area*	cm	Area-weighted mean

Table C.1 continued.

Category	Name	Metric Description	Units	Aggregation
Population Patterns	Pop_Dnsity	Population density (population/watershed area) from 2010 US Census Bureau data	1/km^2	Area-weighted mean
	Pop_Dns_M	…*within Margins*	1/km^2	Area-weighted mean
	Pop_Dns_HH	…*within Highest High Tide area*	1/km^2	Area-weighted mean
	POP10	Total human population in watershed	count	Total sum
	POP10_M	…*within Margins*	count	Total sum
	POP10_HH	…*within Highest High Tide area*	count	Total sum
	StructCnt	Number of addressable structures in wshed d	count	Total sum
	StrCnt_M	…*within Margins*	count	Total sum
	StrCnt_HH	…*within Highest High Tide area*	count	Total sum
	StructDens	Number of addressable structures divided by watershed area	1/km^2	Area-weighted mean
	StrDens_M	…*within Margins*	1/km^2	Area-weighted mean
	StrDens_HH	…*within Highest High Tide area*	1/km^2	Area-weighted mean
Point Sources	OBDcnt	Number of overboard discharge sites within the watershed	count	Total sum
	OBDcnt_M	…*within Margins*	count	Total sum
	OBDcnt_HH	…*within Highest High Tide area*	count	Total sum
	OBDdens	Number of overboard discharge sites within the watershed divided by watershed area	1/km^2	Area-weighted mean
	OBDdens_M	…*within Margins*	1/km^2	Area-weighted mean
	OBDdens_HH	…*within Highest High Tide area*	1/km^2	Area-weighted mean
	CSOcnt	Number of ombined sewer overflow sites	count	Total sum
	CSOcnt_M	…*within Margins*	count	Total sum
	CSOcnt_HH	…*within Highest High Tide area*	count	Total sum
	CSOdens	Number of combined sewer overflow sites within the watershed divided by w'shed area	1/km^2	Area-weighted mean
	CSOdens_M	…*within Margins*	1/km^2	Area-weighted mean
	CSOdens_HH	…*within Highest High Tide area*	1/km^2	Area-weighted mean
	PDES_Fcnt	Number of active MaineDEP Pollutant Discharge Elimination System Facilities within watershed	count	Total sum
	PDES_Fc_M	…*within Margins*	count	Total sum
	PDES_Fc_HH	…*within Highest High Tide area*	count	Total sum
	PDES_Fdens	Number of active MaineDEP Pollutant Discharge Elimination System Facilities within watershed divided by watershed area	1/km^2	Area-weighted mean
	PDES_Fd_M	…*within Margins*	1/km^2	Area-weighted mean
	PDES_Fd_HH	…*within Highest High Tide area*	1/km^2	Area-weighted mean
	PDES_Ocnt	Number of active MaineDEP Pollutant Discharge Elimination System Outflows within watershed	count	Total sum
	PDES_Oc_M	…*within Margins*	count	Total sum
	PDES_Oc_HH	…*within Highest High Tide area*	count	Total sum
	PDES_Odens	Number of active MaineDEP Pollutant Discharge Elimination System Outflows within watershed divided by watershed area	1/km^2	Area-weighted mean
	PDES_Od_M	…*within Margins*	1/km^2	Area-weighted mean
	PDES_Od_HH	…*within Highest High Tide area*	1/km^2	Area-weighted mean

Table C.1 continued.

Category	Name	Metric Description	Units	Aggregation
Estuary Mixing	Da_Ea_r	Watershed area divided by estuary area	unitless (ratio)	Da_km2/EstA_km2
	Qmax_Ev_r	Ratio of maximum watershed runoff volume to estuary volume	unitless (ratio)	Qmax_m3/EstV_m3
	Qmin_Ev_r	Ratio of minimum watershed runoff volume to estuary volume	unitless (ratio)	Qmin_m3/EstV_m3
	Est_open	Openness of the estuary to the surrounding water body (ratio of estuary mouth length to total estuary perimeter)	unitless (ratio)	Calculated from outlet line and estuary polygon
	Bear_Out	Bearing of the estuary mouth	decimal degrees	Calculated from outlet line
	Est_circ	Shape factor: circularity of the estuary.	unitless (ratio)	Calculated from estuary polygon

<u>Areas Omitted from Analysis:</u> Some areas of the coast were omitted from the analysis due to the influence of large interior watershed runoff or the inability of unidirectional flow delineations to capture complex or temporally variable in-estuary flow paths. These regions are described in the following figures and table.

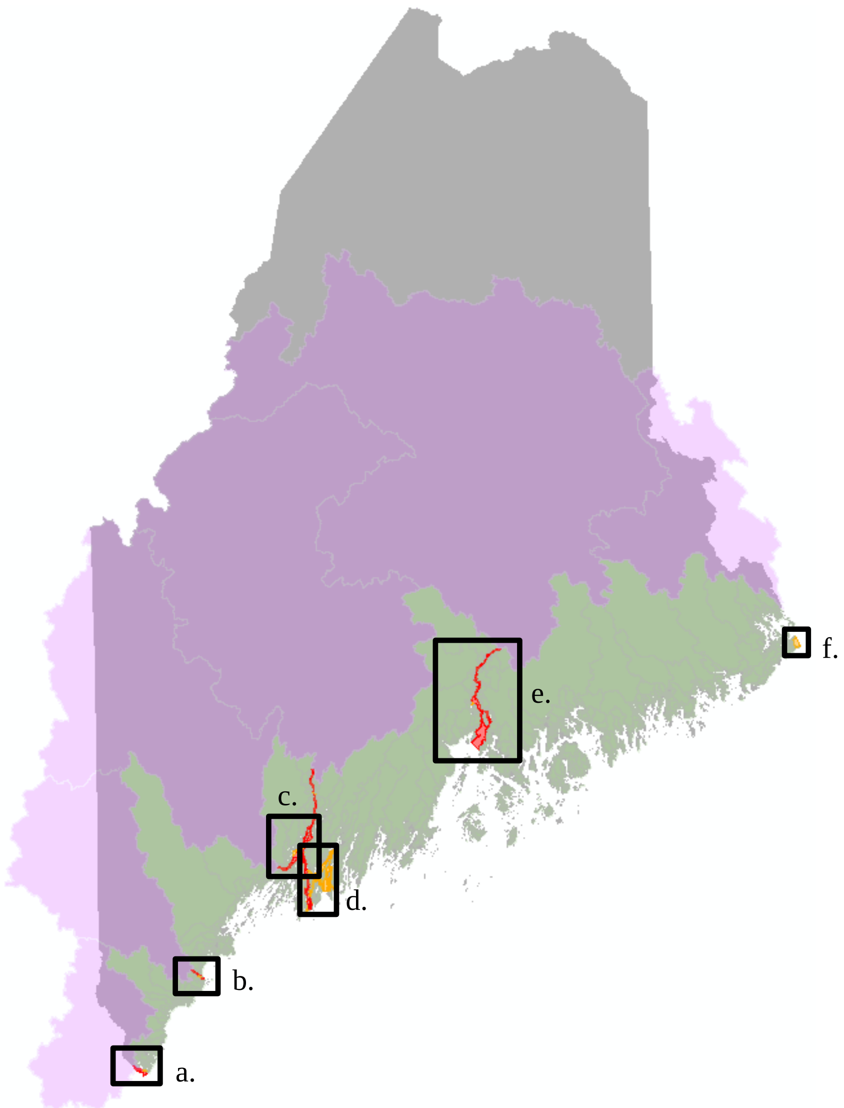

Figure C.1. Maps of Maine highlighting caution areas where the expert system tool returns incorrect (red) or potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. Six large interior watersheds omitted from tool are shown in purple. See following pages for inset box figures.

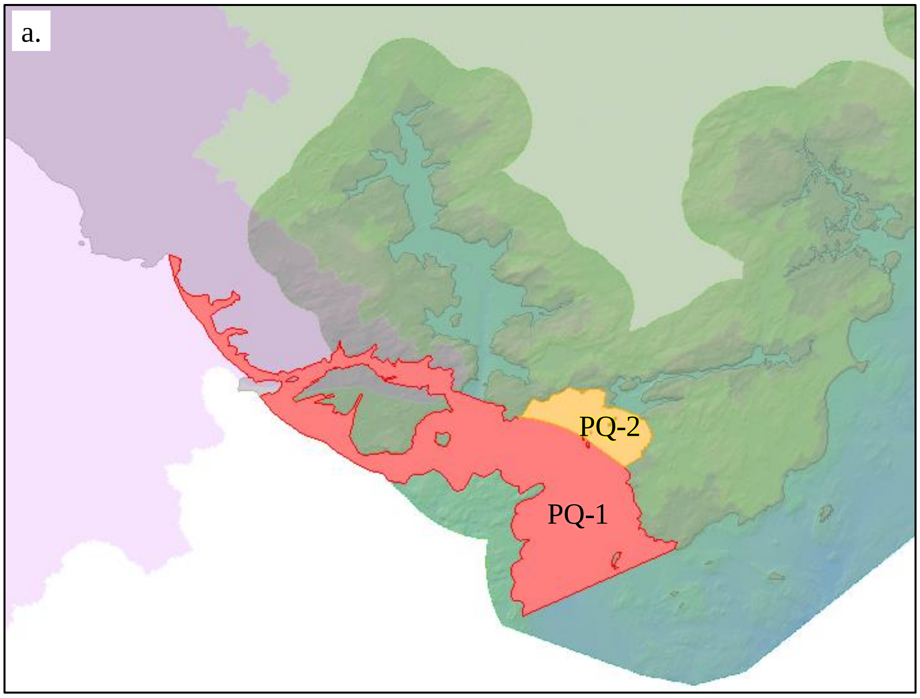

Figure C.1a. Map of lower Piscataqua River estuary where the expert system tool returns incorrect (red) or potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. Large interior watersheds omitted from tool are shown in purple. See Table C.2 for caution area descriptions.

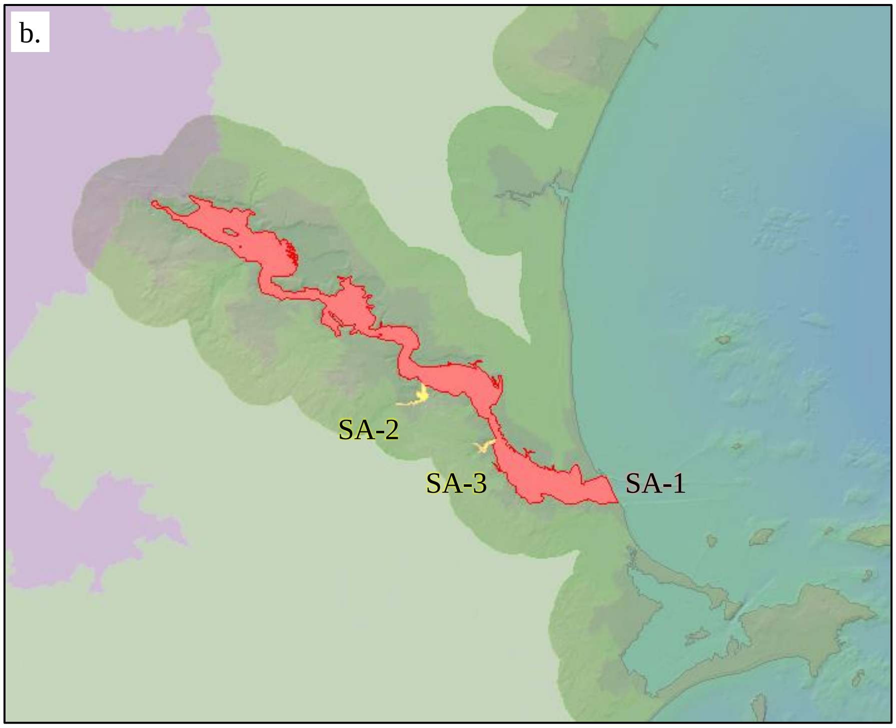

Figure C.1b. Map of lower Saco River estuary where the expert system tool returns incorrect (red) or potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. Large interior watersheds omitted from tool are shown in purple. See Table C.2 for caution area descriptions.

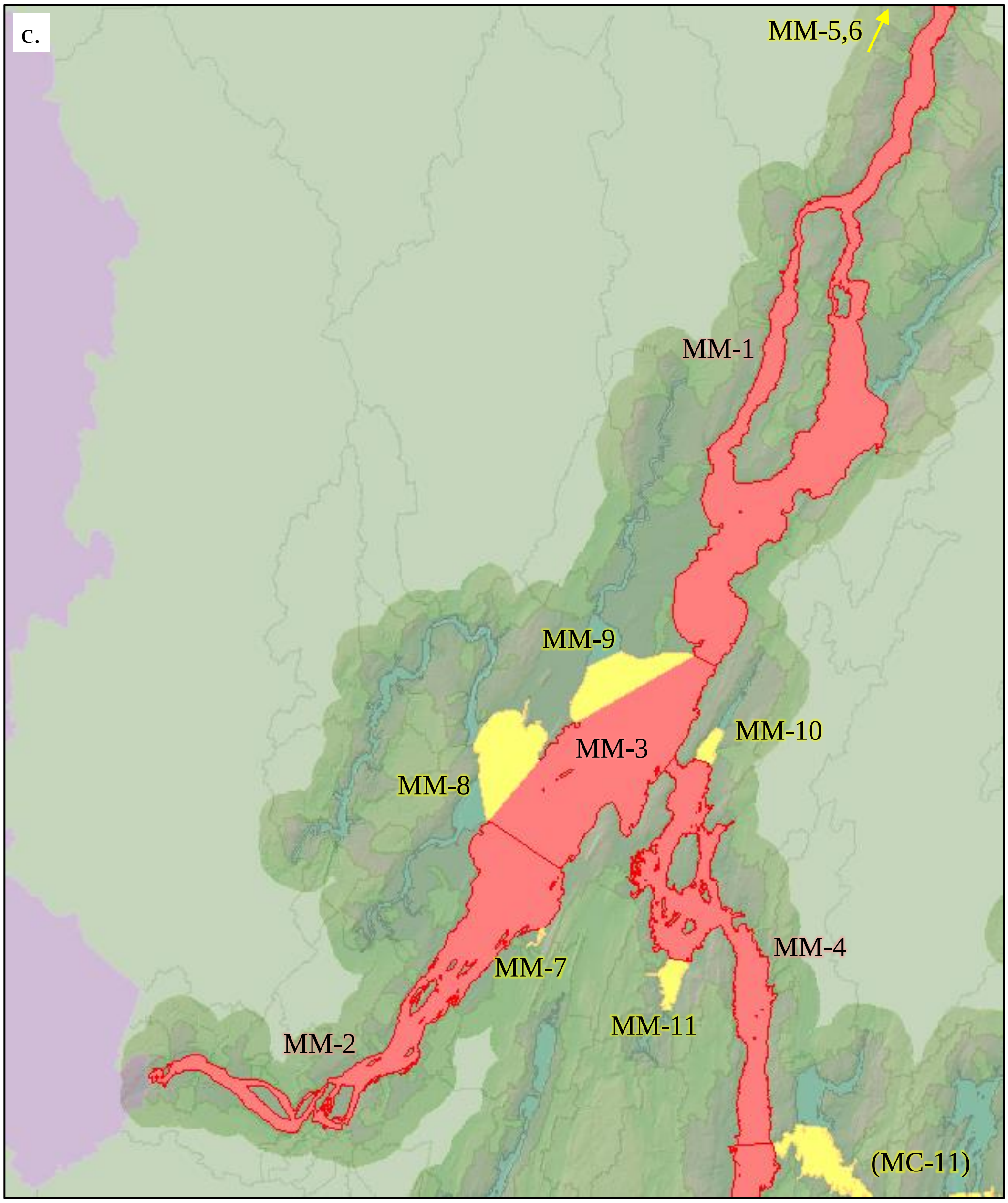

Figure C.1c. Map of Merrymeeting Bay and surroundings where the expert system tool returns incorrect (red) or potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. Large interior watersheds omitted from tool are shown in purple. See Table C.2 for caution area descriptions.

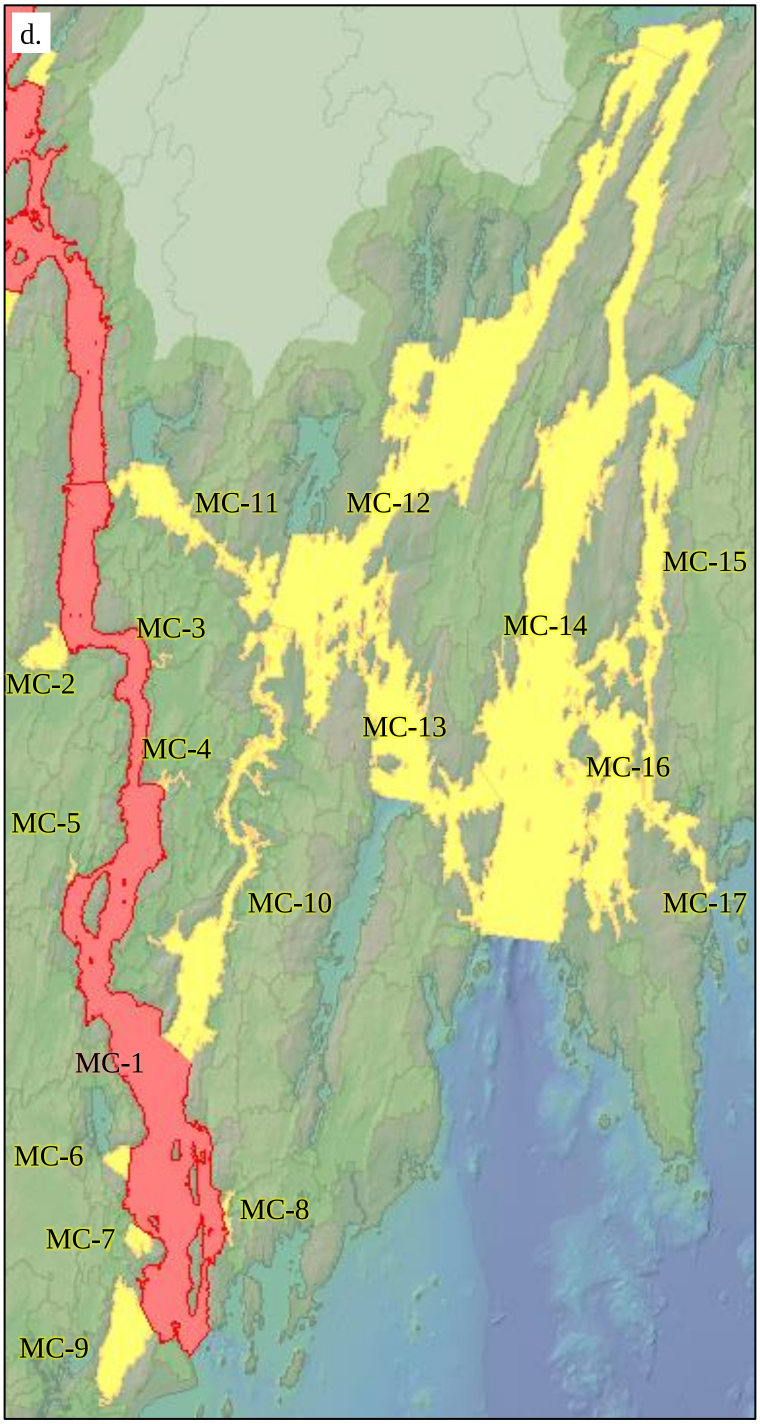

Figure C.1d. Map of ria estuaries of Midcoast Maine where the expert system tool returns incorrect (red) or potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. See Table C.2 for caution area descriptions.

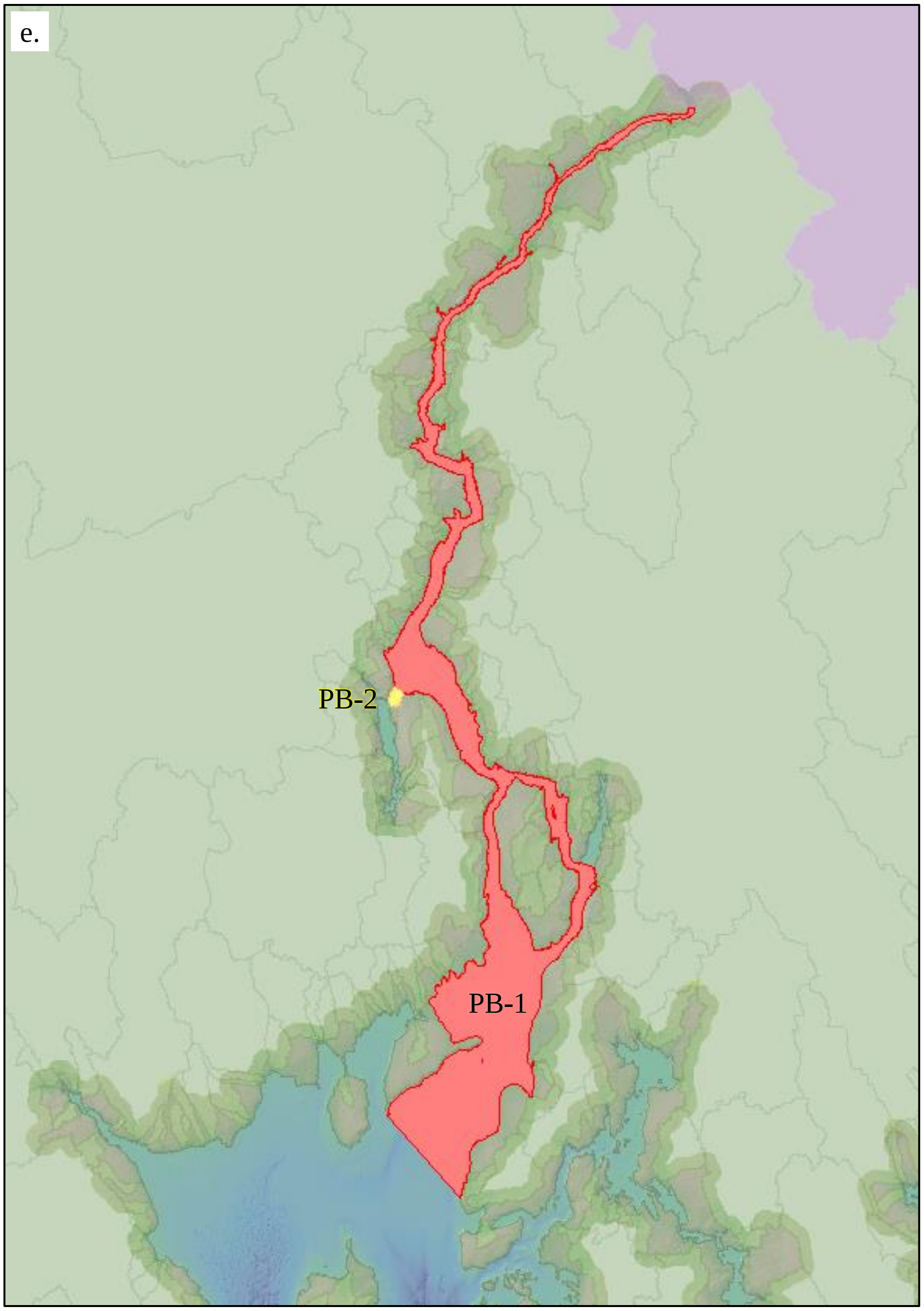

Figure C.1e. Map of lower Penobscot River estuary where the expert system tool returns incorrect (red) or potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. Large interior watersheds omitted from tool are shown in purple. See Table C.2 for caution area descriptions.

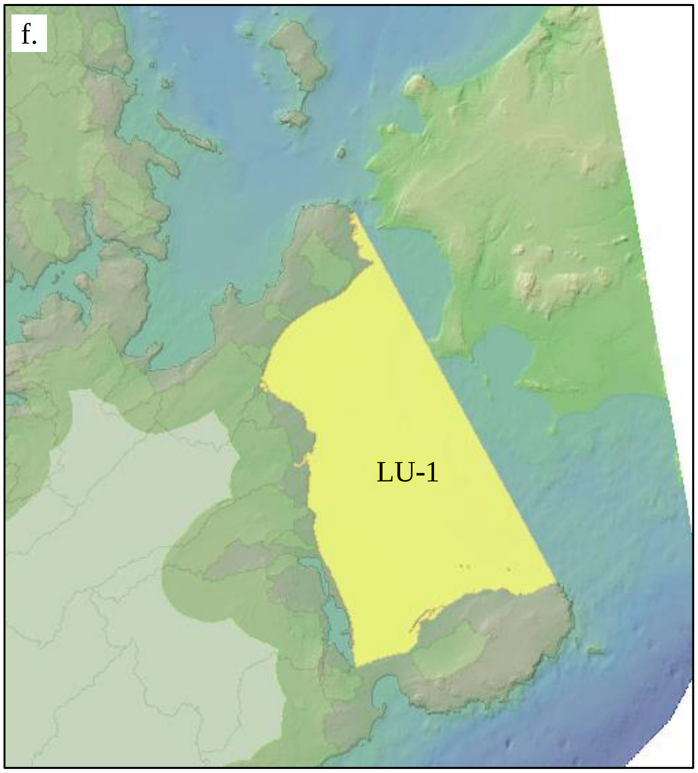

Figure C.1f. Map of Lubec Channel where the expert system tool returns potentially misleading (yellow) results. Nontidal watersheds included in tool are shown in green. See Table C.2 for caution area descriptions.

Table C.2. Explanations for advisory areas in which expert system tool may return incorrect or potentially misleading results.

Site #	Advisory	Note
MC-1	Warning – Incorrect Results	This area directly receives flow from the Kennebec and Androscoggin River watersheds, large Maine river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.
MC-2 - MC-9	Caution – Potentially Misleading	This area may be affected by flow from the Kennebec and Androscoggin River watersheds, two of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.
MC-10 - MC-13	Caution – Potentially Misleading	Flow directions through this area may be temporally variable. Delineations are best estimates of normal drainage during ebb tides, based on mud flat channel curvatures. Use of tool in Hockomock Bay / Montsweag Bay region may result in misleading returns.
MC-14	Caution – Potentially Misleading	Contributing watershed area to this section of the Sheepscot River may be temporally variable. Use of tool here may result in misleading returns.
MC-15	Caution – Potentially Misleading	Flow directions through this area may be temporally variable. Delineations are best estimates of normal drainage during ebb tides, based on mud flat channel curvatures. Use of tool here may result in misleading returns.
MC-16	Caution – Potentially Misleading	Flow mixing from the Sheepscot River main channel is not captured in flow path delineations in this area. Use of tool here may result in misleading returns.
MC-17	Caution – Potentially Misleading	Net flow directions through this area are unclear and may be temporally variable. Use of tool here may result in misleading returns.
MM-1	Warning – Incorrect Results	This area directly receives flow from the Kennebec River watershed, one of six large river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.
MM-2	Warning – Incorrect Results	This area directly receives flow from the Androscoggin River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.
MM-3	Warning – Incorrect Results	Merrymeeting Bay receives flow from the Kennebec and Androscoggin River watersheds, large Maine river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.

Table C.2 continued

Site #	Advisory	Note
MM-5 MM-6	Caution – Potentially Misleading	This area may be affected by flow from the Kennebec River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.
MM-7	Caution – Potentially Misleading	This area may be affected by flow from the Androscoggin River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.
MM-8 - MM-11	Caution – Potentially Misleading	This area may be affected by flow from the Kennebec and Androscoggin River watersheds, two of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.
PB-1	Warning – Incorrect Results	This area directly receives flow from the Penobscot River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.
PB-2	Caution – Potentially Misleading	This area may be affected by flow from the Penobscot River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.
PQ-1	Warning – Incorrect Results	This area directly receives flow from the Piscataqua River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.
PQ-2	Caution – Potentially Misleading	This area may be affected by flow from the Piscataqua River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.
SA-1	Warning – Incorrect Results	This area directly receives flow from the Saco River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here will result in incorrect returns, including drastically underrepresented runoff volume.
SA-2 SA-3	Caution – Potentially Misleading	This area may be affected by flow from the Saco River watershed, one of six large Maine river watersheds not included in this dataset. Use of tool here may result in misleading returns, depending on river stage and flow dynamics.

APPENDIX D

TECHNICAL MANUAL: EXPERT SYSTEM IMPLEMENTATION

The following pages are a reproduction of the *"WPES, How Do I…?" Quick Guide to the "Estuary Builder" GIS Tool – ArcGIS Pro 3 Version*. This document and its companion for the ArcMap 10 version were written as user manuals for the "Estuary Builder" GIS tool that is the software implementation of the coastal setting identification expert system framework developed for this dissertation research. The manual gives additional technical detail not included in the chapter text.

WPES, How Do I…?

Quick Guide to the

"Estuary Builder" GIS Tool

ArcGIS Pro 3 Version – Beta Release

Prepared by Bea E. Van Dam

Watershed Process and Estuary Sustainability Research Group
University of Maine, Orono, ME
umaine.edu/watershedresearch

Funding support for production of this publication has been provided by Maine Sea Grant Program Development grant DV-22-08 (NOAA NA22OAR4170121).

This document is formatted for print in booklet form on letter size paper.

Table of Contents

Introduction

This document is a guide to using the ArcGIS Pro 3 beta release of the "Estuary Builder" GIS Tool for delineating and classifying coastal Maine embayments.

The Estuary Builder was developed with input from Maine Department of Marine Resources Shellfish Sanitation and Management as part of the author's doctoral dissertation research into the vulnerability of different coastal Maine estuarine settings to bacteria pollution events. The tool enables a user to draw an outlet line across the mouth of any embayment or sub-embayment along Maine's tidal coastline and performs two primary tasks.

First, it delineates the drainage area contributing surface runoff through the embayment mouth and aggregates data about the landscape and embayment.

These data include factors relating to terrain, embayment morphology, land use, soil infiltration, and human population and infrastructure, and serve as proxy metrics for potential bacteria sources in the landscape and estuary, delivery to coastal shellfishing flats via surface water runoff, and residence time in coastal waters.

Second, the tool takes these aggregated data and uses a clustering algorithm to identify which of several archetypal coastal setting categories the combined landscape/estuary unit belongs to. Because setting definitions are based upon bacteria source, delivery, and residence time proxy metrics, delineated estuary units are expected to have similar bacteria vulnerability profiles and pollution event responses to other members of the same cluster.

Further information about the background of the tool and its research underpinnings are available through the Watershed Process and Estuary Sustainability (WPES) Research Group website.[A]

Instructions provided in this document assume that the user has a working knowledge of Esri's ArcGIS Pro. ArcGIS Pro help resources from Esri are available online.[B] Users who prefer to work in ArcMap (ArcGIS Desktop) can find the ArcMap version of the tool and accompanying documentation at the Estuary Builder page of the WPES website. Users who are new to ArcGIS products should consider using the ArcMap version, for which they can also consult the "WPES, How Do I...?" Quick Guide to ArcMap Basics.[C]

[A] umaine.edu/watershedresearch/estuarybuilder
[B] www.esri.com/en-us/arcgis/products/arcgis-pro/resources
[C] umaine.edu/watershedresearch/publications

Installation

This version of the Estuary Builder tool is designed to operate within Esri's ArcGIS Pro 3 software. Installation of the Estuary Builder tool requires that the user have sufficient Administrator privileges to update their ArcGIS Pro Python environment, as well as a license for the Spatial Analyst extension for ArcGIS. Prospective users without these system requirements should talk to their organizational IT departments, or may contact the WPES Research Group to inquire about potential access to the tool via virtual machine. Installation of the ArcGIS Pro version of the Estuary Builder tool requires 3.44 GB of available storage – 1.74 GB for the tool and associated spatial data, and 1.70 GB for the updated Python environment.

Download

The Estuary Builder tool, its associated files and spatial data, and a project file (map document) preconfigured for tool use are available for download as a single zipped file through the Estuary Builder webpage. A list of files included in the download and their descriptions can be found in Appendix A – Data Files.

1. Click download link for ArcGIS Pro 3 version of tool at Estuary Builder webpage.[D]
2. Unzip the downloaded folder to a location of your choosing, such as your Documents folder.

Update Python Environment

The Estuary Builder tool is written in the Python scripting language and relies on several Python "modules" or "packages" (collections of premade functions) for performing geoprocessing routines, most notably Esri's ArcPy library that mirrors geoprocessing tools found in the ArcToolbox. ArcGIS Pro uses Python 3, which is installed on your computer during ArcGIS installation. While a number of commonly used Python modules come bundled with this installation, the module used for cluster analysis

[D] umaine.edu/watershedresearch/estuarybuilder (scikit-learn) is not part of the default library and must be installed separately.

1. Open the EstuaryBuilder.arpx project file included in the Estuary Builder tool download in ArcGIS Pro.
2. Click “Project” in the ArcGIS Pro ribbon (top left of map window).
3. Click “Package Manager” in commands menu (left side of window).
4. Click gear icon (top right of window) to open “Environment Manager” pop-up window.
5. Click icon at upper right of pop-up to clone the default “arcgispro-py3” Python environment.
    - 5.1. (Optional) In “Destination” field, browse to set destination folder to location of unzipped Estuary Builder folder.
    - 5.2. (Optional) Change environment name from “arcgispro-py3-clone” to “py3-EstuaryBuilder” or other descriptive name.
    - 5.3. Click “OK” to clone.

      Note: This will take several minutes.
    - 5.4. Double-click newly cloned environment to activate and click “OK” to close Environment Manager pop-up window.
6. Click “ Add Packages” to bring up list of Python modules available for installation.
7. Type “scikit-learn” into search box.
8. Change “Version” dropdown to 1.0.2 and click “ Install” button.
    - 8.1. Check “I agree to the terms and conditions” box in pop-up window and click “Install.”

Activate Spatial Analyst

Several of the ArcPy functions used by the Estuary Builder tool require ArcGIS Pro’s Spatial Analyst extension to be active.

1. From “Project” screen in ArcGIS Pro ribbon, click “Licensing” in commands menu (left side of window).
2. Under “Esri Extensions,” check that Spatial Analyst is licensed (indicated with “Yes”).

3. If you (or your organization) have a valid Spatial Analyst license but it does not show as actively licensed, click "Configure your licensing options" to open "Licensing" pop-up window.[E]
   3.1. Select correct "License Type," "License Level," and (for Concurrent Use License) enter correct "License Manager."

   Note: Contact your organization's IT or other ArcGIS point of contact for these values.[F]

   3.2. Ensure "Licensed" box for Spatial Analyst is checked and click "OK" to close pop-up.

Initial Tool Setup

Certain default parameters can be set in the tool Properties window, most importantly the file path to the Python code upon which the tool runs.

In the file paths shown in the steps below, replace the ellipsis (...) with the location to which you unzipped the tool folder, e.g., `C:\Users\[Username]\Documents`.

1. With EstuaryBuilder.arpx map document open in ArcGIS Pro, navigate to tool location in Catalog Pane (`...\EstuaryBuilder_ArcPro3_beta\EstuaryBuilder.atbx`)

   Note: EstuaryBuilder_ArcPro3_beta folder will be set as home folder and should also appear under "Favorites" in Catalog Pane.
2. Right-click tool name and select 'Properties' from dropdown.
3. In 'Execution' tab, embedded code should appear that begins with
```
#### Import modules
import arcpy as a
```
   3.1. If this does not appear, click folder icon to browse to `...\EstuaryBuilder_ArcPro3_beta\ToolData\EstuaryBuilder_ArcPro3_Beta.py`
4. In 'Parameters' tab,
   4.1. Ensure "Default" field for "Data Folder" row points to `...\EstuaryBuilder_ArcPro3_beta\ToolData`

[E] pro.arcgis.com/en/pro-app/latest/help/analysis/spatial-analyst/basics/enabling-the-spatial-analyst-extension.htm

[F] UMaine affiliates: umservices.umaine.edu/software/arcgis/index.cgi

4.2. Ensure “Default” field for “Workspace” row points to `…\EstuaryBuilder_ArcPro3_beta\Output`

5. Click ‘OK’ to close Properties window.

Delineating Estuary Units

The Estuary Builder tool allows users to delineate estuary units using outlet lines previously saved on your computer or by interactively drawing outlet lines when the tool is run. The initial setup steps (1-5) are identical for the two methods.

1. With ArcMap open, navigate to tool location in Catalog Pane (`…\EstuaryBuilder_ArcPro3_beta\EstuaryBuilder.atbx`)

   Note: If EstuaryBuilder.arpx map document is open, EstuaryBuilder_ArcPro3_beta folder will be set as home folder and should also appear under “Favorites” in Catalog Pane.

2. Double-click tool name (or right-click > “Open…”) to open Geoprocessing pane containing Estuary Builder tool.

3. Ensure “Data Folder” file path is pointing to `…\EstuaryBuilder_ArcPro3_beta\ToolData`

   Note: To change the default file path populating this field when the tool opens, see “Initial Tool Setup,” page 3.

4. Ensure “Workspace” file path is pointing to the folder you wish delineated estuary unit polygon shapefile to be saved into.

   Note: By default this is the …\EstuaryBuilder_ArcPro3_beta\Output folder, but users may point to wherever they prefer for an individual tool run.

5. Enter desired outlet lines for delineation of estuary units using **one** of the methods below.

Delineation Using Existing Line File

6. Click folder icon to browse to location of shapefile containing outlet line(s).

7. Select shapefile and click “OK” to close “Outlet Line” dialog.

8. Click “ Run” in Estuary Builder tool window to run tool.

9. Click “View Details” under progress bar to open tool dialog.

   Continue to “Understanding Tool Output” section, page 5.

Delineation with Interactively Drawn Outlets

6. Ensure "Edge Snapping" is enabled in map document.
    6.1. To open snapping options, click "Snapping" icon next to cartographic scale at bottom of map window or type "Snapping" in Command Search field at top of window.
7. Click "Create Lines" icon , then select "Line" icon from set of feature options that appears below.
8. Draw outlet line across embayment or sub-embayment mouth in Map Window.
    8.1. Single-click shoreline at one side of embayment mouth to start line, making sure to allow cursor to snap to shoreline edge or vertex.
    8.2. Double-click shoreline at opposite side of embayment mouth to end line, again snapping to shoreline.
9. Repeat step 8 for any additional estuary units you want the tool to delineate.
10. When all outlet lines have been drawn, click " Run" in Estuary Builder tool window to run tool.
11. Click "View Details" under progress bar to open tool dialog.

    Continue to "Understanding Tool Output" section.

Understanding Tool Output

Tool Dialog

When the Estuary Builder tool is run, users can click "View Details" to open a dialog box with progress messages. The start time of the operation will appear at the top of the scrollable text box.

The next line to appear indicates how many estuary units are being processed. Below that is an estimate of when the tool will finish running, which is based on an average processing time of 3.5 minutes per estuary unit. Because actual tool performance varies with the capabilities of the computer on which it is run, users with more powerful computers may see the operation finish before the estimated time.

Next, verbal progress indicators will appear for each estuary unit being processed, as the tool delineates the contributing area, aggregates intersecting non-tidal watershed and margin area polygons, and tabulates metrics for the combined watershed polygon. The final line for each is the cluster designation indicating the coastal setting "type" that best fits the delineated estuary unit.

Once all estuary units have been processed and clustered, a success dialog will appear at the bottom of the text box and a completion time will appear at the top of the dialog box. The dialog box can then be safely closed.

Watershed Polygon

Running the Estuary Builder tool creates a polygon shapefile "EstuaryPolygons_[MMDD]_[HHmm]" and adds it to the active map window. The date and time suffix are for ease of file management and to avoid name conflicts within the output folder. The polygons contained in this shapefile represent the total contributing watershed area above each outlet line.

Attribute Table

The set of source, delivery, and residence time proxy metrics tabulated for each estuary unit can be found in the polygon shapefile's attribute table. To open the attribute table, right-click the layer name in ArcGIS Pro's Contents Pane and select "Open Attribute Table."

Metric descriptions can be found in Appendix B – Attribute Descriptions.

Cluster Designation

The Estuary Builder tool uses a Gaussian mixture model clustering algorithm to assign each delineated estuary unit to a coastal setting "type" based on its landscape and embayment metrics related to potential bacteria sources, delivery mechanisms, and estuary residence time. Estuary units within the same setting share core morphological, land use, runoff generation, and anthropogenic characteristics and as such are expected to exhibit broadly similar pollution vulnerability profiles.

Cluster designations are indicated for each estuary unit in the tool dialog when the Estuary Builder runs, as well as in the final column of the output polygon attribute table.

Important Notice: The clustering algorithm built into this beta version of the Estuary Builder tool is a preliminary version. Cluster assignment for a particular location may differ in the final release of the Estuary Builder tool. The instruction manual accompanying the final tool release will include an appendix with setting descriptions for each cluster number, as well as examples of one or more well-monitored estuary sites exemplifying the setting type. This has been deliberately omitted from this document to avoid confusion between beta and final cluster designations.

Tool Limitations

Users of the Estuary Builder tool should be aware of its limitations when interpreting output and understand that as a decision support tool it is an aid to management decisions and not a substitute for expert human judgment. The tool provides indications of general potential vulnerability to bacteria pollution and is not a predictive model for assessing how a delineated estuary unit will respond to an individual pollution event.

Caution Areas

Certain areas of the coast are known to yield incorrect or potentially misleading results in the Estuary Builder tool. The tool is not coded to prohibit use in these areas, so it is the responsibility of the user to avoid them using the "Caution Areas" layer of the EstuaryBuilder.arpx map document. Clicking on a caution area using ArcGIS Pro's "Explore" tool (in the "Map" ribbon) will display additional information about each area.

Red areas indicate sections of the coast where the tool should not be used for data aggregation or clustering. Six very large inland river watersheds

(Piscataqua, Saco, Androscoggin, Kennebec, Penobscot, and St. Croix) were omitted from the analysis behind the Estuary Builder tool due to their size. Use of the tool in coastal waters that directly receive flows from these watersheds will yield incorrect results, including drastically underrepresented watershed areas and runoff volumes.

Orange areas indicate sections of the coast where caution should be exercised when using the tool due to the potential for misleading results. Some of these are areas that may be influenced by discharge from the six omitted watersheds, while areas of the Midcoast with hydrological connections to the Hockomock Bay / Montsweag Bay region have temporally variable or otherwise complex flow patterns that can result in unreliable delineations of "upstream" areas.

Unidirectional Flow

ArcGIS routines for drainage area delineation from digital elevation models are built on the assumption of unidirectional (i.e., downhill), converging surface water flow paths. Once surface flow from the landscape reaches the coastline, flow lines are routed through and out of coastal estuaries using seafloor bathymetry. This inescapable limitation of the underlying software means that the Estuary Builder can only recognize areas that are "upstream" of the drawn embayment outlet mouth and cannot incorporate the effects of tidal dynamics or other mixing that may bring bacteria and pollutants deeper into an estuary from outside the drawn outlet. When placing outlet lines, users should look for "natural" embayment or sub-embayment divisions like constrictions or breaks in plan-form curvature and avoid splitting contiguous bodies of water where mixing is likely.

Appendix A – Data Files

The following is a list of files included in the zipped Estuary Builder download. Individual files are indicated in italics, while folders/containers are indicated in bold.

- *Estuary Builder Guide – ArcGIS Pro Beta.pdf* – this instruction manual.
- *EstuaryBuilder.aprx* – an ArcGIS Pro project file (map document) preconfigured for tool use.
- **EstuaryBuilder.atbx** – an ArcGIS Pro "Toolbox" containing the Estuary Builder tool.
- **ToolData**
    - *ArcMapBetaClusters.pkl* – a Python "pickle" file containing the clustering algorithm used to assign delineated estuary units to their respective landscape setting clusters.
    - *EstuaryBuilder_ArcPro3_Beta.py* – a Python file containing the code used to run the Estuary Builder tool.
    - **EBData_ArcPro3_beta.gdb** – an ArcGIS "Geodatabase" containing files necessary for the tool to run. Included data layers are listed below:
        - *_1_3_BathyDEM_FA* – flow accumulation raster for the Maine coast with resolution of 1/3 arcsecond. Used in drainage area delineation routine.
        - *_1_3_BathyDEM_FA* – flow direction raster for the Maine coast with resolution of 1/3 arcsecond. Used in drainage area delineation routine.
        - *BathyTopoElevations* – elevation / bathymetry raster for the Maine coast. Used as optional base map layer in map document.
        - *BathyTopoHillshade* – hillshade version of BathyTopoElevations raster. Used as optional base map layer in map document.
        - *CautionAreas* – polygon layer indicating sections of the coast where tool use may result in erroneous outcomes. Used as map document layer.
        - *CellsMargins* – polygon layer covering "Margin" watershed area, broken into ~2 million 30m x 30m cells with landscape metrics. Used in data aggregation for delineated estuary units.

- *EstuaryOutfalls* – point layer containing CSO, OBD, and PDES locations that fall within coastal waters. Used in data aggregation for delineated estuary units.
- *GoM_bathymetry* – polygon layer with simplified bathymetry for coastal Maine waters. Used in data aggregation for delineated estuary units.
- *GulfOfMaine* – polygon layer showing Maine coastal waters in the tool domain, sharing coastline with "MaineBloc" layer. Used as optional base map layer in map document.
- *HighHighNT* – polygon layer covering areas of non-tidal river/stream watersheds that are inundated during highest astronomical tide, with landscape metrics. Used in data aggregation for delineated estuary units.
- *MaineBloc* – polygon layer representing the State of Maine. Used as definitive coastline and head of tide boundaries and as map document layer.
- *ME_WetlandsGoM_E2* – polygon layer containing estuarine and marine wetland areas that fall outside the state's land boundaries. Used in data aggregation for delineated estuary units.
- *NewOutlet* – line layer that serves as template for outlet lines drawn during tool run.
- *Template* – polygon layer that serves as template for delineated estuary unit polygons created during tool run.
- *WshedsNonTidal* – polygon layer containing watershed polygons and landscape metrics for 2,195 rivers and streams that reach Maine's coastal waters, with watersheds delineated from head of tide (i.e. boundary of "MaineBloc" layer). Used in data aggregation for delineated estuary units.

- **Output** – An empty folder; included as the default location into which delineated estuary unit shapefiles are created during tool run.

Appendix B – Attribute Descriptions

The table below provides descriptions and units of measure for the source, delivery, and residence time metrics reported in the attribute table for estuary units delineated by the Estuary Builder tool. Data sources for metrics not derived directly from elevation data are denoted in brackets in the "Description" field and listed after the table. Some metric names in the polygon attribute tables are followed by _M or _H. These suffixes indicate that the metric has been calculated specifically for the "Margin" watershed area or for the coastal area inundated during Highest High tide, respectively.

Metric	Description	Units
Est_ID_	A numeric identifier for each estuary unit.	-
Da_km2	Total (terrestrial) watershed drainage area.	km^2
FMargin	Fraction of the watershed comprised of "margin" watershed area.	x/1
FHighHigh	Fraction of the watershed inundated during Highest High tide. [a]	x/1
FHiHi_M	Fraction of the watershed inundated during Highest High tide that is also within "margin" watershed area. [a]	x/1
Qmax_m3	Maximum potential precipitation runoff from watershed during 2" rain event (assuming no infiltration or storage capacity).	m^3
Qmin_m3	Minimum potential precipitation runoff from watershed during 2" rain event (subtracting maximum soil storage, see SSTORAGEME).	m^3
Zmax	Maximum elevation in watershed drainage area.	m
EstA_km2	Total surface area of delineated embayment.	km^2
EstD_m	Mean depth of delineated embayment.	m
EstV_m3	Total volume of delineated embayment.	m^3
DD_3 (_2, _05)	Natural drainage density of watershed area based on source area to channel initiation of 0.3 (0.2, 0.05) km^2.	1/km

11

Metric	Description	Units
DD_3_Rds (_2_, _05_)	Drainage density of watershed area incorporating roads as preferential flow paths in addition to natural channels. [b]	1/km
FDEVELOPED	Fraction of watershed that is in developed land cover types. [c]	x/1
FFARM	Fraction of watershed that is used for non-forestry agriculture (crops, etc.). [c]	x/1
FRURAL	Fraction of watershed that is in forested, shrubland, open land, or barren rock land cover types. [c]	x/1
FSTORAGE	Fraction of watershed that is in open water (i.e. ponds) or marsh land cover types. [c]	x/1
FConserve	Fraction of watershed in conservation areas. [d]	x/1
FTidalWet	Fraction of watershed that is tidal wetland. [e]	x/1
SOILDRAINS	Soil drainage score based on average soil drainage class by area in watershed, normalized to range 0-1 where 1 is well-drained. [f]	-
GOODDRAIN	Fraction of watershed with soil drainage class "well drained" to "excessively drained." [f]	x/1
MODDRAIN	Fraction of watershed with soil drainage class "moderately well drained." [f]	x/1
POORDRAIN	Fraction of watershed with soil drainage class "somewhat poorly drained" to "very poorly drained." [f]	x/1
SSTORAGEME	Mean available water storage depth for watershed, based on available water capacity and soil thickness. [f]	cm
Pop10	Total population in watershed. [g]	count
Pop_Dnsity	Population density in watershed.	$1/km^2$
StructCnt	Total addressable structures in watershed. [h]	count
StructDens	Addressable structure density in watershed.	$1/km^2$
OBDcnt	Number of active overboard discharge sites in estuary unit. [i]	count
OBDdns	Number of OBD sites divided by watershed area.	$1/km^2$

Metric	Description	Units
CSOcnt	Number of active combined sewer overflow discharge sites in estuary unit. [j]	count
CSODens	Number of CSO sites divided by watershed area.	$1/km^2$
PDES_Fcnt	Number of Pollutant Discharge Elimination System facilities in estuary unit. [k]	count
PDES_Fdens	Number of PDES-F sites divided by watershed area.	$1/km^2$
PDES_Ocnt	Number of Pollutant Discharge Elimination System outfalls in estuary unit. [l]	count
PDES_Odens	Number of PDES-O sites divided by watershed area.	$1/km^2$
Da_Ea_r	Ratio of (watershed) drainage area to estuary surface area.	-
Qmax_Ev_r	Ratio of maximum watershed precipitation runoff during 2" storm (see "Qmax_m3") to embayment volume.	-
Qmin_Ev_r	Ratio of minimum watershed precipitation runoff during 2" storm (see "Qmin_m3") to embayment volume.	-
Est_open	Openness of the delineated embayment, calculated as a ratio of the length of the outlet line to the perimeter of the water body.	-
Bear_Out	Bearing of the estuary outlet. This value is the azimuth (measured clockwise from north) of a line perpendicular to the drawn outlet line, pointing away from the delineated embayment.	°
Est_circ	Circularity of the delineated embayment (water body only), calculated using the Polsby-Popper test: $4\pi(\text{area})/\text{perimeter}^2$	-
Slope_Mean	Mean slope of land in delineated watershed area.	%
Slope_Max	Maximum slope of land in delineated watershed area.	%
Cluster	Cluster number (landscape setting type) assigned to the delineated estuary unit by the tool's clustering algorithm.	-

[a] "Maine Highest Astronomical Tide Line" (Maine GeoLibrary)
[b] "Maine E911 Roads Feature" (Sept. 2019) (Maine GeoLibrary)
[c] "National Land Cover Database 2016" (US Geological Survey)
[d] "Maine Conserved Lands" (Sept. 2019) (Maine GeoLibrary)
[e] "National Wetlands Inventory" (US Fish & Wildlife Service)
[f] "Soil Survey Geographic Database (SSURGO)" (US Dept. of Agriculture)
[g] "Maine Census Blocks 2010" (Maine GeoLibrary)
[h] "Maine E911 Addresses Feature" (Sept. 2019) (Maine GeoLibrary)
[i] "Overboard Discharges" (Sept. 2019) (Maine Dept. of Environmental Protection)
[j] "MaineDEP CSO" (Maine GeoLibrary)
[k] "MaineDEP Pollutant Discharge Elimination System Facility" (Maine GeoLibrary)
[l] "MaineDEP Pollutant Discharge Elimination System Outfall" (Maine GeoLibrary)

APPENDIX E

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

Table E.1. Table showing inclusion of proxy metrics in the six sets (i-vi) of Gaussian mixture model (GMM) clusters identified for setting vulnerability analysis. Proxy metric descriptions are available in Appendix C.

Metric	i	ii	iii	iv	v	vi
Da_km2	✓	✓	✓		✓	
Da_km2_M	✓		✓		✓	✓
Da_km2_HH	✓					
FMargin	✓	✓				
FHighHigh	✓					
FHighHi_M	✓					
Qmax_m3	✓	✓				
Qmin_m3	✓	✓				
Zmax	✓	✓				
EstA_km2	✓	✓	✓	✓		
EstD_m	✓	✓	✓	✓		
EstV_m3	✓	✓	✓	✓		
DD_3	✓	✓				
DD_3_M	✓		✓			
DD_3_HH	✓					
DD_2	✓	✓				
DD_2_M	✓		✓			
DD_2_HH	✓					
DD_05	✓	✓				
DD_05_M	✓		✓			
DD_05_HH	✓					
DD_3_Rds	✓	✓				
DD_3_R_M	✓		✓			
DD_3_R_HH	✓					
DD_2_Rds	✓	✓				
DD_2_R_M	✓		✓			
DD_2_R_HH	✓					
DD_05_Rds	✓	✓				
DD_05_R_M	✓		✓			
DD_05_R_HH	✓					
FDEVELOPED	✓	✓			✓	
FDEVELO_M	✓		✓		✓	✓
FDEVELO_HH	✓					
FFARM	✓	✓			✓	
FFARM_M	✓		✓		✓	✓
FFARM_HH	✓					
FRURAL	✓	✓			✓	
FRURAL_M	✓		✓		✓	✓
FRURAL_HH	✓					
FSTORAGE	✓	✓				
FSTORAG_M	✓		✓			
FSTORAG_HH	✓					
FConserve	✓	✓			✓	
FConser_M	✓		✓		✓	✓
FConser_HH	✓					
FTidalWet	✓	✓			✓	✓
SOILDRAINS	✓	✓				
SOILDRS_M	✓		✓			
SOILDRS_HH	✓					
GOODDRAIN	✓	✓				
GOODDR_M	✓		✓			
GOODDR_HH	✓					
MODDRAIN	✓	✓				
MODDR_M	✓		✓			
MODDR_HH	✓					
POORDRAIN	✓	✓				
POORDR_M	✓		✓			
POORDR_HH	✓					
SSTORAGEME	✓	✓				
SSTORME_M	✓		✓			
SSTORME_HH	✓					
Pop_Dnsity	✓	✓			✓	
Pop_Dns_M	✓		✓		✓	✓
Pop_Dns_HH	✓					

Table E.1 continued.

Metric	i	ii	iii	iv	v	vi
POP10	✓	✓			✓	
POP10_M	✓		✓		✓	✓
POP10_HH	✓					
StructCnt	✓	✓			✓	
StrCnt_M	✓		✓		✓	✓
StrCnt_HH	✓					
StructDens	✓	✓			✓	
StrDens_M	✓		✓		✓	✓
StrDens_HH	✓					
OBDcnt	✓				✓	
OBDcnt_M	✓				✓	✓
OBDcnt_HH	✓					
OBDdens	✓				✓	
OBDdens_M	✓				✓	✓
OBDdens_HH	✓					
CSOcnt	✓				✓	
CSOcnt_M	✓				✓	✓
CSOcnt_HH	✓					
CSOdens	✓				✓	
CSOdens_M	✓				✓	✓
CSOdens_HH	✓					
PDES_Fcnt	✓				✓	

Metric	i	ii	iii	iv	v	vi
PDES_Fc_M	✓				✓	✓
PDES_Fc_HH	✓					
PDES_Fdens	✓				✓	
PDES_Fd_M	✓				✓	✓
PDES_Fd_HH	✓					
PDES_Ocnt	✓				✓	
PDES_Oc_M	✓				✓	✓
PDES_Oc_HH	✓					
PDES_Odens	✓				✓	
PDES_Od_M	✓				✓	✓
PDES_Od_HH	✓					
Da_Ea_r	✓	✓		✓		
Qmax_Ev_r	✓	✓		✓		
Qmin_Ev_r	✓	✓		✓		
Est_open	✓	✓	✓	✓		
Bear_Out	✓	✓	✓	✓		
Est_circ	✓	✓	✓	✓		
Slope_Mean	✓	✓				
Slp_Mean_M	✓		✓			
Slope_Max	✓	✓				
Slp_Max_M	✓		✓			

Table E.2. Coefficients and R-squared values for linear regressions of geometric mean bacteria counts within estuary monitoring stations vs individual metric values.

Metric	Coef	R^2
Da_km2	0.00	0.06
Da_km2_M	-0.02	0.00
Da_km2_HH	-0.03	0.00
FMargin	-1.28	0.04
FHighHigh	-3.89	0.01
FHighHi_M	0.16	0.00
Qmax_m3	0.00	0.06
Qmin_m3	0.00	0.07
Zmax	0.01	0.04
EstA_km2	-0.03	0.00
EstD_m	-0.17	0.03
EstV_m3	0.00	0.00
DD_3	0.32	0.01
DD_3_M	0.61	0.01
DD_3_HH	0.09	0.04
DD_2	0.17	0.01
DD_2_M	0.49	0.01
DD_2_HH	0.08	0.03
DD_05	0.03	0.00
DD_05_M	0.19	0.01
DD_05_HH	0.08	0.04
DD_3_Rds	-0.02	0.00
DD_3_R_M	0.17	0.02
DD_3_R_HH	0.09	0.04
DD_2_Rds	0.00	0.00
DD_2_R_M	0.18	0.02
DD_2_R_HH	0.07	0.03
DD_05_Rds	-0.03	0.00
DD_05_R_M	0.20	0.03
DD_05_R_HH	0.07	0.04
FDEVELOPED	0.59	0.00
FDEVELO_M	5.14	0.05
FDEVELO_HH	3.30	0.02
FFARM	2.82	0.00
FFARM_M	2.80	0.01
FFARM_HH	4.79	0.00

Metric	Coef	R^2
FRURAL	-2.11	0.01
FRURAL_M	-2.37	0.03
FRURAL_HH	-1.03	0.00
FSTORAGE	2.67	0.01
FSTORAG_M	0.41	0.00
FSTORAG_HH	-0.50	0.00
FConserve	-1.31	0.01
FConser_M	-1.27	0.01
FConser_HH	-0.51	0.00
FTidalWet	0.60	0.00
SOILDRAINS	-2.73	0.04
SOILDRS_M	-2.30	0.04
SOILDRS_HH	-1.70	0.02
GOODDRAIN	-1.60	0.03
GOODDR_M	-1.67	0.04
GOODDR_HH	-1.53	0.03
MODDRAIN	0.83	0.00
MODDR_M	1.88	0.01
MODDR_HH	0.32	0.00
POORDRAIN	1.71	0.03
POORDR_M	1.48	0.02
POORDR_HH	1.28	0.02
SSTORAGEME	1.00	0.02
SSTORME_M	0.98	0.03
SSTORME_HH	0.49	0.03
Pop_Dnsity	0.00	0.00
Pop_Dns_M	0.01	0.08
Pop_Dns_HH	0.01	0.05
POP10	0.00	0.14
POP10_M	0.00	0.00
POP10_HH	0.00	0.00
StructCnt	0.00	0.14
StrCnt_M	0.00	0.00
StrCnt_HH	-0.01	0.00
StructDens	0.00	0.00
StrDens_M	0.01	0.07

Metric	Coef	R^2
StrDens_HH	0.00	0.00
OBDcnt	-0.03	0.00
OBDcnt_M	-0.04	0.00
OBDcnt_HH	-0.25	0.00
OBDdens	-0.39	0.00
OBDdens_M	-0.19	0.00
OBDdens_HH	0.00	0.00
CSOcnt	3.22	0.15
CSOcnt_M	4.43	0.03
CSOcnt_HH	3.22	0.15
CSOdens	5603.14	0.15
CSOdens_M	73.58	0.16
CSOdens_HH	2.08	0.16
PDES_Fcnt	0.45	0.05
PDES_Fc_M	-0.05	0.00
PDES_Fc_HH	-0.11	0.00
PDES_Fdens	-0.40	0.00
PDES_Fd_M	-0.10	0.00
PDES_Fd_HH	-0.02	0.00
PDES_Ocnt	0.11	0.03
PDES_Oc_M	-0.03	0.00
PDES_Oc_HH	0.00	0.00
PDES_Odens	-0.36	0.00
PDES_Od_M	-0.15	0.00
PDES_Od_HH	-0.02	0.00
Da_Ea_r	0.00	0.13
Qmax_Ev_r	0.00	0.01
Qmin_Ev_r	0.00	0.01
Est_open	-2.47	0.01
Bear_Out	0.00	0.00
Est_circ	-1.20	0.01
Slope_Mean	-0.12	0.03
Slp_Mean_M	-0.08	0.02
Slope_Max	0.01	0.01
Slp_Max_M	-0.01	0.01

Table E.3. Proxy metric values and cluster assignment for five possible estuary unit delineations within the Baileys Mistake embayment.

Metric	I	II	III	IV	V	Max Pct. Change
Da_km2	8.66	8.09	3.67	3.56	1.62	434%
Da_km2_M	2.86	2.30	0.50	0.40	0.39	636
Da_km2_HH	0.08	0.07	0.06	0.05	0.03	155
FMargin	0.33	0.28	0.14	0.11	0.24	196
FHighHigh	0.01	0.01	0.02	0.01	0.02	141
FHighHi_M	0.01	0.01	0.03	0.03	0.02	161
Qmax_m3	439,782.0	411,175.0	186,334.0	181,076.0	82,304.3	434
Qmin_m3	290,563.0	270,716.0	12,4814.0	120,821.0	52,455.2	454
Zmax	71.40	71.40	71.40	71.40	62.17	15
EstA_km2	1.74	1.15	0.35	0.11	0.08	2,113
EstD_m	5.29	3.40	1.11	0.22	0.14	3,669
EstV_m3	9,233,220.0	3,905,130.0	388,060.0	25,763.4	11,068.8	83,317
DD_3	1.08	1.13	0.92	0.95	0.63	79
DD_3_M	0.72	0.81	0.25	0.31	0.12	587
DD_3_HH	14.07	17.33	19.75	21.00	16.54	49
DD_2	1.48	1.53	1.30	1.34	0.89	72
DD_2_M	0.97	1.03	0.36	0.46	0.27	281
DD_2_HH	15.90	19.58	22.40	23.82	21.18	50
DD_05	2.88	2.97	3.03	3.11	2.98	8
DD_05_M	2.09	2.22	1.25	1.58	1.42	77
DD_05_HH	18.12	22.32	24.99	26.58	23.96	47
DD_3_Rds	2.09	2.21	2.61	2.67	2.44	28
DD_3_R_M	1.64	1.96	1.63	1.97	1.82	21
DD_3_R_HH	14.66	18.05	20.37	21.66	17.62	48
DD_2_Rds	2.49	2.61	2.99	3.06	2.70	23
DD_2_R_M	1.89	2.17	1.75	2.12	1.97	24
DD_2_R_HH	16.48	20.30	23.02	24.48	22.26	49
DD_05_Rds	3.89	4.05	4.72	4.84	4.79	24
DD_05_R_M	3.01	3.37	2.64	3.24	3.12	27
DD_05_R_HH	18.71	23.04	25.61	27.23	25.04	46
FDEVELOPED	0.01	0.01	0.02	0.02	0.03	184
FDEVELO_M	0.01	0.02	0.03	0.04	0.04	238
FDEVELO_HH	0.05	0.06	0.06	0.06	0.10	118
FFARM	0.00	0.00	0.00	0.00	0.00	229
FFARM_M	0.00	0.00	0.00	0.00	0.00	172
FRURAL	0.85	0.85	0.89	0.89	0.91	8
FRURAL_M	0.83	0.83	0.86	0.87	0.89	7
FRURAL_HH	0.16	0.15	0.11	0.07	0.11	123
FSTORAGE	0.14	0.14	0.09	0.09	0.05	176
FSTORAG_M	0.15	0.15	0.11	0.08	0.07	124

Table E.3 continued.

Metric	I	II	III	IV	V	Max Pct. Change
FConserve	0.16	0.16	0.02	0.02	0.02	923%
FConser_M	0.21	0.23	0.06	0.08	0.08	255
FConser_HH	0.10	0.05	0.02	0.02	0.02	436
FTidalWet	0.01	0.01	0.03	0.03	0.05	310
SOILDRAINS	0.57	0.57	0.53	0.52	0.45	27
SOILDRS_M	0.53	0.51	0.48	0.45	0.45	18
SOILDRS_HH	0.19	0.17	0.14	0.12	0.09	103
GOODDRAIN	0.73	0.72	0.63	0.63	0.50	47
GOODDR_M	0.65	0.59	0.54	0.49	0.50	32
GOODDR_HH	0.17	0.11	0.08	0.04	0.06	287
MODDRAIN	0.04	0.04	0.02	0.02	0.03	103
MODDR_M	0.08	0.10	0.04	0.00	0	-
MODDR_HH	0.02	0.03	0.00	0.00	0	-
POORDRAIN	0.23	0.24	0.35	0.35	0.47	104
POORDR_M	0.27	0.31	0.42	0.51	0.50	85
POORDR_HH	0.80	0.86	0.92	0.95	0.94	18
SSTORAGEME	1.72	1.74	1.68	1.69	1.84	10
SSTORME_M	1.84	1.91	1.57	1.66	1.66	22
SSTORME_HH	1.99	2.33	2.62	2.73	3.06	53
Pop_Dnsity	5.65	5.78	4.51	4.38	4.24	36
Pop_Dns_M	6.38	7.02	5.34	4.38	4.45	60
Pop_Dns_HH	1.92	1.50	0.53	0.20	0.28	838
POP10	49	47	17	16	7	600
POP10_M	18	16	3	2	2	800
StructCnt	35	34	18	17	8	338
StrCnt_M	18	17	3	2	2	800
StrCnt_HH	1	1	1	1	1	0
StructDens	4.04	4.20	4.91	4.77	4.94	22
StrDens_M	6.30	7.41	5.98	5.02	5.15	47
StrDens_HH	1.63	2.01	1.06	1.13	1.85	89
Da_Ea_r	4.96	7.06	10.50	31.03	20.55	525
Qmax_Ev_r	0.05	0.11	0.48	7.03	7.44	15,511
Qmin_Ev_r	0.03	0.07	0.32	4.69	4.74	14,959
Est_open	0.08	0.04	0.15	0.14	0.14	291
Bear_Out	161.86	275.35	121.16	119.72	95.56	188
Est_circ	0.22	0.23	0.24	0.20	0.20	22
Slope_Mean	7.40	7.01	7.47	7.22	8.31	19
Slp_Mean_M	8.71	7.65	11.52	10.32	10.39	51
Slope_Max	115	84	84	60	60	92
Slp_Max_M	115	84	84	56	56	105
Cluster	**5**	**1**	**3**	**4**	**4**	-

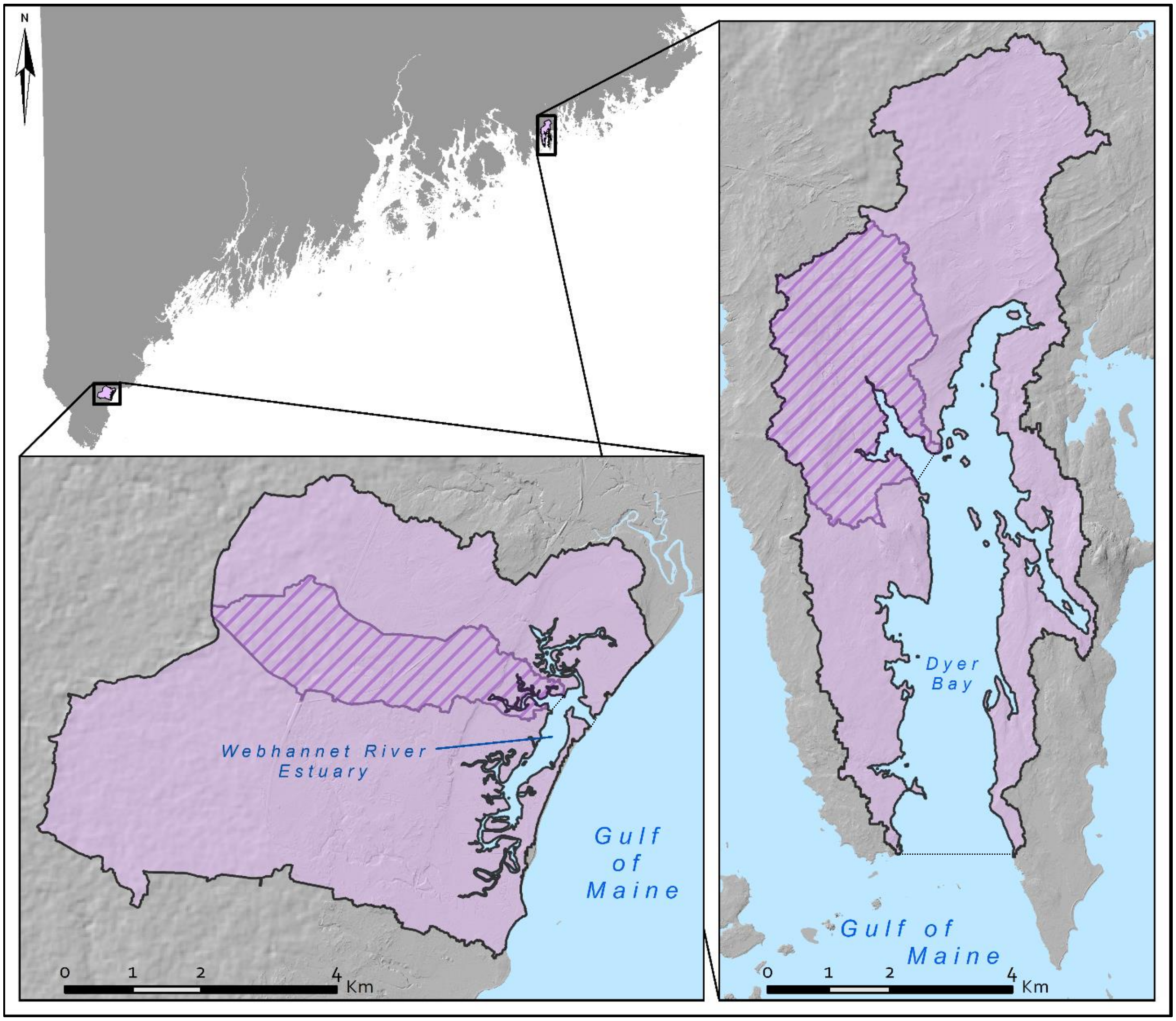

Figure E.1. Illustration of sub-embayments. Examples of delineated watershed areas for nested sub-embayments, or "inner" embayments (hatched), within larger embayment complexes ("outer" embayments) (solid) in a riverine estuary system (Webhannet River, Wells; left) and a bay system (Dyer Bay, Steuben; right).

BIOGRAPHY OF THE AUTHOR

Bea Van Dam was born in Hartford, Connecticut. She was raised in Scarborough, Maine and graduated from Scarborough High School in 2006. She attended the University of Maine and graduated *summa cum laude* in 2011 with a Bachelor of Science degree in Parks, Recreation and Tourism and minors in Sociology and Economics. After gaining professional experience in municipal geographic information systems (GIS) and trail surface water runoff and erosion management, she returned to the University of Maine to pursue graduate studies in the field of hydrology. She was inducted into the Honor Society of Phi Kappa Phi in 2023.

Bea is a candidate for the Doctor of Philosophy degree in Earth and Climate Sciences from the University of Maine in August 2023.