Title: Minimum Sampling, Maximum Insight: Tracking Environmental Trends in a Tidal Estuary

Authors: *Barbara Spiecker[1], Kalle Matso[2], Chris Whitney[2], Easton R. White[1]*

Affiliations:

[1]Department of Biological Sciences, University of New Hampshire, Durham, New Hampshire, USA 03824.

[2]Piscataqua Region Estuaries Partnership, University of New Hampshire, Durham, New Hampshire, USA 03824.

*Corresponding Author: Barbara Spiecker (barbara.spiecker@gmail.com)

Running Head: Optimizing Estuarine Monitoring Design

Abstract:

Long-term environmental monitoring is essential for detecting ecological trends and managing dynamic systems. In estuarine environments, where monitoring is often constrained by cost and logistics, efficient resource allocation is key to sustaining effective programs. We developed a framework to optimize spatial and temporal sampling in the Great Bay Estuary (New Hampshire/Maine, USA), identifying the minimum number of years and sites needed to detect long-term trends. Using 23 years of data on five water quality parameters from 10 sites, we applied a resampling-based trend detection algorithm to estimate minimum sampling effort. Our results show that trend detectability varies by parameter and location, with each requiring different sampling durations. These thresholds also depend on the user-defined level of statistical power (e.g., 80% vs. 100%). For example, nitrogen trends were detectable with as few as five years of data, while dissolved oxygen required up to seven. Additionally, 8–9 sites were sufficient to achieve 80% statistical power, suggesting spatial redundancy at some locations. Variance partitioning revealed that autocorrelation, slope error, and data variability each influenced sampling effort for reliable trend detection. Parameters such as dissolved oxygen and water temperature—both highly autocorrelated—required longer time series, while those with lower slope precision, like nitrogen, were sensitive to measurement error. These findings underscore the value of adaptive monitoring designs that align sampling strategies with the statistical and ecological characteristics of individual parameters. Our approach provides a flexible, data-driven framework for refining estuarine monitoring programs, enabling managers to maintain ecological insight while optimizing resource use.

Impact Statement: Monitoring environmental change is critical for managing estuaries, yet limited resources often constrain long-term data collection. Using two decades of water quality data from New Hampshire/Maine's Great Bay Estuary, we developed a method to identify the minimum number of years and sites needed to detect long-term trends. Our approach shows that different environmental parameters require different monitoring strategies, and that some sites may be redundant. This data-driven framework helps optimize sampling design, allowing programs to reduce costs without sacrificing scientific insight. While demonstrated in one estuary, the approach should be tested in other systems to evaluate broader applicability. By aligning monitoring effort with the characteristics of each parameter, managers can make informed decisions, allocate resources efficiently, and support long-term monitoring.

Keywords: environmental monitoring; estuarine systems; trend detection; sampling optimization; statistical power

Introduction

Long-term monitoring programs are the backbone of modern ecological research and ecosystem management (Possingham et al. 2012, Giron-Nava et al. 2017). Despite the cost and logistical complexities of maintaining long-term monitoring programs, past work has highlighted their importance (Maguran et al. 2010). Given the difficulty of long-term monitoring, the monitoring programs must be designed to address research questions of interest properly and for ecosystem management. Yet, many ecological monitoring programs are likely underpowered in terms of the number of years of sampling and their spatial scales (Legg and Nagy 2006, Field et al. 2007, Rhodes and Jonzen 2011, White 2019). Conversely, some monitoring programs may be spending unnecessary effort and resources (Caughlan 2001, Morant et al. 2020). These issues are potentially compounded by the fact that monitoring data may now be used for purposes that were not originally envisioned.

Past work on the design of ecological monitoring programs has focused on the minimum number of years required to detect changes in population abundance (Caughlan 2001, Field et al. 2007, Lindenmayer and Likens 2010, Legg et al. 2017, White 2019). There is also a long history of studying the spatial scale of sampling in ecology (Legendre et al. 2002). There has been a renewed interest in this work given access to larger datasets, the combining of data from different monitoring programs, and the cost constraints of monitoring (Fletcher Jr. et al. 2019). The renewed interest in spatial monitoring questions is also evident with the proliferation of software, including the spsurvey R package, that helps facilitate the spatial design of monitoring programs (Barry et al. 2017, Regular et al. 2020). For example, Bashevkin (2022) examined how reduced sampling of fish populations in the San Francisco Bay Estuary would affect inference. He found that 10-20% reductions in the number of stations monitored would not alter the ability to detect long-term trends. However, Bashevkin (2022), and other work, have not assessed the variation in monitoring requirements across multiple variables within the same system.

By their very nature, estuaries are dynamic ecosystems given their tidal fluxes and inputs from freshwater and terrestrial systems (Kalra et al. 2020, Mulukutla et al. 2021, Boynton et al. 2022). Thus, high-resolution monitoring, both temporally and spatially, may be required to capture the dynamics of estuarine systems. Given this variability, there is also potential for individual locations within any estuary to exhibit a great degree of variability from one another (PREP 2023a). Given high levels of variability spatially and temporally in estuaries, it is important to carefully design their monitoring programs, considering the number of sampling locations and the frequency in which sampling should occur. In addition, the information gained from monitoring environmental changes in estuaries also must be weighed in comparison to the cost of maintaining several monitoring sites over time. All these considerations are likely to vary between environmental parameters. For example, even though temperature and dissolved oxygen may be collected on similar timescales, such as every 15 minutes by a sensor, the trends and variability in each may mean the data required for detecting change may be unique to each variable.

To investigate sampling questions, we examined long-term monitoring data within Great Bay Estuary (GBE). GBE is dynamic tidal estuary on the New Hampshire coast fed by seven rivers (Jones 2000). The extreme tidal fluxes provide various types of habitats (e.g., mudflats, saltmarshes) that support numerous species (Jones 2000, Cook et al. 2019). There have been various water quality monitoring programs in GBE over the past few decades, largely in the form of monthly grab samples across the estuary. The grab samples are expensive to collect and process. Therefore, GBE presents an interesting case study to investigate questions of sampling effort optimization. In addition to the grab samples, approximately eight of the estuarine stations have automated sensors collecting parameters such as temperature, salinity and dissolved oxygen; however, these sondes do not collect information on nutrient species, a critical parameter in this estuary.

The Great Bay Estuary receives nitrogen loading from both point sources (i.e. 13 waste water treatment facilities (WWTF) draining to GBE) as well as numerous nonpoint sources throughout the watershed. As such, the Environmental Protection Agency (EPA) Region 1 issued a total nitrogen general permit for these WWTFs that release effluent to the Great Bay (EPA Great Bay TN General Permit). These increased nitrogen inputs drive growth of phytoplankton and, along with increases in suspended solids, has been partly responsible for a decrease in eelgrass (*Zostera marina*), a key indicator species in GBE (PREP 2023a). Environmental monitoring in GBE, both sondes and traditional grab samples, aim to track these variables that are believed to be affecting the health of eelgrass.

In this paper, we examine data on environmental parameters (e.g., dissolved oxygen, temperature) from 2000-2023 in Great Bay Estuary, New Hampshire. Using non-random resampling techniques, we tested several interrelated questions: 1) What is the minimum number of years needed to detect long-term trends for each environmental parameter?, 2) What is the minimum number of monitoring sites required to detect long-term trends for each environmental parameter?, 3) Which monitoring sites provide the most unique trend information, and which are most redundant relative to estuary-wide trends?, and 4) What factors influence the minimum sample size estimation? We hypothesized that parameters that were less variable between years and across sites, as well as those that had strong temporal trends, would require less data. We provide general recommendations for sampling these types of systems and potential next steps for research.

Methods

Environmental Data

We extracted environmental data from the Great Bay Estuary in New Hampshire, USA, using the Piscataqua Region Estuaries Partnership (PREP) database at: http://data.prepestuaries.org/data-explorer/. This particular dataset includes five environmental parameters: dissolved oxygen (both concentration and saturation), nitrite and nitrate levels, suspended solids, and water temperature. Data were collected year-round through monthly grab samples at 10 sites within the Great Bay Estuary (Fig. 1). For more details on data and sampling methods, see the PREP 2023 State of Our Estuaries Extended Report (PREP 2023b). We then calculated the yearly averages for each environmental parameter at every sampled site within the estuary. Quality assurance project plans on various monitoring programs in GBE can be found at: https://scholars.unh.edu/prep/.

Data Cleaning

We identified active sites with sufficient years sampled for each environmental parameter using the following criteria: (1) at least 15 years of data collected starting in 2000. For each parameter, 2000 is the earliest year commonly sampled across sites. (2) at least one year sampled between 2019 and 2023. This is to ensure the sites are recently sampled (or in other words, active). (3) less than 25% missing data (or less than ¼ of the years unsampled). This selection yielded 10 active sites (Fig. 1).

After identifying the sampling sites, we evaluated the most suitable method for imputing missing values, grouping the data by site and environmental parameter. We used Multivariate Imputation by Chained Equations (MICE) to compare three methods: Predictive Mean Matching (PMM), Classification and Regression Trees (CART), and Lasso Linear Regression (Lasso). Among these, PMM provided the best fit to the original data. Therefore, missing values in the dataset were imputed using PMM.

Minimum Time

We conducted non-random resampling of the Great Bay Estuary dataset spanning from the year 2000 to 2022 using the Broken Window algorithm (Bahlai et al. 2021) to determine the minimum time required for trend detection (Fig. 2). This method, involves partitioning the data into a series of smaller samples for comparative analysis (White 2019, White and Bahlai 2021, Bahlai et al. 2021). We subdivided the 22-year dataset into contiguous chunks of varying lengths for each environmental parameter. These subsamples included one 22-year subsample, two 21-year subsamples, three 20-year subsamples, and so forth. Subsequently, we computed the population trend for each subsample by determining the slope through a linear regression model, with year as a fixed effect, site as a random effect, and environmental values as the response variable. Each subsample slope would then be compared to the slope derived from the complete 22-year dataset, which served as the "true trend" for comparison.

We assessed whether a sample trend accurately represented the complete trend by evaluating if the sample slope fell within 1 standard deviation of the complete slope, indicating representativeness in magnitude (Fig. 2). The proportion of subsamples of a specific length exhibiting the same overall magnitude as the complete time series constituted the statistical power (typically set at 0.8, Cohen 1992). We identified which subsample length was required to achieve certain thresholds of statistical power: 0.8 (with 80% of the subsample slopes matching the complete slope) and 1 (with 100% of them matching the complete slope). This subsample length would become the minimum time series length required to detect trends.

To explore why certain environmental parameters exhibited shorter or longer minimum times for trend detection, we conducted variance partitioning using four metrics: 1) slope strength, which reflects the magnitude of change; 2) slope standard error, which measures the precision of the trend estimate; 3) data variability, measured as the standard deviation, which captures the natural fluctuation in the data; and 4) lag-1 autocorrelation, which accounts for the influence of temporal dependence between observations. We also calculated the coefficient of variation for each environmental parameter to allow standardized comparison across parameters with different units and scales.

For the variance partitioning analysis, we first ran linear regression models for each environmental parameter at each site to extract variance components, along with the minimum time at the 100% representativeness threshold associated with each parameter-site combination. These site-level variance components served as replicates for each parameter and were used as predictor terms in a second regression model. The response variable was the minimum time required. We quantified the relative contribution of each variance metric by calculating the percentage of variance it explained, determined by dividing the sum of squares for each term by the total sum of squares in the model.

Minimum Site

To determine the minimum number of sites necessary for trend detection, we adapted the Broken Window algorithm (as detailed in the Minimum Time section) to systematically partition sampled sites into groups of varying sizes for each environmental variable. The protocol follows a two-step process. First, for each level of site reduction (starting with the removal of 10% of sites), we generated all possible combinations of the remaining sites. For each combination, we applied a linear regression model to the environmental data, with the model structure depending on the number of sites included. If a combination contained three or more sites, we fit a mixed-effects model with year as a fixed effect, site as a random effect, and the environmental parameter as the response. If a combination included fewer than three sites, we fit a simpler linear model without random effects, as fewer than three levels do not provide sufficient information to estimate variance reliably in a random term. This process was repeated for increasing levels of site removal in 10% increments (e.g., 20%, 30%, and so on).

In the second step, following the same approach outlined in the Minimum Time section, we compared the slope from each reduced model to the slope from the full model, which included all sites and treated site as a random effect, to assess whether the subset trend accurately represented the complete trend. A reduced model was considered representative if its slope fell within one standard deviation of the full-model slope. For each level of site reduction, we calculated the proportion of subset combinations that met this criterion. We then identified the minimum number of sites required to achieve 80% representativeness (i.e., 80% of subset slopes falling within one standard deviation of the full-model slope) and 100% representativeness (i.e., all subset slopes within this threshold). These thresholds defined the minimum number of sites necessary to reliably detect trends for each environmental parameter.

We ran a variance partitioning analysis for minimum site requirements using a similar approach to the one described in the minimum time section. However, instead of modeling each environmental parameter-site combination, we treated each environmental parameter as a single replicate and used the number of sites required to reach the 100% trend representativeness threshold as the response variable. Due to the limited number of replicates and associated statistical power, we included only three variance components as predictors: lag-1 autocorrelation, slope standard error, and data standard deviation. These components were extracted at the parameter level and entered into a single regression model to evaluate their relative influence on the minimum number of sites required. To quantify the contribution of each predictor, we calculated the percentage of variance explained by dividing the sum of squares for each term by the total sum of squares in the model.

Site Uniqueness

To determine which site(s) provided the most or least unique information for trend detection at the 80% representativeness threshold, we used linear regression analysis across all environmental parameters and sites. For each site, we compared the slope from a reduced model (with the site removed) to the slope from the full model (including all sites). The absolute difference between the reduced model slope and the full model slope was used as a measure of that site's deviation from the estuary-wide trend. These slope differences were then standardized (z-scored) to enable comparison across parameters.

The most unique site was defined as the one with the largest standardized slope difference (i.e., the most positive z-score), indicating that its removal caused the greatest deviation from the full-model trend. A positive z-score reflects that the site's data differ meaningfully from the estuary-wide pattern and contribute uniquely to trend detection.

In contrast, the least unique site was defined as the one with the smallest standardized slope difference (i.e., the most negative z-score), indicating strong agreement with the full-model trend. A negative z-score reflects that removing the site caused a smaller-than-average change in the overall slope, suggesting that its contribution is largely redundant with other sites and adds limited additional information to the network.

To identify the most representative combination of sites when a smaller number of sites is considered sufficient for monitoring (as determined by the minimum site analysis), we applied linear regression analysis to all possible combinations of that site count. We then compared the slope from each subset model to the slope from the full model. The combination with the smallest standardized slope difference was considered the most representative of the estuary-wide trend.

Data and Code

All data, analyses, and visualizations were performed in RStudio (version 2024.09.1+394). All code used in this study is available at https://github.com/QuantMarineEcoLab/prep-sampling-optimization.

Results

Our results indicate that trends vary across sites by parameter, and that each parameter requires a different minimum sampling effort for trend detection. Importantly, the estimated minimum depends on the user-defined threshold for representativeness—in this case, 80% or 100%.

Environmental Parameters

Across the 2000–2020 period, most environmental parameters exhibited weak or non-significant linear trends (Fig. 3). Dissolved oxygen concentration, dissolved oxygen saturation, nitrite + nitrate, and water temperature all showed minimal directional change over time, with none reaching statistical significance ($p > 0.05$). While some exhibited slight declining or flat trends, interannual variability was relatively high. The exception was suspended solids, which displayed a statistically significant increasing trend ($p = 0.001$), indicating a consistent rise in concentration over time.

To better understand spatial variability in water quality trends, we examined how site-level patterns differed across the estuary (Fig. 4). For dissolved oxygen saturation, nitrite + nitrate, and water temperature, most sites (> 7) exhibited negative z-scores, indicating broadly consistent temporal trends across the estuary. In contrast, dissolved oxygen concentration and suspended solids showed greater site-level variability, with more positive z-scores (> 5) suggesting more spatially localized trends.

Across parameters, the Route 9 Bridge and Squamscott River consistently emerged as the most unique sites, with positive z-scores (ranging from 1.6 to 2.7) indicating trends that deviated notably from those observed elsewhere. Conversely, Adams Point frequently exhibited negative z-scores (ranging from –0.19 to –0.86), reflecting patterns more closely aligned with the estuarine average and likely more representative of system-wide behavior.

Minimum Time

Monitoring duration should be at least 7 years to ensure reliable trend detection across all parameters (Table 1). However, the minimum required duration varies by parameter. For instance, nitrite + nitrate achieved 100% representativeness with just 5 years of data, whereas both dissolved oxygen metrics and water temperature required the full 7 years. Furthermore, the difference between the thresholds we used—80% and 100% representativeness—was typically minimal, varying by only about one year.

We identified four key drivers influencing the minimum time required for trend detection: slope magnitude, slope standard error, data standard deviation, and lag-1 autocorrelation. Although minimum durations were broadly similar across parameters, the primary drivers behind those durations differed. Lag-1 autocorrelation was the strongest driver for both dissolved oxygen metrics, which also had the lowest coefficients of variation. For dissolved oxygen saturation, slope standard error emerged as the second most influential factor. In contrast, water temperature, nitrite + nitrate, and suspended solids exhibited high residual variance, with nearly half of their variation unexplained. Water temperature was influenced almost equally by data standard deviation and autocorrelation. Nitrite + nitrate was primarily driven by slope standard error and autocorrelation, while suspended solids were influenced solely by data standard deviation. Interestingly, nitrite + nitrate and suspended solids had the highest coefficients of variation (exceeding 100%), indicating that the standard deviation was greater than the mean. Despite this high variability, these parameters required the shortest monitoring periods to detect trends.

Minimum Site

To achieve 100% trend representativeness across all environmental parameters, monitoring all 10 sites is necessary. However, if 80% representativeness is deemed sufficient, only 8 to 9 sites are required (Table 2). For dissolved oxygen concentration, monitoring just 8 sites is adequate (Table 3), with Adams Point and Oyster River identified as the most redundant—and therefore the least essential to include (Table 4). For the remaining four environmental parameters, 9 sites are sufficient to maintain trend accuracy, with Adams Point once more appearing as the least necessary site in two of those cases (nitrite + nitrate and suspended solids). In the case of dissolved oxygen saturation and water temperature, Route 9 Bridge (Central Avenue) and Squamscott River, respectively, were identified as the least essential sites.

We identified three key drivers influencing the number of sites required to detect long-term trends at the 100% representativeness threshold: lag-1 autocorrelation, slope standard error, and data standard deviation. Across all environmental parameters, these predictors contributed relatively evenly to the variance in minimum site requirements. Lag-1 autocorrelation accounted for the largest share, explaining 38% of the variance, while slope standard error and data standard deviation each contributed 26%. The remaining 9% of the variation was unexplained by the model.

Discussion

Long-term environmental monitoring programs are essential for understanding ecological trends and managing dynamic systems like the Great Bay Estuary (Wolfe et al. 1987, Cloern and Jassby 2012, Kennish 2019). Efficient resource allocation is also critical for the success of these programs, particularly in estuarine systems where cost and logistical constraints often limit the scope of data collection (Caughlan and Oakley 2001). This study offers a framework for optimizing spatial and temporal monitoring by identifying the minimum effort needed to detect long-term trends across multiple environmental parameters in the Great Bay Estuary.

Temporal Monitoring

Our findings reveal that the minimum number of years required to detect long-term trends varies by environmental parameter, largely depending on the strength and variability of the signal. While most parameters required up to seven years of data to reach 100% trend representativeness, nitrite + nitrate achieved this threshold in just five years. Interestingly, both nitrite + nitrate and suspended solids exhibited the highest coefficients of variation, suggesting that high variability does not necessarily extend the time needed for trend detection. These results suggest that factors beyond overall variability, such as autocorrelation and slope standard error, play a more critical role in determining the monitoring duration needed.

In particular, the influence of autocorrelation underscores the need for caution when estimating sample size and interpreting trend reliability. Autocorrelation within ecological time series reduces the amount of independent information, making it appear as though the sample size is larger than it truly is. This misrepresentation inflates the apparent degrees of freedom, leading to overconfident slope estimates and potentially premature conclusions about trend significance. Sturludóttir (2015) further demonstrated that failing to account for autocorrelation can dramatically inflate type I error rates in changepoint detection. For example, with a true autocorrelation of $\rho = 0.5$ in a 50-point time series, the false positive rate increased from 10%—when autocorrelation was properly modeled—to 60% when it was ignored. These findings highlight the critical importance of properly modeling temporal dependence when designing monitoring programs and interpreting trend analyses.

The influence of different factors—such as slope strength, slope standard error, data variability, and autocorrelation—on monitoring duration reflects the distinct temporal behavior of each parameter. For example, dissolved oxygen metrics were strongly influenced by lag-1 autocorrelation, indicating that values in one year are closely linked to those in the previous year. This temporal inertia suggests that the factors affecting dissolved oxygen levels are persistent over time rather than driven by random or isolated events. Such persistence is likely due to recurring seasonal patterns and consistent environmental conditions, including thermal stratification (Keeling et al. 2010, Kwiatkowski et al. 2020) and chronic nutrient loading and accumulation (Klump et al. 2018, Hanson et al. 2023). This temporal dependency makes it harder to detect long-term trends without longer time series.

In contrast, parameters like nitrite + nitrate were more influenced by slope standard error, suggesting that while interannual values may fluctuate more widely, the trend itself follows a clear, consistent trend, thus allowing for shorter monitoring periods. Nitrogen compounds such as nitrite and nitrate typically exhibit short-term variability due to factors like seasonal biological uptake, rainfall-driven runoff, and land use activity within the watershed (Hubertz and Cahoon 1999, Chen et al. 2005, Jani and Toor 2018). For example, concentrations often spike after storm events that transport fertilizers and other nitrogen sources into the estuary, but decline during periods of intense phytoplankton growth, when nitrogen is assimilated to support biomass production (Chen et al. 2005, Glibert and Burkholder 2006). Given that nitrite + nitrate concentrations follow a relatively consistent long-term trajectory despite high short-term variability, slope standard error may be a more relevant indicator of trend detectability—as capturing the reliability of the trend direction becomes more important than accounting for raw variability alone.

Suspended solids, on the other hand, were primarily influenced by overall data variability (i.e., standard deviation), rather than by temporal structure or trend strength. This parameter exhibited high short-term fluctuations, likely driven by episodic events such as stormwater runoff, tidal resuspension, or construction-related sediment inputs (Burton and Johnston 2010, Corbett 2010, Phlips et al. 2020). Despite this variability, suspended solids showed a clear and steady increase over time, likely reflecting ongoing anthropogenic inputs to the estuary. Similar to nitrite + nitrate, its high variability did not obscure the underlying trend, which enabled early detection of long-term change.

While water temperature was influenced by both data standard deviation and autocorrelation, this reflected a combination of high short-term variability and persistent long-term patterns. Fluctuations can be driven by seasonal cycles, tidal changes, storm-driven freshwater inflows, and extreme events such as marine heatwaves and cold snaps (Shi et al. 2024). Despite this variability, temperature trends change slowly, exhibiting high temporal autocorrelation due to the large heat capacity of water and the gradual influence of climate warming, both of which promote continuity across years (Lefcheck et al. 2017). As a result, detecting long-term temperature trends requires extended monitoring periods, even in the presence of frequent short-term fluctuations.

Spatial Monitoring

Similar to our results on temporal sampling requirements, spatial optimization analyses revealed that different environmental parameters require different numbers of sites for accurate trend detection. While all 10 sites are needed at the 100% representativeness threshold, 8–9 sites are sufficient when an 80% threshold is acceptable. Dissolved oxygen concentration, for instance, could be effectively monitored with only 8 sites, and Adams Point and Oyster River were identified as the most redundant. Here, redundant means that removing the sites did not have a major effect on the overall temporal trend. For other parameters, including nitrite + nitrate and suspended solids, Adams Point again appeared among the least essential, reinforcing the idea that certain sites may be overrepresented in current monitoring efforts.

To understand what drives these differences in minimum site requirements across parameters, we conducted a variance partitioning analysis. Lag-1 autocorrelation, slope standard error, and data standard deviation each contributed almost equally to explaining the number of sites needed. This even distribution suggests that no single statistical property drives spatial sampling needs. Rather, a combination of temporal structure, trend precision, and overall variability shapes how many sites are required to achieve trend representativeness. This outcome is expected, as the minimum site analysis is based on the minimum time algorithm and treats environmental parameters as replicates. As such, it reflects both shared and unique statistical characteristics across parameters.

Patterns of site uniqueness—measured by the influence of individual sites on estuary-wide trend estimates—also varied by parameter and were shaped by ecological dynamics. For parameters like dissolved oxygen saturation, nitrite + nitrate, and water temperature, the majority of sites exhibited negative z-scores, suggesting that site-level trends closely mirrored the overall estuarine trend. This consistency may reflect parameters that are more uniformly influenced by system-wide drivers such as seasonal cycles and broad-scale hydrodynamics, rather than localized conditions. For example, DO saturation is a temperature-corrected metric and thus tends to reflect broader oxygen availability patterns, while water temperature itself is strongly governed by seasonal insolation and mixing. Nitrite + nitrate, while subject to short-term fluctuations, may show uniform long-term declines across sites due to coordinated reductions in upstream nutrient loading and estuary-wide uptake by primary producers.

In contrast, dissolved oxygen concentration and suspended solids showed more site-specific variability, with more positive z-scores, indicating that removing individual sites more often altered the estuary-wide trend. This localized behavior may reflect site-specific biological activity (e.g., eelgrass photosynthesis, respiration, and organic matter degradation) and differences in physical conditions, such as depth, mixing, and sediment resuspension. Suspended solids, in particular, are highly sensitive to localized inputs and disturbances, including river discharge, stormwater runoff, and proximity to vegetated habitats like eelgrass beds, which promote sediment settling.

Two sites—Route 9 Bridge and Squamscott River—stood out as particularly unique, consistently showing the highest positive z-scores (ranging from 1.6 to 2.7 across parameters), suggesting that their trends diverged significantly from the system-wide average. This divergence is likely due to their location in highly developed subwatersheds, where impervious surfaces, stormwater runoff, and altered flow regimes increase variability in parameters like nutrients and suspended sediments. In contrast, Adams Point often had the most negative z-scores (ranging from –0.19 to –0.86), suggesting that it may serve as a reasonable proxy for estuary-wide trends—likely due to its central location at a hydrodynamic chokepoint.

Implications for Monitoring Programs

Our findings highlight the potential value of adaptive monitoring designs that align sampling effort with the statistical and ecological characteristics of each parameter. By identifying parameters and sites that require less intensive monitoring, programs can reallocate resources toward more variable or less predictable parameters, enhancing efficiency without compromising scientific value.

Monitoring additional environmental parameters typically does not increase field effort substantially, as multiple parameters are measured during the same sampling event. For example, water quality monitoring often involves collecting data on a suite of parameters simultaneously—whether through grab samples or deployed sondes. As a result, choosing not to monitor a parameter solely based on its short minimum time requirement may not lead to significant cost savings. Instead, it may be more effective to base the overall temporal monitoring strategy on the parameter that requires the longest time series to detect trends. Doing so ensures sufficient data are collected for all parameters of interest without fragmenting the sampling timeline or creating inconsistent records across variables.

In contrast, reducing the number of monitoring sites can meaningfully lower effort and cost. While adding more sites can improve statistical precision, each additional site contributes less to overall trend detection after a certain point, resulting in diminishing returns relative to the added cost (Fairweather 1991, Caughlan and Oakley 2001). Our spatial optimization results support this principle: some sites, like Adams Point, contribute little unique information and are redundant across multiple parameters. Removing such sites from routine sampling could significantly reduce personnel time, transportation costs, and logistical complexity—especially in estuarine systems where site access is resource-intensive. When redundancy spans multiple parameters, the justification for streamlining becomes even stronger.

However, statistical redundancy does not always equate to management irrelevance. In the case of Adams Point, for example, its location at a narrow constriction between Great Bay and the rest of the estuary, makes it a strategically important site for capturing water quality as it flows out of the bay during ebb tides to downstream ecosystems. Additionally, the presence of laboratory facilities nearby makes it one of the most logistically convenient sites to sample. These factors underscore that practical, ecological, or management priorities may, at times, outweigh purely statistical optimization in long-term monitoring design.

There are also opportunities to reduce costs by re-evaluating sampling methods (Hawker et al. 2022). For instance, automated data sondes have been shown to offer a cost-effective alternative to manual grab sampling in many contexts. These sondes enable high-frequency, continuous data collection with reduced labor demands and the potential for real-time data transmission, which can significantly cut operational costs and expand spatial monitoring without increasing field effort (Kumar et al. 2024, Rozemeijer et al. 2025). However, these advantages come with trade-offs. Sondes require a high initial investment and can incur substantial long-term maintenance costs, including calibration, sensor replacement, and data quality control. They are also limited in the range of parameters they can measure in situ, especially for nutrients, trace metals, toxins, and microbial contaminants, which are crucial in regulatory compliance and public health monitoring. Their readings can also be affected by optical interferences like turbidity, biofouling, or colored dissolved organic matter (Downing et al. 2012, Robinson 2024).

In contrast, grab sampling remains the gold standard for many of water quality parameters due to its high analytical precision, flexibility in accommodating a wide range of laboratory techniques, and regulatory acceptance for parameters that require confirmatory lab testing (Erickson et al. 2013, Kmush et al. 2022). Moreover, grab samples can capture complex water chemistry interactions that sondes cannot detect and provide critical context in event-based monitoring or source tracking. Therefore, a hybrid strategy—using sondes for baseline, continuous monitoring and strategic grab samples for periodic, targeted analysis—strikes an effective balance between cost-efficiency, temporal resolution, and comprehensive water quality assessment.

Our findings also suggest that different statistical drivers—particularly autocorrelation and slope standard error—have distinct implications for monitoring design. When autocorrelation strongly influences trend detection, as observed for dissolved oxygen parameters, it signals that year-to-year measurements are not fully independent. This means that simply collecting more data points over time may not increase statistical power as much as expected, especially if measurements are temporally clustered or influenced by persistent seasonal processes. To mitigate this, monitoring programs may benefit from ensuring temporal spacing that captures independent variability—for instance, sampling across seasons, hydrological conditions, or climatic regimes. Additionally, distributing sampling effort spatially across sites with distinct hydrodynamic or ecological conditions can reduce the risk of temporal autocorrelation dominating the signal.

In contrast, when slope standard error emerges as a dominant factor—as seen for nitrite + nitrate—it implies that the precision of the estimated trend is the limiting factor for detection. This suggests that improving consistency in measurement techniques, reducing sampling noise, or increasing sample size during high-variability periods could sharpen trend estimates. In such cases, attention to quality assurance, sensor calibration, and minimizing measurement error may yield greater benefits than simply extending the duration of monitoring.

Together, these insights emphasize that optimizing a monitoring design is not just about “how long” or “how often” to sample, but also about how well the data structure aligns with the characteristics of each parameter. Programs should not assume that the same sampling design will be equally effective across all parameters—instead, tailoring strategies based on underlying statistical behavior can improve efficiency and reliability in detecting meaningful ecological trends.

Overall, these findings support a flexible, data-driven approach to long-term environmental monitoring—one that balances the need for rigorous trend detection with the practical realities of staffing, funding, and logistics.

Limitations and Future Directions

This study focused on detecting linear long-term trends using annual average data—a simplification that helps identify broad directional changes over time but does not fully capture the complexity of estuarine systems (Rigal et al. 2020). Estuaries are shaped by interacting processes that vary on multiple timescales, including seasonal cycles, tidal fluctuations, storm events, and anthropogenic inputs (McLusky and Elliott 2004). Aggregating data into annual averages smooths out this intra-annual variability (White and Hastings 2020), potentially masking important ecological signals such as seasonal hypoxia, nutrient pulses, or life-stage-specific biological responses. Moreover, linear trend models assume steady, incremental change and may overlook non-linear dynamics, threshold effects, or regime shifts—phenomena that are increasingly relevant in the context of climate change and coastal development (Ellis and Post 2004, McGlathery et al. 2013). For example, Bruel and White (2021) used a resampling approach to understand the temporal data requirements to detect changepoints in biodiversity data. They showed that additional temporal sampling during periods of rapid change can be more effective than sampling in equally spaced intervals. For management questions that require an understanding of fine-scale variability—such as identifying critical habitat windows, detecting early warning signs of stress, or evaluating compliance with water quality standards—annual averages and linear models may be insufficient.

These limitations are further compounded when sampling is infrequent. Low-frequency designs may fail to capture short-term or episodic events, leading to systematic underestimation of key processes. For example, Anderson et al. (2024) found that missing storm-driven nutrient pulses in wetlands substantially underestimated nutrient export, mischaracterizing ecosystem function. In dynamic coastal systems, such underestimation could result in flawed assessments or missed signals of degradation. These findings underscore the importance of considering both sampling frequency and data resolution when designing monitoring programs, especially for parameters influenced by short-term hydrologic variability.

Additionally, while our analysis offers insight into sampling design in the Great Bay Estuary, its applicability to other systems remains an open question. Future studies should apply this framework to other estuaries with different physical, chemical, hydrodynamic, and biological characteristics to determine how broadly these findings can be generalized. Comparative analyses across multiple systems could reveal whether certain parameters or sampling strategies consistently perform well or whether design recommendations must be tailored to local conditions. Incorporating systems that vary in size, salinity gradient, watershed development, and monitoring history would also help refine guidelines for transferable and adaptive monitoring programs.

Expanding this work to include seasonal resolution, event-based sampling, or multi-system comparisons would provide a more nuanced understanding of how monitoring design can support adaptive management in dynamic coastal environments. As environmental change accelerates, designing monitoring programs that are both efficient and responsive to diverse ecological conditions will be increasingly critical.

Acknowledgements

We would like to thank the countless number of volunteers and staff that have maintained the data in Great Bay Estuary over the decades. Some funding for this project came from XXXX and YYY. We would also like to thank members of the Quantitative Marine Ecology Lab at the University of New Hampshire and ZZ anonymous reviewers for their feedback on previous versions of this manuscript.

Authors Contributions

B. Spiecker contributed to project conceptualization, data curation, formal analysis, and led the writing and editing of the manuscript. E. White contributed to conceptualization, funding acquisition, and manuscript writing and editing. K. Matso contributed to funding acquisition and manuscript review and editing. C. Whitney contributed to data curation and manuscript review and editing.

Competing Interests

This work was supported by the Piscataqua Region Estuaries Partnership (PREP), whose data were used in the study. Co-authors Barbara Spiecker and Easton White received research funding from PREP. Co-authors Kalle Matso and Chris Whitney are employed by PREP as Director and Data Scientist, respectively. PREP receives funding from the U.S. Environmental Protection Agency and also from municipalities in the Piscataqua Region Watershed.

Data Availability

The data that support the findings of this study are openly available in Piscataqua Watershed Data Explorer at http://data.prepestuaries.org:8510/. All code used in this study is available at https://github.com/QuantMarineEcoLab/prep-sampling-optimization.

Funding Statement

This work was supported by the Piscataqua Region Estuaries Partnership (PREP) with funding from the U.S. Environmental Protection Agency and the Infrastructure and Investment Jobs Act. PREP is one of 28 National Estuary Programs, created by the 1987 amendment to the Clean Water Act. PREP's role is to work with partners to protect and restore the waters of the Great Bay Estuary and the Hampton-Seabrook Estuary.

Literature Cited

Anderson, K. J., B. Adhikari, O. F. Schloegel, R. Marques Mendonca, M. P. Back, N. Brocato, J. A. Cianci-Gaskill, S. E. McMurray, C. Bahlai, D. M. Costello, and L. Kinsman-Costello. 2024. We know less about phosphorus retention in constructed wetlands than we think we do: A quantitative literature synthesis. Ecological Indicators 169:112969.

Bahlai, C. A., E. R. White, J. D. Perrone, S. Cusser, and K. Stack Whitney. 2021. The broken window: An algorithm for quantifying and characterizing misleading trajectories in ecological processes. Ecological Informatics 64:101336.

Barry, J., D. Maxwell, S. Jennings, D. Walker, and J. Murray. 2017. Emon: an R-package to support the design of marine ecological and environmental studies, surveys and monitoring programmes. Methods in Ecology and Evolution 8:1342–1346.

Bashevkin, S. M. 2022. A Framework for Evaluating the Effects of Reduced Spatial or Temporal Monitoring Effort. San Francisco Estuary and Watershed Science 20.

Boynton, W. R., M. a. C. Ceballos, C. L. S. Hodgkins, D. Liang, and J. M. Testa. 2022. Large-Scale Spatial and Temporal Patterns and Importance of Sediment–Water Oxygen and Nutrient Fluxes in the Chesapeake Bay Region. Estuaries and Coasts 46:356–375.

Bruel, R., and E. R. White. 2021. Sampling requirements and approaches to detect ecosystem shifts. Ecological Indicators 121:107096.

Burton, G. A., and E. L. Johnston. 2010. Assessing contaminated sediments in the context of multiple stressors | Environmental Toxicology and Chemistry | Oxford Academic 29:2625–2643.

Caughlan, L. 2001. Cost considerations for long-term ecological monitoring. Ecological Indicators 1:123–134.

Caughlan, L., and K. L. Oakley. 2001. Cost considerations for long-term ecological monitoring. Ecological Indicators 1:123–134.

Chen, L., H. Peng, B.-J. Fu, J. Qiu, and S. Zhang. 2005. Seasonal variation of nitrogen-concentration in the surface water and its relationship with land use in a catchment of northern China. Journal of environmental sciences (China) 17:224–231.

Cloern, J. E., and A. D. Jassby. 2012. Drivers of change in estuarine-coastal ecosystems: Discoveries from four decades of study in San Francisco Bay. Reviews of Geophysics 50.

Cohen, J. 1992. A power primer. Psychological Bulletin 112:155–159.

Cook, S. E., T. C. Lippmann, and J. D. Irish. 2019. Modeling nonlinear tidal evolution in an energetic estuary. Ocean Modelling 136:13–27.

Corbett, D. R. 2010. Resuspension and estuarine nutrient cycling: insights from the Neuse River Estuary 7.

Downing, B. D., B. A. Pellerin, B. A. Bergamaschi, J. F. Saraceno, and T. E. C. Kraus. 2012. Seeing the light: The effects of particles, dissolved materials, and temperature on in situ measurements of DOM fluorescence in rivers and streams. Limnology and Oceanography: Methods 10:767–775.

Ellis, A. M., and E. Post. 2004. Population response to climate change: linear vs. non-linear modeling approaches. BMC Ecology 4:2.

Erickson, A. J., P. T. Weiss, and J. S. Gulliver. 2013. Water Sampling Methods. Page Optimizing Stormwater Treatment Practices: A Handbook of Assessment and Maintenance. Springer, New York, NY.

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring 42:555–567.

Field, S. A., P. J. O. Connor, A. J. Tyre, and H. P. Possingham. 2007. Making monitoring meaningful. Austral Ecology 32:485–491.

Fletcher Jr., R. J., T. J. Hefley, E. P. Robertson, B. Zuckerberg, R. A. McCleery, and R. M. Dorazio. 2019. A practical guide for combining data to model species distributions. Ecology 100:e02710.

Giron-Nava, A., C. C. James, A. F. Johnson, D. Dannecker, B. Kolody, A. Lee, M. Nagarkar, G. M. Pao, H. Ye, D. G. Johns, and G. Sugihara. 2017. Quantitative argument for long-term ecological monitoring. Marine Ecology Progress Series 572:269–274.

Glibert, P. M., and J. M. Burkholder. 2006. The Complex Relationships Between Increases in Fertilization of the Earth, Coastal Eutrophication and Proliferation of Harmful Algal Blooms. Pages 341–354 *in* E. Granéli and J. T. Turner, editors. Ecology of Harmful Algae. Springer Berlin Heidelberg.

Hanson, P. C., R. Ladwig, C. Buelo, E. A. Albright, A. D. Delany, and C. C. Carey. 2023. Legacy Phosphorus and Ecosystem Memory Control Future Water Quality in a Eutrophic Lake. Journal of Geophysical Research: Biogeosciences 128:e2023JG007620.

Hawker, D. W., J. Clokey, S. G. Gorji, R. Verhagen, and S. L. Kaserzon. 2022. Monitoring techniques–Grab and passive sampling. Pages 25–48 *in* T. Dalu and N. T. Tavengwa, editors. Emerging Freshwater Pollutants. Elsevier.

Hubertz, E. D., and L. B. Cahoon. 1999. Short-term variability of water quality parameters in two shallow estuaries of North Carolina. Estuaries 22:814–823.

Jani, J., and G. S. Toor. 2018. Composition, sources, and bioavailability of nitrogen in a longitudinal gradient from freshwater to estuarine waters. Water Research 137:344–354.

Jones, S. 2000. A Technical Characterization of Estuarine and Coastal New Hampshire. New Hampshire Estuaries Project.

Kalra, T. S., N. K. Ganju, and J. M. Testa. 2020. Development of a submerged aquatic vegetation growth model in the Coupled Ocean–Atmosphere–Wave–Sediment Transport (COAWST v3.4) model.

Keeling, R. F., A. Körtzinger, and N. Gruber. 2010. Ocean Deoxygenation in a Warming World.

Kennish, M. J. 2019. The National Estuarine Research Reserve System: A Review of Research and Monitoring Initiatives. Open Journal of Ecology 09:50.

Klump, J. V., S. L. Brunner, B. K. Grunert, J. L. Kaster, K. Weckerly, E. M. Houghton, J. A. Kennedy, and T. J. Valenta. 2018. Evidence of persistent, recurring summertime hypoxia in Green Bay, Lake Michigan. Journal of Great Lakes Research 44:841–850.

Kmush, B. L., D. Monk, H. Green, D. A. Sachs†, T. Zeng, and D. A. Larsen. 2022. Comparability of 24-hour composite and grab samples for detection of SARS-2-CoV RNA in wastewater. FEMS Microbes 3:xtac017.

Kumar, M., K. Khamis, R. Stevens, D. M. Hannah, and C. Bradley. 2024. In-situ optical water quality monitoring sensors—applications, challenges, and future opportunities 6.

Kwiatkowski, L., O. Torres, L. Bopp, O. Aumont, M. Chamberlain, J. R. Christian, J. P. Dunne, M. Gehlen, T. Ilyina, J. G. John, A. Lenton, H. Li, N. S. Lovenduski, J. C. Orr, J. Palmieri, Y. Santana-Falcón, J. Schwinger, R. Séférian, C. A. Stock, A. Tagliabue, Y. Takano, J. Tjiputra, K. Toyama, H. Tsujino, M. Watanabe, A. Yamamoto, A. Yool, and T. Ziehn. 2020. Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections 17.

Lefcheck, J. S., D. J. Wilcox, R. R. Murphy, S. R. Marion, and R. J. Orth. 2017. Multiple stressors threaten the imperiled coastal foundation species eelgrass (Zostera marina) in Chesapeake Bay, USA. Global Change Biology 23:3474–3483.

Legendre, P., M. R. T. Dale, M.-J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. Ecography 25:601–615.

Legg, C. J., and L. Nagy. 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management 78:194–199.

Legg, J., M. Ndalahwa, J. Yabeja, I. Ndyetabula, H. Bouwmeester, R. Shirima, and K. Mtunda. 2017. Community phytosanitation to manage cassava brown streak disease. Virus Research 241:236–253.

Lindenmayer, D. B., and G. E. Likens. 2010. The science and application of ecological monitoring. Biological Conservation 143:1317–1328.

Maguran, A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. Somerfield, and A. D. Watt. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology and Evolution 25:574–582.

McGlathery, K. J., M. A. Reidenbach, P. D’Odorico, S. Fagherazzi, M. L. Pace, and J. H. Porter. 2013. Nonlinear Dynamics and Alternative Stable States in Shallow Coastal Systems. Oceanography 26:220–231.

McLusky, D. S., and M. Elliott. 2004. The Estuarine Ecosystem: Ecology, Threats and Management. OUP Oxford.

Morant, J., J. A. González-Oreja, J. E. Martínez, P. López-López, and I. Zuberogoitia. 2020. Applying economic and ecological criteria to design cost-effective monitoring for elusive species. Ecological Indicators 115:106366.

Mulukutla, G. K., W. M. Wollheim, J. E. Salisbury, R. O. Carey, T. K. Gregory, and W. H. McDowell. 2021. High-Frequency Concurrent Measurements in Watershed and Impaired Estuary Reveal Coupled DOC and Decoupled Nitrate Dynamics. Estuaries and Coasts 45:445–461.

Phlips, E. J., S. Badylak, N. G. Nelson, and K. E. Havens. 2020. Hurricanes, El Niño and harmful algal blooms in two sub-tropical Florida estuaries: Direct and indirect impacts. Scientific Reports 10:1910.

Possingham, H. P., B. A. Wintle, R. A. Fuller, and L. N. Joseph. 2012. The conservation return on investment from ecological monitoring. Biodiversity monitoring in Australia:49–58.

PREP. 2023a. 2023 State of Our Estuaries. Piscataqua Region Estuaries Partnership.

PREP. 2023b. 2023 SOOE Extended Version. Piscataqua Region Estuaries Partnership.

Regular, P. M., G. J. Robertson, K. P. Lewis, J. Babyn, B. Healey, and F. Mowbray. 2020. SimSurvey: An R package for comparing the design and analysis of surveys by simulating spatially-correlated populations. PLOS ONE 15:e0232822.

Rhodes, J. R., and N. Jonzen. 2011. Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time? Ecography 34:1040–1048.

Rigal, S., V. Devictor, and V. Dakos. 2020. A method for classifying and comparing non-linear trajectories of ecological variables. Ecological Indicators 112:106113.

Robinson, R. L. 2024. Investigating Biofouling on Long-Term, In Situ, Optic Sensors: Impacts on Optical Measurement Integrity and Insight Into Community Dynamics. M.Sc., University of Windsor, Ontario, Canada.

Rozemeijer, J., P. Jordan, A. Hooijboer, B. Kronvang, M. Glendell, R. Hensley, K. Rinke, M. Stutter, M. Bieroza, R. Turner, P. E. Mellander, P. Thorburn, R. Cassidy, J. Appels, K. Ouwerkerk, and M. Rode. 2025. Best practice in high-frequency water quality monitoring for improved management and assessment; a novel decision workflow. Environmental Monitoring and Assessment 197:1–23.

Shi, J., C. Hu, and E. Stabenau. 2024. Temperature Response of South Florida Estuaries to the 2023 Heatwave. Estuaries and Coasts 47:1388–1401.

Sturludóttir, E. 2015, May 6. Statistical analysis of trends in data from ecological monitoring. Thesis.

White, E. R. 2019. Minimum time required to detect population trends: the need for long-term monitoring programs. BioScience 69:40–46.

White, E. R., and C. A. Bahlai. 2021. Experimenting With the Past to Improve Environmental Monitoring. Frontiers in Ecology and Evolution 8:572979.

White, E. R., and A. Hastings. 2020. Seasonality in ecology: Progress and prospects in theory. Ecological Complexity 44:100867.

Wolfe, D. A., M. A. Champ, D. A. Flemer, and A. J. Mearns. 1987. Long-term biological data sets: Their role in research, monitoring, and management of estuarine and coastal marine systems. Estuaries 10:181–193.

Tables

Table 1. Minimum time required to monitor each environmental variable across sites (out of 22 years). *Minimum Years (80% correct)* corresponds to the minimum number of years with 80% of sub-sample slopes matching the long-term slope and *Minimum Years (100% correct)* corresponds to the minimum number of years with a 100% match.

Variable	Minimum Years (80% correct)	Minimum Years (100% correct)
Dissolved Oxygen Saturation (%)	6	7
Dissolved Oxygen (mg/L)	6	7
Nitrite + Nitrate, dissolved (mg/L)	5	5
Suspended Solids (mg/L)	6	6
Water Temperature (°C)	6	7

Table 2. Percentage of variance explained by five variance metrics in relation to the minimum number of years required to detect trends for each environmental parameter. The variance metrics include: **slope** (overall trend magnitude), **slope standard error** (SE; precision of the trend estimate), **data standard deviation** (SD, variability in observed values), **lag-1 autocorrelation** (temporal dependence across years), and **residuals** (unexplained variation from the trend model). The two rows labeled "**Minimum Years (80%)**" and "**Minimum Years (100%)**" indicate the number of years required to achieve 80% and 100% trend representativeness, respectively. The final row shows the **coefficient of variation**, which provides a normalized measure of variability relative to the mean.

Metric	Dissolved Oxygen (mg/L)	Dissolved Oxygen Saturation (%)	Water Temperature (°C)	Nitrite + Nitrate, dissolved (mg/L)	Suspended Solids (mg/L)
Slope	10.7%	0.7%	0.2%	0.1%	10.5%
Slope SE	1.2%	35.8%	3.9%	22.6%	9.9%
Data SD	7%	1.8%	21.2%	5.5%	34%
Autocorrelation	71.5%	51%	24.3%	26.7%	5%
Residuals	9.6%	10.7%	50.4%	45.2%	40.6%
Minimum Years (80%)	**6**	**6**	**6**	**5**	**6**
Minimum Years (100%)	**7**	**7**	**7**	**5**	**6**
Coefficient of Variation	11.6%	9.4%	12.6%	101.6%	112.6%

Table 3. Minimum number of sites required to monitor each environmental variable across years (out of 10 sites). *Minimum Sites (80% correct)* corresponds to the minimum number of sites with 80% of sub-sample slopes matching the long-term slope and *Minimum Sites (100% correct)* corresponds to the minimum number of sites with a 100% match.

Variable	Minimum Sites (80% correct)	Minimum Sites (100% correct)
Dissolved Oxygen Saturation (%)	9	10
Dissolved Oxygen (mg/L)	8	9
Nitrite + Nitrate, dissolved (mg/L)	9	10
Suspended Solids (mg/L)	9	10
Water Temperature (°C)	9	10

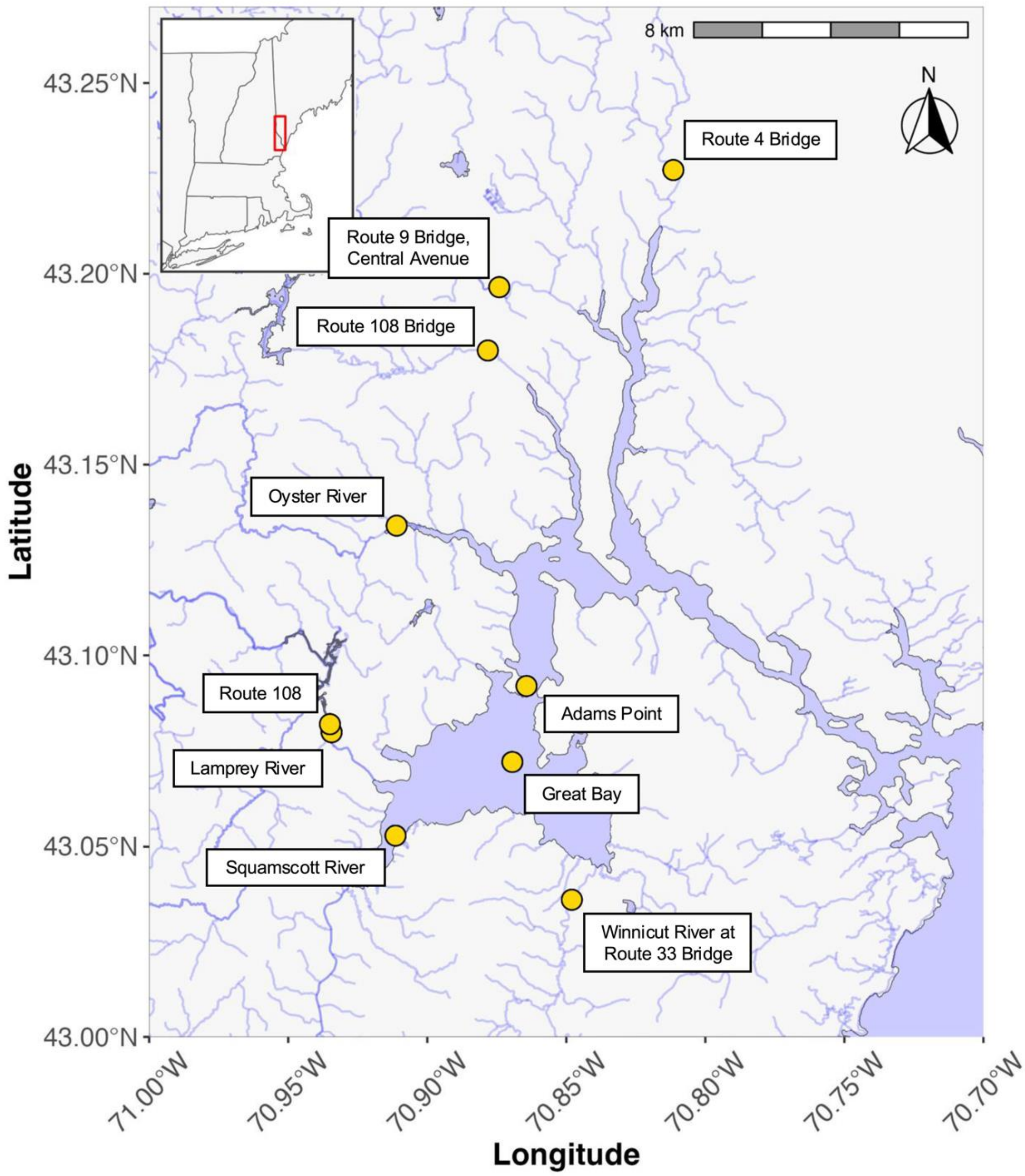
Table 4. Least unique site combinations at the 80% trend representativeness threshold for each environmental parameter. The table identifies the combination with the smallest slope difference relative to the full dataset. It includes the minimum number of sites needed, the total number of sampled sites, the site(s) removed (or deemed unnecessary when that minimum is used), the resulting subsample slope compared to the full-sample slope, and the absolute difference between these slopes.

Variable	Minimum # of Sites	Total Sampled Sites	Sites Removed	Subsample Slope	Full Sample Slope	Slope Difference
Dissolved Oxygen Saturation (%)	9	10	Route 9 Bridge (Central Avenue)	-0.012994	-0.012705	0.000289
Dissolved Oxygen (mg/L)	8	10	Adams Point, Oyster River	-0.011696	-0.011692	0.000004
Nitrite + Nitrate, dissolved (mg/L)	9	10	Adams Point	-0.002673	-0.002629	0.000045
Suspended Solids (mg/L)	9	10	Adams Point	0.023013	0.022327	0.000686
Water Temperature (°C)	9	10	Squamscott River	-0.002736	-0.001952	0.000785

Table 5. Most and least unique sites for each environmental parameter. The most unique site was identified as the one whose removal resulted in the largest difference from the full-model trend, indicating that it contributed the most unique information to the overall signal. Conversely, the least unique site exhibited the smallest difference when removed, suggesting strong agreement with the full-model trend and limited additional contribution beyond what was captured by other sites.

Variable	Most Useful Site	Least Useful Site
Dissolved Oxygen Saturation (%)	Squamscott River	Route 9 Bridge (Central Avenue)
Dissolved Oxygen (mg/L)	Route 9 Bridge (Central Avenue)	Lamprey River
Nitrite + Nitrate, dissolved (mg/L)	Route 9 Bridge (Central Avenue)	Adams Point
Suspended Solids (mg/L)	Squamscott River	Adams Point
Water Temperature (°C)	Route 9 Bridge (Central Avenue)	Squamscott River

751 # Figures



752

753 **Figure 1. Map of sampling locations at the Great Bay Estuary, New Hampshire, USA.**

754

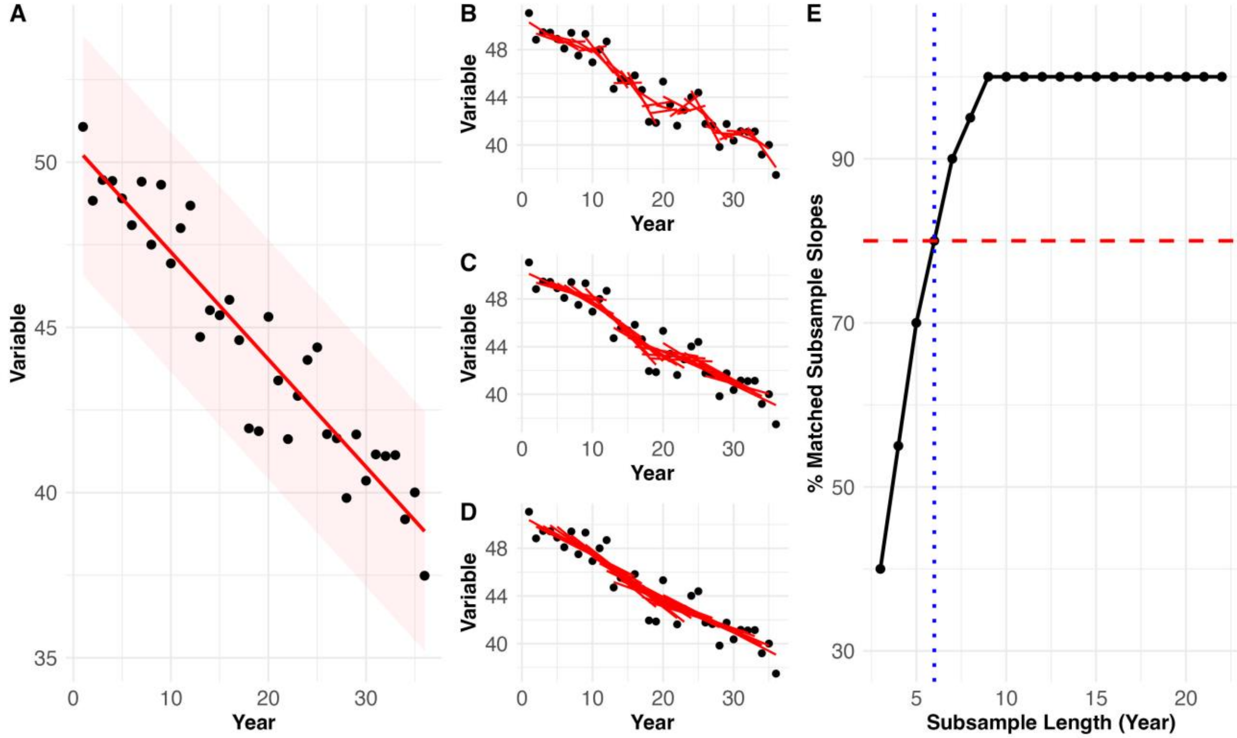


Figure 2. Illustration of the Broken Window algorithm used to determine the minimum time required for trend detection. **Panel A** shows the full time series with a linear regression line representing the “complete” slope and a shaded band indicating ±1 standard deviation (SD). **Panels B–D** illustrate the variability in trend estimates across all possible moving windows of 5 years **(B)**, 10 years **(C)**, and 15 years **(D)**, with each red line representing a linear model fit to one subsample. **Panel E** shows the percentage of subsample slopes that fall within ±1 SD of the complete slope across increasing window lengths. In this example, the intersection of the dashed horizontal (red) and vertical (blue) lines identifies the minimum number of years required to detect a trend with 80% statistical power.

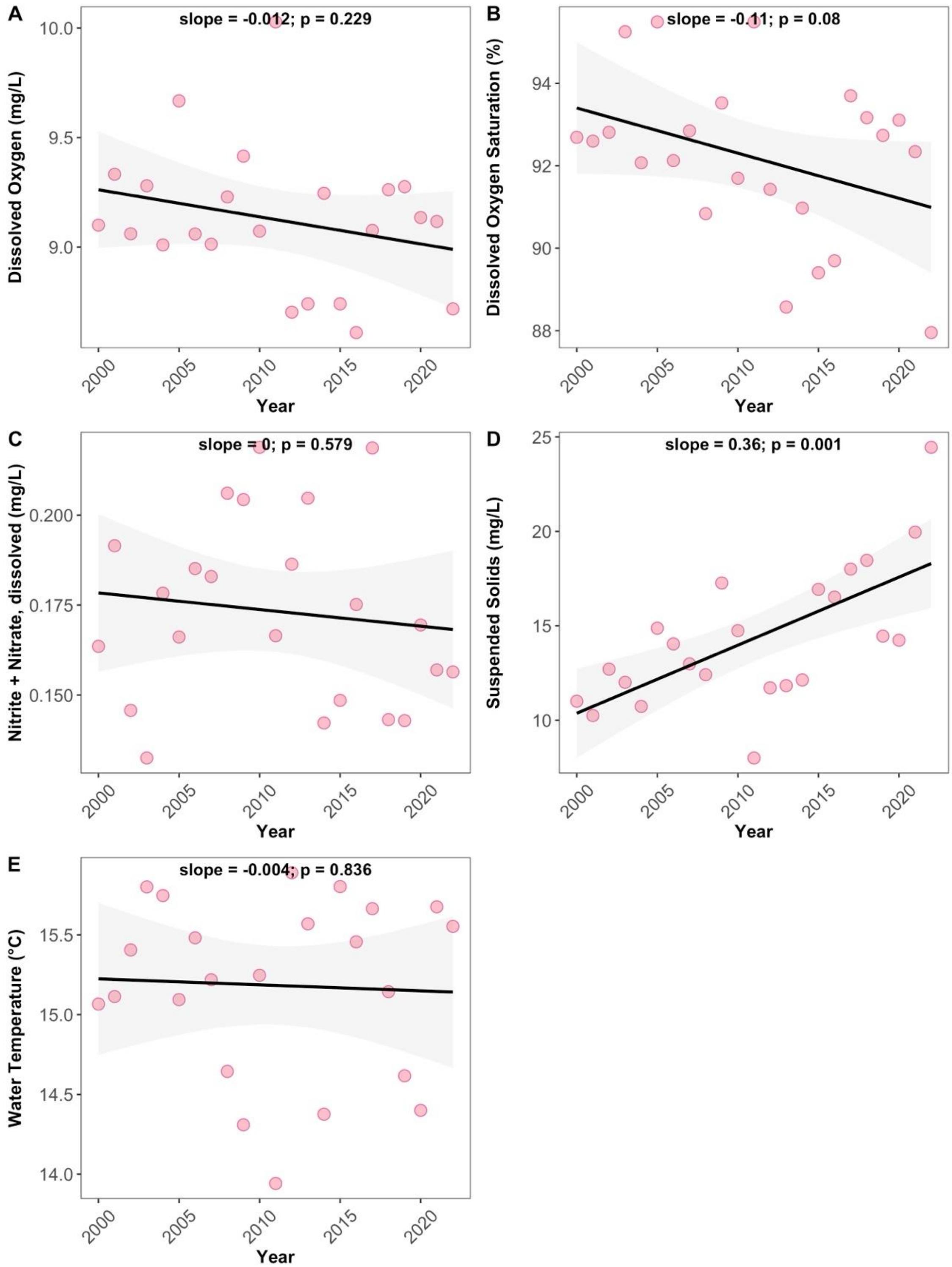


Figure 3. Long-term time series for each environmental parameter averaged across sites. Slopes are unstandardized, with units corresponding to each parameter. **(A)** Dissolved Oxygen

Concentration (mg/L). **(B)** Dissolved Oxygen Saturation (%). **(C)** Dissolved Nitrite + Nitrate (mg/L). **(D)** Suspended Solids (mg/L). **(E)** Water Temperature (°C).

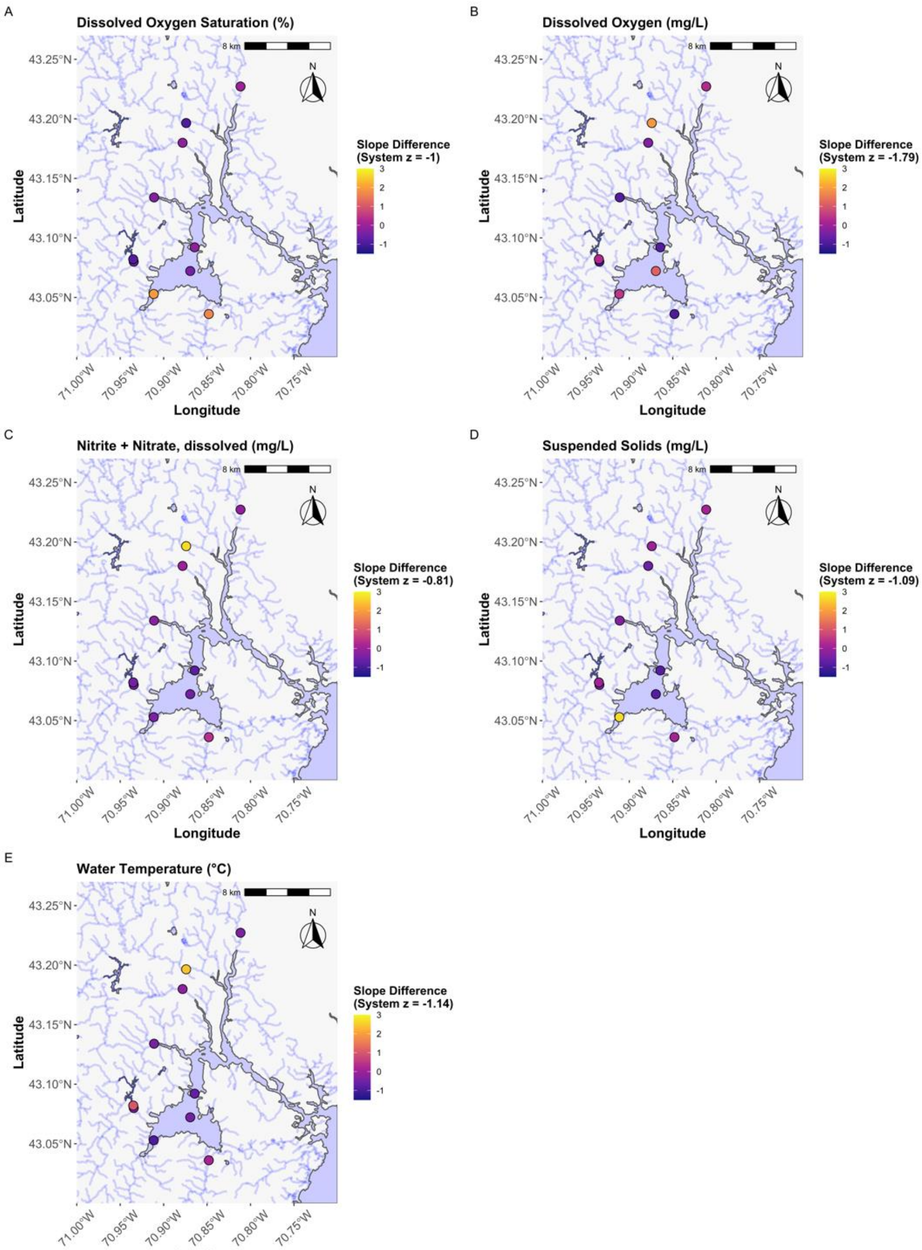


Figure 4. Spatial heterogeneity in site-level temporal trends for each environmental parameter. These values represent relative deviation in site-level trends from estuary-wide behavior, represented as z-score standardized slope differences. “System z” denotes the system-wide minimum possible z-score—i.e., the standardized slope difference corresponding to a perfect match with the full model trend (slope difference = 0). Positive *z*-scores indicate greater

divergence from the estuary-wide trend, while negative values reflect stronger alignment with overall system behavior. **(A)** Dissolved Oxygen Concentration (mg/L). **(B)** Dissolved Oxygen Saturation (%). **(C)** Dissolved Nitrite + Nitrate (mg/L). **(D)** Suspended Solids (mg/L). **(E)** Water Temperature (°C).