Modeling the combined effects of ocean warming and ocean acidification on *Zostera marina* meadows in Massachusetts

Julia Wepppler

Modeling the combined effects of ocean warming and ocean acidification on *Zostera marina* meadows in Massachusetts

Julia Weppler

Advisor: Jesse Swann-Quinn, Ph.D.

Abstract

Climate change poses significant threats to seagrass ecosystems, predominantly *Zostera marina* on the east coast of the United States, through rising ocean temperatures and shifting carbonate chemistry. This study uses an extreme gradient boosting (XGBoost) model to predict changes in *Z. marina* bed area under combined ocean warming (OW) and ocean acidification (OA) scenarios along the Massachusetts coast. A multi-source time series was compiled of *Z. marina* coverage, sea surface temperature, and pH to train the model. XGBoost was selected for its ability to capture complex nonlinear relationships and interactions in ecological data (Zhang et al, 2019; Ahn et al, 2023). The model achieved a reasonable fit, with a mean absolute error (MAE) of ~28% of annual area change, and R^2 of approximately 0.73. The model predicts a modest annual decline (a range of 0.9–3% per year) in *Z. marina* area under current warming and acidification trends in the Massachusetts coasts. Temperature-related features, such as annual maximum and mean sea surface temperature (SST) emerged as the strongest predictors of *Z. marina* decline, whereas acidity (pH) had a comparatively minor influence. These findings align with experimental studies showing that while elevated dissolved CO_2 can enhance seagrass growth, its benefits may be outweighed by thermal stressors (Collier et al., 2018; Egea et al., 2018). *Z. marina* meadows provide critical ecological services through blue carbon sequestration, habitat for fisheries and coastal protection; their climate-driven losses carry ecological and economic consequences for blue economies (Zarnoch et al., 2017). The findings of this study suggest that ocean warming is likely to be a dominant stressor on *Z. marina* in the coming decades, and management efforts should prioritize mitigating heat stress and improving water quality. The study also demonstrates the utility of XGBoost in environmental modeling, offering an effective tool for predicting ecosystem responses to multiple stressors and informing conservation strategies, even in instances with limited data availability.

Introduction

Seagrass meadows are among the most productive and valuable coastal ecosystems due to their role in blue economies and habitat provisioning. *Zostera marina*, the dominant seagrass in temperate North Atlantic and Pacific waters, provides a suite of ecological functions that support coastal biodiversity and economies. *Z. marina* beds serve as nursery and foraging habitat for countless fish and invertebrate species, stabilize sediments and shorelines, improve water clarity, and sequester carbon in biomass and sediments (Zarnoch et al., 2017). These services translate into substantial economic value: Zarnoch et al. found that the presence of *Z. marina* can enhance commercial fishery yields up to $178,000 per hectare per year in some regions, and globally *Z. marina* carbon sequestration is valued at billions of dollars per year (2017). Because of these benefits, *Z. marina* meadows are often highlighted in blue economy initiatives and coastal management plans as natural assets that contribute to sustainable fisheries, tourism, and climate change mitigation.

Despite their importance, *Z. marina* meadows have experienced widespread declines in recent decades. Global seagrass coverage is estimated to be shrinking by approximately 7% per year since 1990, a loss rate comparable to coral reefs and tropical rainforests (Waycott et al., 2009). By the early 2000s, at least 29% of historical seagrass area had been lost worldwide (Waycott et al., 2009). Multiple *Zostera* species are now considered threatened—about 15% of all seagrass species are at risk of extinction according to the IUCN (Short et al., 2011). These declines have been attributed to a range of anthropogenic stressors. Locally, coastal development and pollution have degraded water quality, as nutrient over-enrichment fuels algal blooms that reduce water clarity, and sediment runoff smothers seagrass, inhibiting the light availability essential for *Z. marina* growth (Zarnoch et al., 2017). Boat propeller scarring, dredging, and other physical disturbances have also damaged many *Z. marina* meadows (Barañano et al., 2022). Additionally, disease outbreaks have caused periodic die-offs historically, and the introduction of invasive herbivores, such as the European green crab, has exacerbated losses by uprooting *Z. marina* beds (Matheson et al., 2016).

Among the emerging drivers of *Z. marina* decline are the global-scale stressors associated with climate change: ocean warming (OW) and ocean acidification (OA). Rapid increases in sea surface temperature (SST), especially in coastal zones and semi-enclosed seas, are pushing *Z. marina* populations closer to their thermal tolerance limits (Yan et al., 2024). *Z. marina* is a cold-temperate species adapted to cool waters; when summer temperatures exceed roughly 25–30 °C for extended periods, the species experiences reduced photosynthetic capacity, respiration increases, and heat-induced wasting and mortality (Collier et al., 2018). Marine heatwaves have already caused significant *Z. marina* losses—for instance, heat stress combined with a green crab population explosion led to a >20% decline of *Z. marina* in New Hampshire's Great Bay Estuary in 2013 (Short, 2014). Over longer timescales, sustained warming trends can inhibit the recovery of *Z. marina* beds, leading to progressive decline. In Massachusetts, a

12-year mapping study from 1995–2007 found that *Z. marina* area in many embayments was diminishing by as much as 3–5% per year (Costello et al., 2011). Subsequent surveys indicate that species coverage has continued to decline into the 2010s in parts of Massachusetts (Novak et al., 2023), signalling an ongoing downward trajectory.

Ocean acidification, in contrast, has a more complex influence on *Z. marina*. As atmospheric CO2 concentrations rise, more CO2 dissolves into seawater, lowering pH and altering carbonate chemistry. Because seagrasses are photosynthetic autotrophs that can utilize dissolved CO2 and bicarbonate, elevated CO2 can act as a fertilizer, increasing capacity for photosynthesis and growth (Egea et al., 2018). Research has shown that *Z. marina* can experience increased carbon fixation and shoot density under elevated CO2 (lower pH) conditions, especially when light and nutrients are not limiting (Egea et al., 2018). For example, Zimmerman et al. (1997) demonstrated that tripling CO2 availability in the water tripled *Z. marina* photosynthetic rates and significantly reduced the light requirement for net positive growth, suggesting that rising CO2 could improve the species' productivity in eutrophic, low-light environments. More recent studies have reinforced aspects of this finding, as Yan et al. (2024) found that *Z. marina* exhibits upregulated pathways for carbon assimilation under elevated CO2 when temperatures remain in a stable range for the region. These outcomes imply that OA could partially offset some negative stressors by boosting *Z. marina* growth and resilience via "CO2 fertilization." Some have even proposed that seagrass meadows themselves can locally mitigate acidification by absorbing CO2 and raising pH during photosynthesis, providing protection to acidification-vulnerable species (Abe, 2022).

However, the net effect of climate change on *Z. marina* cannot be understood via observations in warming and acidifying environments in isolation, and the interaction of OW and OA is critical. There is growing evidence that the positive effects of CO2 enrichment on the species may not fully materialize under stressful warming scenarios. Collier et al. reported that "thermal stress and local pressures outweigh the positive effects of ocean acidification" on seagrass health in tropical environments (2018). Similarly, Yan et al. noted that the beneficial metabolic responses of *Z. marina* to OA were temperature dependent; under combined acidic and high-temperature conditions, the species showed different stress responses and the capacity for increased metabolism with additional CO2 was reduced (2024). Extreme heat can induce physiological damage that additional carbon cannot compensate for, and in such cases, warming may have a greater influence on overall bed health and coverage. The question emerges of how *Z. marina* meadows in Massachusetts will respond to the combined influence of ocean warming and acidification, which are occurring simultaneously in nature. Understanding this combined effect is important for projecting *Z. marina* persistence under future climate scenarios and for guiding conservation efforts.

To address this question, a data-driven modeling approach was used to analyze historical *Z. marina* bed changes along the coast of Massachusetts in relation to observed trends in

temperature and pH. Specifically, this research applies an extreme gradient boosting model (XGBoost) to predict changes in *Z. marina* areal coverage as a function of ocean warming and acidification indicators. XGBoost is a machine learning algorithm that has gained popularity in environmental and ecological modeling for its high predictive accuracy and ability to handle complex interactions in data (NVIDIA, 2025; Ahn et al., 2023). Gradient boosting methods like XGBoost build an ensemble of decision trees in sequence, where each new tree corrects errors made by the previous ensemble, yielding a strong predictive model that can capture nonlinear relationships (NVIDIA, 2025), as depicted in Figure 1. XGBoost's implementation is advantageous when working with smaller datasets, with regularization techniques to prevent overfitting, and efficient handling of missing or sparse data (NVIDIA, 2025). Recent studies in marine ecology have successfully used XGBoost and related ensemble methods for harmful algal bloom forecasting, water quality predictions, and species distribution modeling (Ahn et al., 2023; Zimmerman et al., 2015).
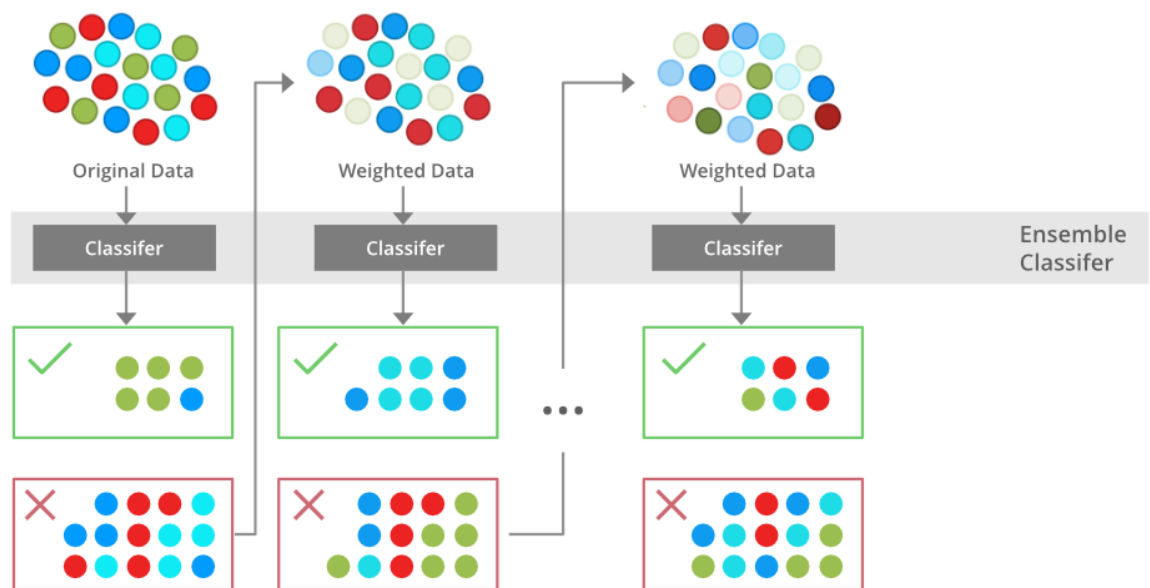

Figure 1: An overview of the training process for extreme gradient boosting (Vema, 2022)

In this paper, a novel approach to modeling impacts of anthropogenic climate change on seagrass habitats is proposed and implemented, with a preliminary analysis to examine the effectiveness and accuracy of modeling *Z. marina* trends under OW and OA. This research aims to combine empirical observations with predictive modeling to better understand how the opposing forces of warming (a stressor) and acidification (a potential resource) might jointly influence *Z. marina* in Massachusetts. The findings provide further insights on climate impacts to a critical coastal habitat and demonstrate the utility of machine learning algorithms in ecological prediction and management.

Methods

Study Area and Data Sources

The focus of this study is the Massachusetts coastline, which harbors extensive *Z. marina* meadows, from the Salem Harbor in northern Massachusetts to the slightly warmer embayments of Cape Cod and the islands in the south. Massachusetts has conducted periodic *Z. marina* mapping since the 1990s (Collier et al., 2011; Commonwealth of Massachusetts, 2025) that documents the spatial extent of *Z. marina* beds in periodic sampling. However, no single continuous time series dataset of *Z. marina* area, temperature, and pH existed for this region. Datapoints were compiled from multiple sources to create a unified dataset for modeling. *Z. marina* coverage (hectares) for various coastal segments and years was obtained from Massachusetts Department of Environmental Protection mapping surveys and academic studies (Collier et al., 2011; Commonwealth of Massachusetts, 2025), as outlined in Table 1. These datapoints provided estimates of *Z. marina* bed area over time at several locations along the Massachusetts coast. SST records were retrieved from the National Oceanic and Atmospheric Administration's National Data Buoyancy Center (NDBC) and regional monitoring programs (Center for Coastal Studies, 2023), focusing on metrics relevant to *Z. marina* health: daily records for mean SST and annual maximum summer SST for the coastal waters at or near each *Z. marina* site and year in the dataset were appended to the appropriate location and year using Python's Pandas library. To represent ocean acidification, data on water column pH (or pCO2 when available) was collected from sources such as the NDBC and Center for Coastal Studies (2023). All data were carefully time-aligned by year, and when site-specific data were available, by location.

Table 1: A high-level overview of the dataset used in the study, containing the years of area observation for each study site, along with the sources for data compilation.

Location	Years Observed	Source
Madaket Harbor	1995, 2002, 2010, 2018	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Morris Island	1995, 2001, 2010, 2018	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Stage Harbor System	1995, 2001, 2010, 2018	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025;

		Costello et al., 2011)
Hamblin and Jehu Ponds	1995, 2001, 2007, 2018	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Waquoit Bay	1995, 2001, 2007, 2017	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Cuttyhunk Harbor	1995, 2001, 2007, 2017	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
West Cape Cod Canal	1995, 2001, 2007, 2017	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Lagoon Pond	1995, 2001, 2007	(Center for Coastal Studies, 2023; Costello et al., 2011)
Cape Pogue	1995, 2001, 2007, 2017	(Center for Coastal Studies, 2023; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Salem Harbor	1995, 2001, 2007, 2016	(NOAA, n.d.; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Lynn Harbor	1995, 2001, 2007, 2016	(NOAA, n.d.; Commonwealth of Massachusetts, 2025; Costello et al., 2011)
Boston Harbor	1995, 2001, 2007, 2016	(MWRA, 2022; Commonwealth of Massachusetts, 2025; Costello et al., 2011)

Field Site Visit

To contextualize the data collected and findings of this study, a site visit was conducted on February 3, 2025, at Blackrock Beach, located in Nahant, Massachusetts. Blackrock Beach is the closest land-based vantage point to the *Z. marina* beds in Broad Sound, corresponding to the observations in Lynn (Table 1), one of the key areas included in the spatial dataset used for model training (Commonwealth of Massachusetts, 2025). The visit, conducted at 9:30 AM following light snowfall, revealed a shoreline that was far narrower than expected, with a beach-to-street distance of under 100 feet even at low tide. Protective cement blocks lined the beachfront, accompanied by hardy shrub species, indicating potential efforts to mitigate coastal erosion and wave attenuation during storm surges. The beach itself was heavily rock-covered with a small tidal pool and marsh grass stands present nearby.

Thin ice patches remained in the intertidal zone and possibly extended across the subtidal beds of interest, inhibiting any efforts to observe the *Z. marina* beds more closely. Water currents at the site were visibly active, with water temperatures near freezing, further limiting physical access to the beds. However, since the research questions guiding this project were addressed through remotely sensed and publicly available water quality datasets, no direct sampling was required. The site visit provided valuable insight into the environmental context in which *Z. marina* exists in Broad Sound, and highlighted the importance of remote and computational methods in temperate seagrass research.

Additionally, the proximity of Northeastern University's Marine Science Center, located just down the road from the beach, suggests a potential opportunity for future collaboration or consultation regarding seagrass health in the area. The presence of active marine research infrastructure nearby may be particularly useful for expanding data access or validating model results in future iterations of this project.

Data Processing

Because the assembled data came from diverse sources with varying temporal coverage, the datasets were standardized and merged into a single flat table for analysis. Each record in the final dataset corresponded to a particular site-year combination, containing (1) baseline *Z. marina* area (ha) at the start of the interval (or previous year's area), (2) change in *Z. marina* area (%) over that year, which was the target variable to be predicted by the model, (3) daily SST records for that year (°C), (4) annual maximum SST (°C), and (5) daily pH records over that year. This resulted in a multidimensional data frame constructed which was again flattened using Pandas during preprocessing. The percent change in area was calculated as $(Area_t - Area_{t-1}) / Area_{t-1}$. Using percent change as the response helped normalize differences in absolute bed size and focused the model on growth and decline rates. The choice of features was informed by *Z. marina* physiology, as temperature stress is often related to

extreme maxima as well as elevated variations and sustained marine heat waves. Additionally, prior bed size could influence growth potential, as small beds can decline faster due to edge effects or Allee effects (Van Tussenbroek et al., 2016). pH was included to capture potential OA effects. All variables were inspected for quality, and obviously erroneous data points were removed or corrected using NumPy for data cleaning. The final dataset was relatively limited in size due to data availability constraints. This small sample size motivated the use of a XGBoost with careful regularization to avoid overfitting.

Model Setup

An extreme gradient boosting regressor was implemented using the XGBoost library. XGBoost was chosen for its efficiency and proven performance on semi-structured data. For this study's model, features included the *Z. marina* baseline area, SST and pH records, maximum annual SST, and the target was the percent change in *Z. marina* area over the following year studied. The dataset was split using LeaveOneGroupOut, where the most recent record for a site was used as a testing set (held out for evaluating performance on unseen data), and the remainder used as a training set. Because the dataset was not large, a repeated k-fold cross-validation on the training set was used for more effective model tuning and evaluation.

XGBoost has several hyperparameters that control model complexity, such as tree depth, number of trees, learning rate, L1/L2 regularization weights, and subsampling rates. Tuning these is an important step when working with a small dataset to balance model flexibility and overfitting. RandomizedSearchCV was utilized to find the optimal hyperparameter set that minimized cross-validation error. In preliminary runs, it was found that very deep trees would overfit the limited data, so the maximum tree depth was restricted to three levels. The final model used an ensemble of several hundred trees with a max depth ~3, a learning rate of approximately 0.05–0.1, and moderate regularization. These settings were found to give the best validation performance. XGBoost's built-in regularization and shrinkage helped prevent the model from simply memorizing the limited data.

Model Evaluation

Performance was assessed using the mean absolute error (MAE) and the coefficient of determination R^2 on the test set, as well as during cross-validation. MAE (in units of percentage points of area change) provides a measure of prediction error, and R^2 indicates the proportion of variance in *Z. marina* change that was accounted for by the model. Given the ecological context, the primary interest was in whether the model could predict the direction of change (increase vs. decrease) and general magnitude correctly, even if exact percentages had uncertainty. Additionally, to interpret the fitted model, feature importance analysis was implemented. XGBoost can provide feature importance scores based on how each feature contributes to reducing prediction error in the trees. Shapley Additive Explanations (SHAP) values were also calculated to understand the directionality of each feature's effect on predictions. SHAP values

help interpret how increasing or decreasing a feature value impacts the predicted outcome for each observation (Ponce-Bobadilla et al., 2024). This was important for ecological context in the results of this study, to confirm whether OW or OA had a larger influence in predicted *Z. marina* coverage changes.

Results

Model Performance

The XGBoost regressor demonstrated adequate predictive skill given the complexity of the problem and limited data. On the held-out test set, the model's MAE was approximately 28% in terms of absolute percent change in *Z. marina* area; on average, the model's prediction of annual change was within 28 percentage points of the observed change. *Z. marina* percent changes in the dataset varied widely year-to-year (including some large declines), and a 28% MAE indicates the model was often within a few tens of percent of the true value. The coefficient of determination R^2 was 0.725, indicating that about 72.5% of the variance in year-to-year *Z. marina* change was explained by the model's features. This R^2 value suggests a significant portion of the signal was captured. For comparison, a baseline model predicting zero change each year would have an R^2 of 0 and a much larger MAE. Cross-validation results were consistent with the test performance, giving confidence that the model generalized and was not overfitted.

The model correctly identified the overall trend of *Z. marina* change over the study period as a decline. When averaging across all sites and years in the simulation, the model's predictions indicated a slight downward slope in *Z. marina* area of approximately 0.9–3% per year. This aligns with observations in Massachusetts that *Z. marina* habitats have been gradually diminishing (Costello et al., 2011). It is important to note that not every site-year was predicted to decline, as evident in Figure 3. This could potentially be explained by successful conservation efforts, a lack of extreme marine heatwave events during that period, or additional ecological factors that were beyond the scope of this study. However, under typical recent conditions (with warming), the central tendency was negative growth, i.e. net loss of *Z. marina*, emerging from the relationships with temperature and other variables, suggesting those factors indeed drive a downward trajectory.

Key Predictors of *Z. marina* Change

An analysis of feature importance revealed that temperature-related variables were the dominant predictors in the model. Both the annual maximum SST and the annual SST records had high importance scores, with maximum temperature being the most influential feature (Figure 2). Furthermore, randomizing the maximum temperature data resulted in the largest drop in model performance, indicating the model relied heavily on that feature. Annual SST records were the second-most important predictor. Together, these two features contributed the majority

of the model's explained variance in *Z. marina* change. The initial *Z. marina* area was the next most important feature in the ranking, suggesting that there is some carryover effect. For instance, larger existing meadows might have more resilience, affecting the percentage change. Lastly, pH had the lowest importance among the features, though it was not zero. In the permutation importance test, shuffling pH values caused only a small increase in error, indicating a relatively minor role, and its gain-based importance in the trees was modest (~0.16 in XGBoost's normalized gain units). This does not mean that pH had no effect, but rather that within the range of variability observed in the data, pH was not a strong driver of year-to-year *Z. marina* fluctuations compared to SST.

The SHAP value analysis helped to explain how these features affected the model's predictions. Lower SSTs corresponded to higher predicted *Z. marina* growth (or lower decline), while higher SSTs led to predictions of greater *Z. marina* loss. Years with unusually high summer maximum temperatures tended to have negative predicted outcomes for *Z. marina*. This aligns with expectations, as heat stress events often cause *Z. marina* die-back (Collier et al., 2018). The SHAP dependence plot showed a threshold-like behavior around the upper 10% of the temperature range. Beyond a certain temperature threshold, the predicted percent change became strongly negative. Conversely, cooler-than-average years were predicted to have neutral or even positive *Z. marina* growth, suggesting that in the absence of heat stress, *Z. marina* can stabilize or expand if other conditions are favorable.
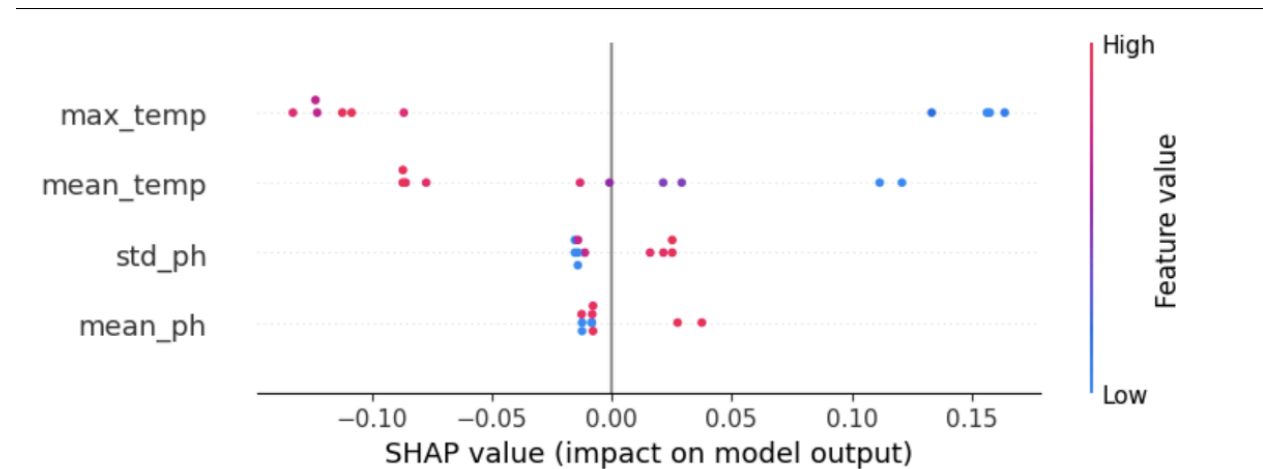

Figure 2: Impacts of each feature on model prediction through SHAP analysis. Negative values indicate a prediction of bed coverage decline, and positive indicate an increase. Values around 0.0 indicate the feature led to no derivation from the mean.

For pH, the SHAP analysis indicated only a slight trend: more acidic conditions were associated with marginally better *Z. marina* outcomes, but the effect size was very small relative to SST impact. This hints at the CO_2 fertilization effect that has been observed in prior research, but in this dataset the variation in pH was limited and the effect was often overshadowed by fluctuations in SST. It's worth noting that the Massachusetts coastal pH variation over the study

period might not have been large enough to elicit a strong growth response. Regardless, the model's inclusion of pH did slightly adjust some predictions, in line with the notion that acidification on its own would tend to favor *Z. marina*, but when high temperatures co-occurred, the temperature effect dominated.

Model Visualization and Validation

Figure 3 describes the predicted percent changes versus the observed values for all available site-years that had corresponding SST and pH data. The points clustered around the 1:1 line reasonably well, with a few outliers, such as 2013. The majority of predictions correctly captured the direction of change (increase vs. decrease). For instance, in known decline years, the model nearly always predicted a decline. The magnitude was sometimes under- or over-estimated, hence the 28% MAE, but there was a positive correlation between predictions and observations. One notable pattern was that the model slightly under-predicted some of the most extreme decline events (e.g., a site that lost ~80% of area in one year might have been predicted to lose ~50%). This is not surprising, as extreme events can be partly due to factors not in the model, such as a disease outbreak or a storm. Conversely, the model occasionally predicted a moderate decline for a site-year that actually had a small gain. These cases often corresponded to instances where temperature was high, driving the model to expect a loss, but local mitigation or restoration efforts may have aided that site in reality. Such discrepancies highlight that additional factors like water quality, restoration, herbivory, etc. also influence outcomes, and their absence from the model can cause some error.
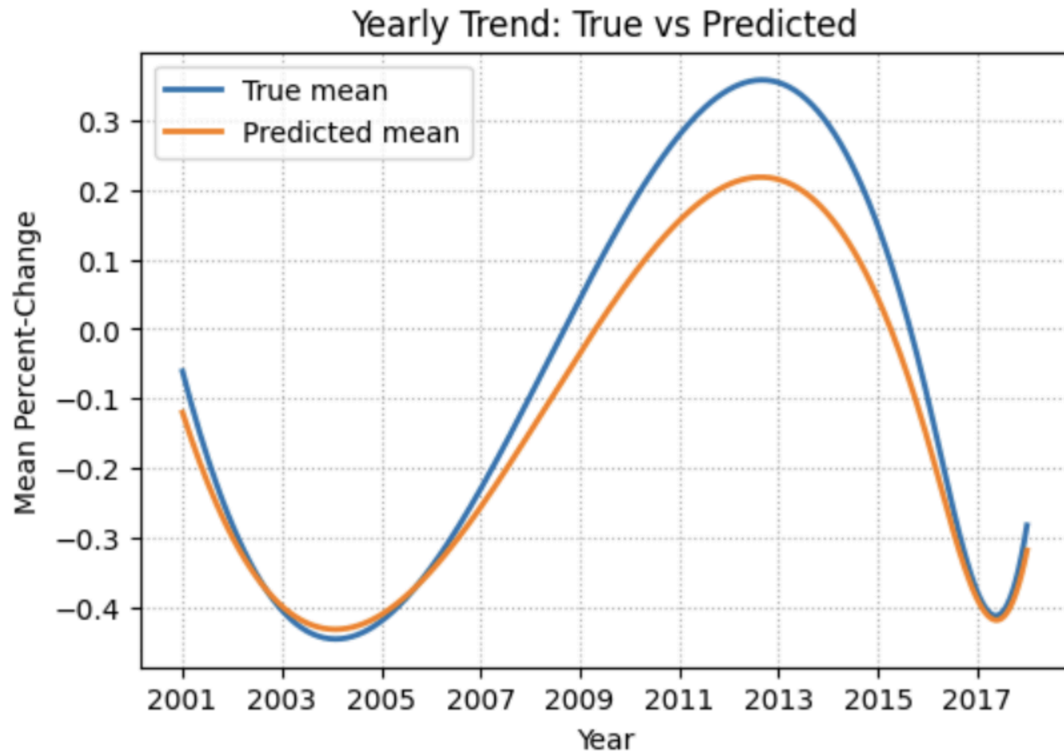
Figure 3: A visual comparison of the true mean (the average percent change in bed coverage for a given year across all available study sites) and the predicted mean (the average of the model's predicted percent changes). It is important to note that not all years had available data for every study site, and so the general shape of the graph does not indicate any patterns in overall *Z. marina* growth or decline.

Discussion

Interpreting the Combined Impact of Warming and Acidification

The modeling results indicate that, in the present Massachusetts coastal environment, ocean warming is a far more consequential driver of *Z. marina* change than ocean acidification. Even though this study considered both factors together, the learned model essentially points to SST—especially extreme heat events—as the determinant of whether *Z. marina* area increases or decreases in a given observed year. The SHAP analysis reinforcing a temperature threshold effect is consistent with the known thermal physiology of *Z. marina*. Past studies have documented that the species experiences heat stress symptoms, like reduced chlorophyll or shoot mortality, when water temperatures exceed roughly 25–30 °C for extended periods (Collier et al., 2018). The model's high sensitivity to the maximum summer temperature suggests that those peak events can be incredibly damaging.

Conversely, ocean acidification's direct effect on *Z. marina* appears to be mild in the current observational record along the Massachusetts coast. This does not contradict the notion that CO2 can stimulate *Z. marina* growth; rather, it may reflect that within the range of pH

observed, *Z. marina* already has sufficient carbon for photosynthesis, or other factors are limiting growth. It's also possible that any growth benefits from OA manifest more in productivity or biomass, which were not measured here, than in immediate area expansion. Such findings are in line with Collier et al. who found that warming stresses overshadowed OA benefits (2018). They are also consistent with the research of Yan et al., who found that OA's benefit to *Z. marina* is context-dependent on temperature (2024). Under moderate temperatures, seagrasses can take advantage of extra CO_2 to bolster certain metabolic processes (Egea et al., 2018), but under high temperature stress, those benefits diminish as the plant reallocates energy to stress responses or simply cannot maintain positive carbon balance regardless of CO_2 availability (Collier et al., 2018). This suggests that future ocean conditions that include both warming and acidification will likely be net harmful to *Z. marina*, unless the warming is kept within tolerable levels. The slight positive effect of acidity in this model might suggest that a comparison between two scenarios with constant warming, the one with higher dissolved CO_2 could be marginally better for *Z. marina*. However, with continued OW, the gains from OA are potentially too small to compensate.

Broader Ecological and Economic Implications

The continued decline of *Z. marina* beds, as projected by both this study's model and observational studies, is alarming for coastal ecosystems and economies. *Z. marina* meadows are critical fish habitat, supporting commercially important migratory fish species, especially during juvenile stages (Chalifour et al., 2019). Thus, a shrinking *Z. marina* extent can translate to reduced fisheries recruitment. Moreover, *Z. marina* contributes to shoreline protection by damping wave energy and stabilizing sediments; its disappearance can make coasts more vulnerable to erosion, especially as sea levels rise (Reidenbach & Thomas, 2018). The blue carbon significance of *Z. marina* is another consideration for the species' significance. Although New England *Z. marina* meadows are not as carbon-rich as tropical mangroves, they still accumulate organic carbon in their soils. Ongoing loss of *Z. marina* not only forfeits future carbon sequestration potential but also can lead to remineralization of stored carbon, releasing CO_2 back to the atmosphere or water (Dahl et al., 2020). A recent estimate suggested that disturbed seagrass beds globally could emit on the order of 1.15 Gt of CO_2 this century if not protected (Krause et al., 2025). Protecting and restoring *Z. marina* could therefore be a strategy to maintain natural carbon sinks as part of blue carbon initiatives.

From a blue economy standpoint, investing in *Z. marina* conservation yields multiple benefits: it helps sustain fisheries, supports biodiversity, and increases climate resilience for coastal and marine habitats. These findings highlight that such benefits are at risk from climate change, and climate adaptation must be part of *Z. marina* management strategies. Coastal municipalities in Massachusetts and elsewhere may need to integrate efforts to reduce local stressors, such as nutrients, sediments, propeller scarring, to improve *Z. marina* health and perhaps increase its thermal tolerance margin. Additionally, identifying cooler refuge areas and

prioritizing those for *Z. marina* protection or planting efforts could preserve some meadows as climate changes. Active restoration of *Z. marina*, in the form of transplanting shoots, might also consider future climate suitability—for example, sourcing thermally tolerant strains of *Z. marina*. The model could be used in a prescriptive manner by plugging in projected temperature increases to see how *Z. marina* might fare, guiding where restoration would be most successful. For instance, if by mid-century Massachusetts summers are on average 2 °C warmer, the model would predict significantly larger annual losses, implying that without intervention many beds could collapse. Conversely, if strong climate action stabilizes temperatures, *Z. marina* might maintain its presence especially if aided by the CO2 fertilization effect.

One interesting implication of this study is the slight positive—though weak—effect of OA in the model. This suggests that if the dominant threats to eelgrass is eliminated or mitigated, there is a potential underlying benefit to *Z. marina* from higher dissolved CO2. In a scenario of aggressive greenhouse gas mitigation, the rate of acidification could slow and perhaps local actions could even improve pH. Managers and conservationists could leverage the presence of seagrassed as OA buffers. While this study looked at the impact of OA on *Z. marina*, the reverse is another ecosystem service to consider.

Utility of XGBoost for Ecological Modeling

This study demonstrates the potential for applications of extreme gradient boosting to ecological data. A notable advantage found is XGBoost's resilience with a small, noisy dataset. Traditional statistical models, such as linear regressions, which struggle to capture the nonlinear effects would have required interaction terms to even approximate the OW and OA interplay (Schuster et al., 2022). The XGBoost model, by contrast, automatically handled these interactions, perhaps implicitly learning that pH had a greater influence when temperature is moderate, but at SSTs the outcome is poor regardless. The algorithm's built-in regularization provided confidence in the results despite the low sample size, as it guards against overfitting by penalizing overly complex trees. Conducting these experiments was made easier by XGBoost's efficiency due to its quick training, allowing for multiple iterations of configuration and input testing. Additionally, the permutation and SHAP analyses provided interpretability to what could otherwise be a "black box" model. In the context of this research, such interpretability is key for gaining acceptance of model recommendations and uncovering the dynamics between OA, OW and *Z. marina* coverage.

Other studies corroborate that XGBoost and related ensemble methods can outperform traditional methods in ecological forecasting. As mentioned, Ahn et al. found gradient boosting models improved harmful algal bloom predictions by capturing nonlinear drivers (2023). Similarly, gradient boosting has been used to predict species distributions under climate change with high accuracy, and to classify habitat from remote sensing data (Zhang et al., 2019). This study adds to the current literature by demonstrating that XGBoost can successfully integrate

climate variables and biological response data. It should be noted that machine learning models require careful validation, and steps were taken to ensure reliability, like cross-validation and examining residuals. Future work could integrate additional data, such as nutrients, light availability, or biotic factors like grazing, to further improve the model. With greater data availability, a comparison to other machine learning algorithms (random forests, neural networks) or hybrid approaches could also be beneficial in future iterations of this research. Moreover, XGBoost could be supplemented with existing climate model outputs for OW and OA in Massachusetts to forecast *Z. marina* changes decades into the future. This would effectively merge statistical learning with scenario modeling, potentially giving resource managers a tool to visualize *Z. marina* trajectories under different emission scenarios.

Limitations and Future Research

While the study's model performs well, there are some limitations to acknowledge. The dataset, being patchworked from various sources, may contain measurement inconsistencies and spatial heterogeneity that the model cannot fully account for. Massachusetts' coast has diverse sub-environments – from the open Gulf of Maine to sheltered bays – and pooling them assumes the SST and pH effects are uniform. In reality, local adaptation of *Z. marina* or varying community interactions could cause deviations. Expanding the dataset to include more regions or a longer time span would help generalize the model. Additionally, it could be valuable to include extreme events explicitly, like an indicator for a marine heatwave year or a major storm, to see how those events correlate with *Z. marina* losses and improve model performance in outlier years.

<div align="center">

Conclusion

</div>

By leveraging an XGBoost machine learning model on an integrated dataset of *Z. marina* area, temperature, and acidity, this study provides a comprehensive assessment of how *Zostera marina* meadows are likely to respond to the dual pressures of ocean warming and acidification. The results indicate that warming overwhelmingly drives *Z. marina* declines, with even slight increases in summer water temperatures causing appreciable losses, while ocean acidification plays a secondary role that is insufficient to counteract heat stress. These findings are corroborated by experimental evidence (Egea et al., 2018; Collier et al, 2018) and suggest that, under current trajectories, *Z. marina* meadows in Massachusetts will continue to diminish, posing risks to the ecosystem services they provide. This work highlights the ecological urgency for climate change mitigation and adaptation. Curbing greenhouse gas emissions to limit ocean warming will be critical for the survival of *Z. marina* and other coastal habitats. In the meantime, local management should focus on building *Z. marina* resilience through other methods, such as improving water quality, reducing direct human impacts, and protecting cooler refuge areas, so that these meadows are in the best possible condition to face changing ocean conditions.

The successful application of XGBoost here illustrates the value of data-driven modeling in ecology. As environmental datasets grow and become more accessible with remote monitoring systems, machine learning offers the ability to synthesize information and forecast changes in ways that complement field studies and experiments. However, careful attention to interpretation is needed, and combining feature importance and SHAP analysis can extract insights from the model. Future research can build on this approach by integrating broader factors, like global climate indices or land use changes, and by applying the model framework to other regions or species of seagrass. Protecting *Z. marina* with the increasing intensity of OW and OA will require both scientific understanding and proactive management. Studies such as this provide a blueprint for using data and advanced computational tools to anticipate ecological changes and inform conservation of blue carbon ecosystems like *Zostera marina*. This research aims to contribute to such efforts by clarifying the dominant threats and helping to outline the need for interventions to preserve Massachusetts' *Z. marina* habitats in the face of climate change.

References

Abe, H., Ito, M., Ahn, H., & Nakaoka, M. (2022). Eelgrass beds can mitigate local acidification and reduce oyster malformation risk in a subarctic lagoon, Japan: A three-dimensional ecosystem model study. *Ocean Modelling, 173*, 101992. https://doi.org/10.1016/j.ocemod.2022.101992

Ahn, J. M., Kim, J., & Kim, K. (2023). Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. *Toxins*, *15*(10), 608. https://doi.org/10.3390/toxins15100608

Barañano Carrion, C., Fernández, E., Morán, P., Urbieta, P., & Martínez, G. (2022). Population dynamics of a fragmented subtidal *Zostera marina* population affected by shell fishing. *Estuarine, Coastal and Shelf Science, 269*, 107818. https://doi.org/10.1016/j.ecss.2022.107818

Center for Coastal Studies. (2023). *Cape Cod Water Quality*. https://www.capecodbay-monitor.org/

Chalifour, L., Scott, D., MacDuffee, M., Iacarella, J., Martin, T., & Baum, J. (2019). Habitat use by juvenile salmon, other migratory fish, and resident fish species underscores the importance of estuarine habitat mosaics. *Marine Ecology Progress Series, 625*, 145–162. https://doi.org/10.3354/meps13064

Collier, C. J., Langlois, L., Ow, Y., Johansson, C., Giammusso, M., Adams, M. P., O'Brien, K. R., & Uthicke, S. (2018). Losing a winner: Thermal stress and local pressures outweigh the positive effects of ocean acidification for tropical seagrasses. *The New Phytologist, 219*(3), 1005–1017. https://doi.org/10.1111/nph.15234

Commonwealth of Massachusetts. (2025). *MassGIS Data: MassDEP Eelgrass Mapping Project*. https://mass-eoeea.maps.arcgis.com/apps/webappviewer/index.html?id=07f8d48c714f4f8 1bec49864ecf252da

Costello, C. T., & Kenworthy, W. J. (2011). Twelve-year mapping and change analysis of eelgrass (*Zostera marina*) areal abundance in Massachusetts (USA) identifies statewide declines. *Estuaries and Coasts, 34*, 232–242. https://doi.org/10.1007/s12237-010-9371-5

Dahl, M., Asplund, M. E., Björk, M., et al. (2020). The influence of hydrodynamic exposure on carbon storage and nutrient retention in eelgrass (*Zostera marina* L.) meadows on the Swedish Skagerrak coast. *Scientific Reports, 10*, 13666. https://doi.org/10.1038/s41598-020-70403-5

Egea, L. G., Jiménez-Ramos, R., Hernández, I., Bouma, T. J., Brun, F. G., & Hwang, J.-S. (2018). Effects of ocean acidification and hydrodynamic conditions on carbon metabolism and dissolved organic carbon (DOC) fluxes in seagrass populations. *PLOS ONE, 13*(2), e0192402. https://doi.org/10.1371/journal.pone.0192402

Krause, J. R., Cameron, C., Arias-Ortiz, A., et al. (2025). Global seagrass carbon stock variability and emissions from seagrass loss. *Nature Communications, 16*, 3798. https://doi.org/10.1038/s41467-025-59204-4

Massachusetts Water Resources Association. (n.d.). *Download Environmental Data*. MWRA. https://www.mwra.com/harbor/download-environmental-data

Matheson, K., McKenzie, C. H., Gregory, R. S., Robichaud, D. A., Bradbury, I. R., Snelgrove, P. V. R., & Rose, G. A. (2016). Linking eelgrass decline and impacts on associated fish communities to European green crab *Carcinus maenas* invasion. *Marine Ecology Progress Series, 548*, 31–45.

NOAA. (n.d.). National Data Buoy Center. https://www.ndbc.noaa.gov/

Novak, A. B., Plaisted, H. K., Hughes, Z. J., Mittermayr, A., & Molden, E. (2023). Eelgrass (*Zostera marina* L.) populations are threatened by high sea-surface temperatures and impaired waters on Nantucket Island, USA. *Marine Pollution Bulletin, 197*, 115689. https://doi.org/10.1016/j.marpolbul.2023.115689

NVIDIA. (2025). *What is XGBoost?* NVIDIA Data Science Glossary. https://www.nvidia.com/en-us/glossary/xgboost/

Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science, 17*(11), e70056. https://doi.org/10.1111/cts.70056

Reidenbach, M. A., & Thomas, E. L. (2018). Influence of the seagrass, *Zostera marina*, on wave attenuation and bed shear stress within a shallow coastal bay. *Frontiers in Marine Science, 5*, Article 397. https://doi.org/10.3389/fmars.2018.00397

Schuster, N. A., Rijnhart, J. J. M., Twisk, J. W. R., & Heymans, M. W. (2022). Modeling non-linear relationships in epidemiological data: The application and interpretation of spline models. *Frontiers in Epidemiology, 2*, 975380. https://doi.org/10.3389/fepid.2022.975380

Short, F. T. (2014). *Eelgrass distribution in the Great Bay Estuary for 2013*. PREP Reports & Publications, 348. https://scholars.unh.edu/prep/348

Short, F. T., Polidoro, B., Livingstone, S. R., Carpenter, K. E., Bandeira, S., Bujang, J. S., et al. (2011). Extinction risk assessment of the world's seagrass species. *Biological Conservation, 144*(7), 1961–1971. https://doi.org/10.1016/j.biocon.2011.04.010

Van Tussenbroek, B. I., Soissons, L. M., Bouma, T. J., et al. (2016). Pollen limitation may be a common Allee effect in marine hydrophilous plants: Implications for decline and recovery in seagrasses. *Oecologia, 182*(2), 595–609. https://doi.org/10.1007/s00442-016-3665-7

Verma, N. (2022, September 7). *XGBoost algorithm explained in less than 5 minutes*. Medium. https://medium.com/@techynilesh/xgboost-algorithm-explained-in-less-than-5-minutes-b561dcc1ccee

Waycott, M., Duarte, C. M., Carruthers, T. J. B., et al. (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the National Academy of Sciences, 106*(30), 12377–12381. https://doi.org/10.1073/pnas.0905620106

Yan, W., Wang, Z., Pei, Y., & Zhou, B. (2024). Adaptive responses of eelgrass (*Zostera marina* L.) to ocean warming and acidification. *Plant Physiology and Biochemistry, 206*, 108257. https://doi.org/10.1016/j.plaphy.2023.108257

Zarnoch, C. B., Hoellein, T. J., Furman, B. T., & Peterson, B. J. (2017). Eelgrass meadows, *Zostera marina* (L.), facilitate the ecosystem service of nitrogen removal during simulated nutrient pulses in Shinnecock Bay, New York, USA. *Marine Pollution Bulletin, 124*(1), 376–387. https://doi.org/10.1016/j.marpolbul.2017.07.061

Zhang, H., Eziz, A., Xiao, J., Tao, S., Wang, S., Tang, Z., Zhu, J., & Fang, J. (2019). High-resolution vegetation mapping using eXtreme Gradient Boosting based on extensive features. *Remote Sensing, 11*(12), 1505. https://doi.org/10.3390/rs11121505

Zimmerman, R. C., Hill, V. J., & Gallegos, C. L. (2015). Predicting effects of ocean warming, acidification, and water quality on Chesapeake region eelgrass. *Limnology and Oceanography, 60*(5), 1781–1804. https://doi.org/10.1002/lno.10139

Zimmerman, R. C., Kohrs, D. G., Steller, D. L., & Alberte, R. S. (1997). Impacts of CO2 enrichment on productivity and light requirements of eelgrass. *Plant Physiology, 115*(2), 599–607. https://doi.org/10.1104/pp.115.2.599