The impact of climate change on Phytoplankton communities in the Coorong Wetlands

By

Thai Quang Le
Master of Data Science
Supervised by
Prof.Sergiy, Prof.Sophie and Dr.Daniel

Thesis
Submitted to Flinders University
for the degree of

Master of Data Science

College of Science and Engineering
20/10/2025

TABLE OF CONTENTS

ABSTRACT

A diverse phytoplankton community preserves the ecological balance in the Ramsar-listed wetland of international status, Coorong wetlands in South Australia. The wetlands are economically significant and essential to South Australia, its society, as well as its cultural heritage. Therefore, the role of climate change in the phytoplankton communities requires an appreciation in the management and maintenance of the ecologies.

During the past several decades, decreased freshwater inputs can be listed among the primary aspects that changed the environmental situation in the Coorong. In addition, growing temperatures, rise in salinity, water chemistry alterations, and repeated impactful weather situations are leading to changes in the abundance of the phytoplankton, which is affecting aquatic food webs and overall water quality. This paper applies machine learning techniques, namely, Random Forest models to analyze environmental data obtained at Flinders University Laboratory in order to measure the effect of climate change on the phytoplankton community structure and abundance.

Such studies will have positive differences to the assumptions of phytoplankton societies in the face of continuous climate stresses. These insights play a fundamental role since they can be relied on to design an adaptive management strategy to ensure that the environmental integrity and resilience of the Coorong are maintained, as well as to inform future climate adaptation studies.

DECLARATION

I certify that this Thesis:

1. does not incorporate without credit any material that has been submitted previously to any university to obtain a degree or diploma.

2. the authorization of Flinders University will be required before the research within is submitted to any other degree or diploma; and

3. to the best of my knowledge and belief does not comprise any material that has been published already or authored by somebody, or the Artificial Intelligence, unless referred to somewhere in the text, and in compliance with Flinders University Academic Integrity policy and direction provided in the topic.

Thai Quang Le

20/10/2025

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors for their invaluable guidance, support, and encouragement throughout the duration of this research. Special thanks to the engineering services at Flinders Lab for providing essential data and technical support crucial to the completion of this project. I am also deeply grateful to the Goyder Institute for Water Research for the scholarship that made this study possible.

My heartfelt thanks go to my family and friends for their constant support, patience, and understanding during my studies. Lastly, I extend my appreciation to everyone who contributed directly or indirectly to this research project.

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

One of the most pressing environmental problems is climate change, which affects aquatic systems throughout the world, and can be fatal to estuarine systems since they are very sensitive to environmental changes caused by altered hydrological conditions (Lopez Abbate et al., 2017). The Coorong wetlands in South Australia are one such threatened ecosystems, where recorded complex interactions among climate drivers with ecological reactions and dynamic interactions of phytoplankton communities subject to environmental stresses occurs (Hemraj et al., 2017).

1.1. Environmental Pressures and Ecosystem Response

The strategic position of the Coorong at the end of Murray-Darling Basin makes it particularly vulnerable to upstream water management operations and adjustment in climatic conditions when they affect river flow. A high sensitivity of the system to low concentrations of inflowing freshwater was observed during the Millennium Drought of 2001-2009 that, under extended hyper salinity intervals, resulted in significant changes in the phytoplankton community structure (van Dijk et al., 2013; Leterme et al., 2015). During this critical period, salinity was registered above 100 PSU causing severe transformation of fresh-water centered communities to salt-arrestant diatoms and dinoflagellate communities (Leterme et al., 2015; Hemraj et al., 2017).

Over the past few decades, warming of both air and water temperatures, the alteration of precipitation patterns, and longer distances with extreme weather events may have been rapidly increasing, which continually alters the ecological dynamics of the system (Hemraj et al., 2017). This has resulted in comparatively complex relationships between salinity, temperature, nutrient status, and hydrological actions directly impacting upon the biomass, composition, and spatial dispersion patterns of the phytoplankton, a part of the wetland system (Leterme et al., 2015; Mosley et al., 2018). The study area and sampling locations are shown in Figure 1.

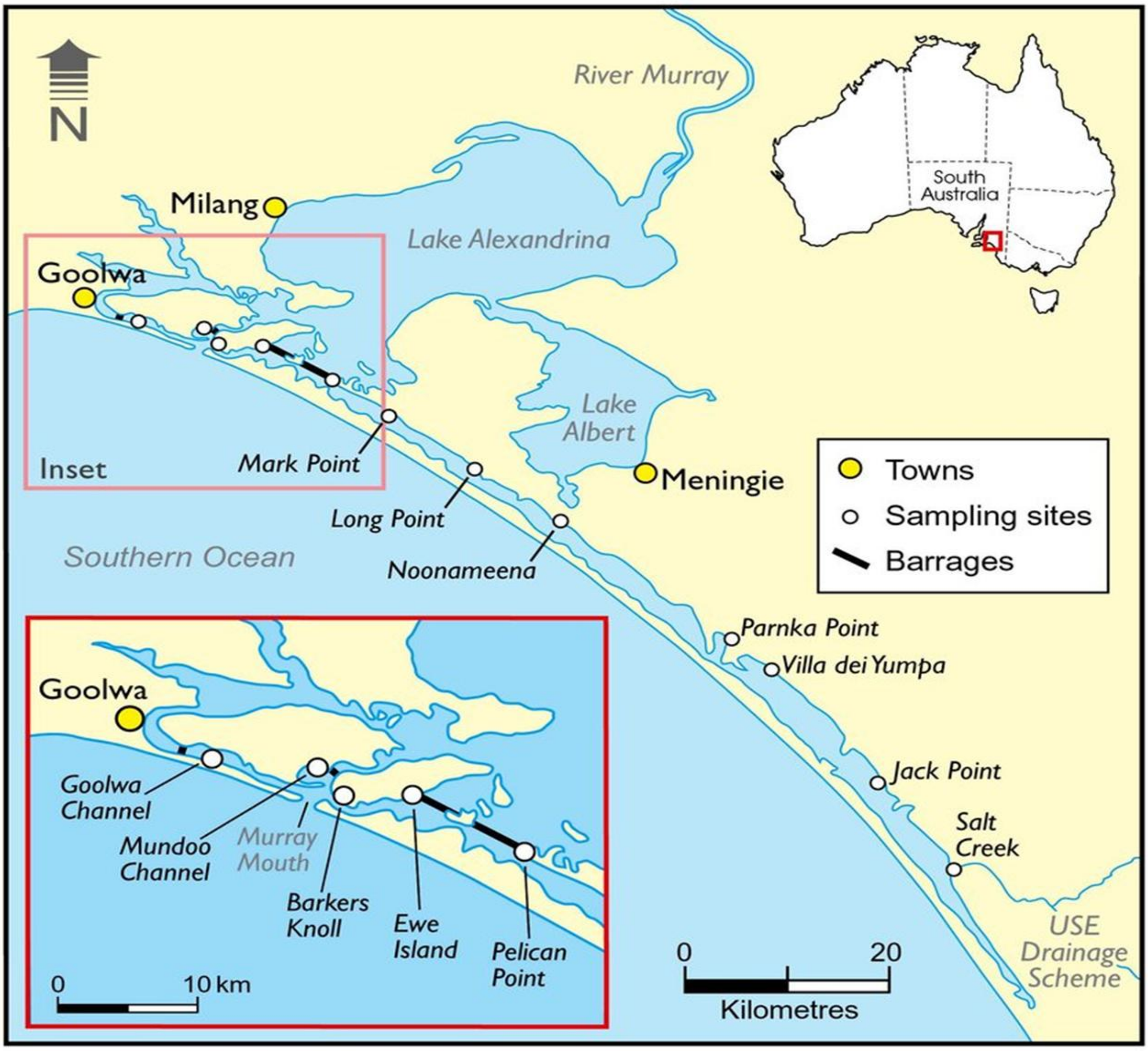


Figure 1: Map of the Coorong Wetlands showing sampling locations and geographic context (Brookes et al., 2009)

1.2. Phytoplankton as Ecological Indicators

Phytoplankton communities represent a sensitive indicator of change in the environment due to the limited time of generation, excessive vulnerability to environmental shifts in water chemistry and central roles within aquatic food webs (Jendyk et al., 2014). In an inverted estuarine system, the water becomes increasingly saline with distance away to the ocean at the inlet but in a normal estuary system, salinity lessens gradually (Wolanski, 2014). In inverse estuaries (e.g. Coorong) the transition of marine-to-freshwater was reported to demonstrate Chlorophyll-a concentrations offering a reliable substitute of phytoplankton biomass and primary productivity in water bodies (Leterme et al., 2015).

The interactions between the phytoplankton community and their environmental driving factors are not simple and involve non-linear interactions of the parameters at varying spatio-temporal scales.

Salinity may take over as the dominant structuring factor and a number of studies have found drastic community shifts as you proceed salty to the north and hypersaline to the south (Hemraj et al., 2017; Leterme et al., 2015). Figure 2 indicates the complicated trophic literature of the Coorong system.

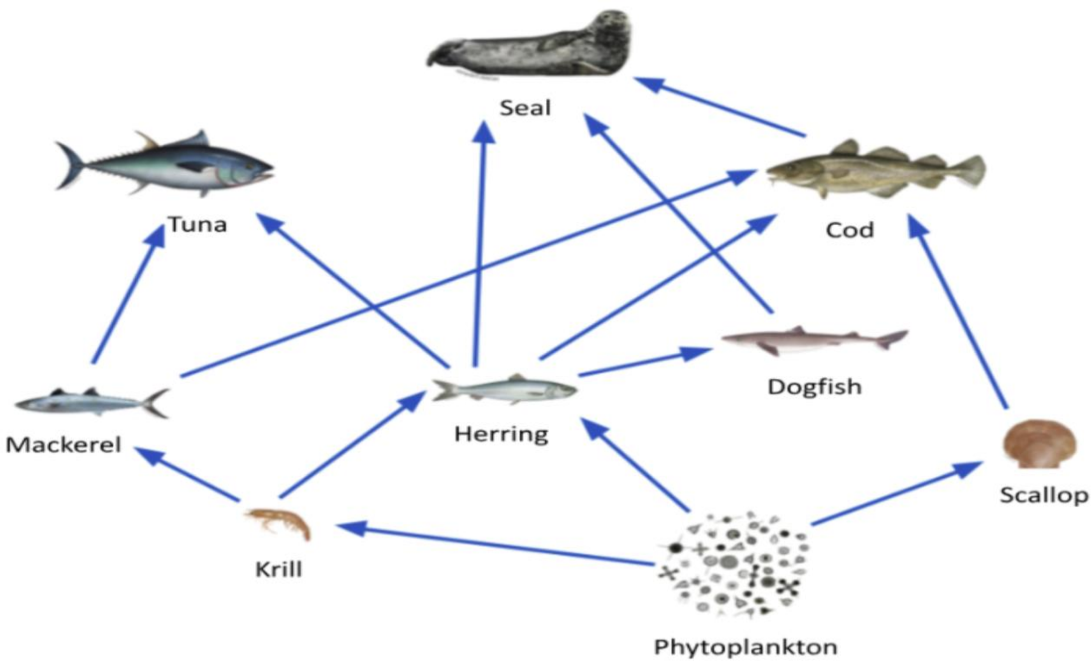


Figure 2: Coorong Wetlands food web diagram showing trophic relationships between species (National Oceanic and Atmospheric Administration, 2019)

1.3. Limitations of Traditional Analytical Approaches

The extensive environmental transformation within the Coorong is not new, and its essential manifestations have altered salinity concentrations throughout droughts (van Dijk et al., 2013), shifts in the composition of the phytoplankton community (Leterme et al., 2015), as well as altered hydrological regimes (Kingsford et al., 2011). The capacity to forecast phytoplankton ecological behaviors to these modifications is still limited by conventional statistical approaches. The non-linear interaction, threshold effect, and multi-scale process with which inverse estuarine systems are embodied are not reflected in simple linear regression models and correlation analysis (Olson et al., 2024). Generalized linear models, multivariate ordination methods, and traditional ecological models have inherent weaknesses, including assumptions about linearity, observations independence and normality of distribution of residuals - assumptions all of which are often not met in complex wetland systems in which species interactions, environmental thresholds, and time-response autocorrelation are common (Soudant et al., 1997; Olson et al., 2024). Changes in phytoplankton are not effectively captured by the application of their constant regression models based on the temporal variation and dynamism of phytoplankton communities (Soudant et al., 1997).

1.4. Machine Learning Applications in Ecological Research

Recent advances in machine learning have had encouraging solutions to overcome the shortcoming of conventional environmental ecological modeling. Examples of ensemble models that have shown to be more effective in ecological response prediction are random Forest, XGBoost and stacked regression models, which arrive at complex, non-linear relationships between biological outcomes and drivers of environmental change (Xu et al., 2023; Liu et al., 2023).

Such tools as explainable artificial intelligence (XAI) including SHAP (SHapley Additive exPlanations) methodology can address the limitations of the black box problem, meaning that complex models achieve high prediction accuracy and yet their decisions remain inexplicable and obscure this phenomenon often linked to the same models (Xu et al., 2023). These tools enable establishing important drivers of the environment and rank their relative importance and the impacts of interactions to learn about science and implement it to the management (Hussein et al., 2024).

1.5. Research Problem and Significance

The research problem in this case is that the available predictive models have not adequately accounted for phytoplankton dynamics in inverse estuaries to environmental change. This is one shortcoming of ecosystem management and conservation planning particularly when conditions of rapidly changing climate demand stronger adaptive styles with high predictive abilities.

The importance of the research problem is not restricted to the Coorong to everything about coastal wetlands ecology. The ability to establish improved prediction models will enhance scientific data on the impacts of climatic changes in inverse estuaries, aid in making evidence-based management choices, and provide methodological systems, which could be extended to other analog systems worldwide. Machine learning and explainable AI have promises of both creating ecological models and remaining interpretable as required in scientific and managerial activity.

1.6. Research Objectives and Questions

The authors develop and evaluate machine learning predictors of phytoplankton biomass in the Coorong Wetlands on the basis of rich datasets of the environment using 16 years of continuous observation (2007-2023). Chlorophyll-a concentration is used as an approximate indicator of phytoplankton level, as chlorophyll-a is the main photosynthetic molecule found in all phytoplankton and is a standardized metric to measure community biomass among taxonomic groups (Leterme et al., 2015). The research addresses four key questions:

- How do multiple environmental factors interact to predict phytoplankton abundance across different habitat zones in the Coorong?
- What are the relative contributions of temperature, salinity, nutrients, and hydrological factors to phytoplankton variability?
- Can ensemble machine learning models improve prediction accuracy compared to traditional statistical approaches?

- How do environmental drivers vary in importance across the five habitat zones as revealed by explainable AI methods?

1.7. Thesis Structure

This thesis discusses the research objectives in a systematic manner using five major chapters. After this introduction, Chapter 2 includes an extensive literature overview including the role of machine learning in the ecology of phytoplankton, explainable AI algorithms in environmental sciences, and gaps in research in the modeling of an inverse estuarine system. Chapter 3 elaborates the methodology plan such as data collection processes, machine learning models, and measuring schemes. Chapter 4 offers the findings of comparing models, analyses of feature importance, and habitat-specific results. Chapter 5 explains implications of findings on ecological literacy and practical management uses, pursues the limitations of the research conducted as well as future research directions.

2. LITERATURE REVIEW

2.1. The background and significance to the literature.

The Coorong's biodiversity and the environmental factors that affect it have been well documented in early research materials. Early work was directed at describing the fundamental elements of the ecosystem and how these elements respond to environmental variables, especially salinity and freshwater inflow patterns (Kingsford et al., 2011; Leterme et al., 2015). These alterations in the phytoplankton community and climate change have continued to be identified, and research indicates a strong relationship between increased temperatures and fluctuating levels of salinity and supply of nutrients and changes in species composition (Mosley et al., 2018; Hemraj et al., 2017).

A turning point in the Coorong was the Millennium drought period, 2001-2009, during which research concentrated on the susceptibility of the ecosystem to salinity changes and fluctuation in fresh water supply (van Dijk et al., 2013; Kingsford et al., 2011). This timeframe recorded significant alterations in the phytoplankton composition under contributions of salt increase (Leterme et al., 2015). Palaeoecological data, including diatom assemblages, have also assisted in building a sense of historical state and an understanding of long-term environmental shift (Haynes et al., 2011).

2.2. Machine Learning Applications in Phytoplankton Ecology

The more recent advances in machine learning have revolutionized phytoplankton ecology study have also given new capabilities to study complex interactions between the environment and phytoplankton with incomparable power. In the analysis of the density of algal cells within the water vessels, Xu et al. (2023) demonstrated a better performance of the Random Forests models, as compared to the traditional statistical methods and achieved moderate predictors over phytoplankton biomass when predicting algal cell density. In their study, very few machine learning algorithms were applied, including Support Vector Regression (SVR) and XGBoost, and Random Forest was found to perform better in most instances.

Secretarialization of approaches has been uniquely lauded in the application of aquatic ecology. Choudhary et al. (2025) endeavored the mutual stacked ensemble- regression models to forecast water quality outcomes at excellent outcomes with R2 of 0.9952 using the mixture of various streamlined algorithms comprising of the XGBoost, CatBoost, Random Forest, and Boosting. All these studies indicate that machine learning strategies can recognize non-linear relationships across the ecology which are at times weakly portrayed by traditional statistical techniques.

2.3. Explainable AI and Model Interpretability in Environmental Science

Explainable artificial intelligence (XAI) methods will become a noteworthy solution to the black box obstacle of machine learning models, which is still apparent in the area of environmental modelling.

Specifically, it is the SHAP (Shapley Additive exPlanations) that has gained prominence with regards to the usage of global and local explanations of model predictions without compromising theoretical underpinnings on the basis of game theory (Xu et al., 2023).

The importance of considering model interpretability is also applicable both at the technical level but also at the practical level within ecosystem management. Conciliable Artificial Intelligence practices permit correlation of essential environmental levels and factor relations causing adaptive management practices. They are useful with complex systems including the inverse estuarine contexts where there are many interacting stressors that respond in complex and non-linear manners.

2.4. Inverse Estuarine Systems and Environmental Gradients

Inverse estuarine systems present special difficulties to modeling phytoplankton communities with salinity decreasing with distance to the ocean outlet (in normal estuaries, salinity should increase with increasing distance to the ocean outlet). The dynamics of the environment and phytoplankton in the Coorong, an inverse estuary in South Australia, characterized the intricate connection among the salinity gradients, the nutrient composition, and the community structure of the habitat zones (Jendyk et al. 2014).

The implications of the alterations in the phenology of the biogeochemical cycles of the estuarine system and the possible effects of the change in the timing on the ecosystem functioning were also found by Testa et al. (2018). They tried to focus on the complex interactions between the nutrient cycle and temperature as well as those biological processes that characterise such dynamic environments.

2.5. Habitat-Specific Community Analysis Methods

The current techniques of phytoplankton community analysis identify the increasing significance in environment-specific responses to environmental gradients. Belluz et al. (2024) depicted application of correlation analysis and hierarchical clustering to demonstrate environmental drivers up to fjord into shelf, and this gives possible methodologies that can be used to investigate habitat-differentiating systems like Coorong.

In order to clarify the habitat of specific analysis, Dang et al. (2025) merged the data of the traits and the experienced environmental thresholds and applied Threshold Indicator Taxa Analysis (TITAN) to determine the extents of the ecological thresholds of single populations of phytoplankton. This method provides objective criteria to provide phytoplankton groups based on local environmental needs which enhance the use of previous taxonomic classifications in process-based modelling.

2.6. Limitations of Traditional Statistical Approaches

Conventional statistical frameworks are constrained with regard to implementation in sophisticated phytoplankton ecological systems. In spite of its strong sensitivity to significant parameters, Olson et al. (2024) demonstrated that traditional ecosystems models are unable to resolve observed patterns in primary production by the lakes and phytoplankton stoichiometry. Their study indicated the underlying incompatibility between supply and demand relationships that limit the predictive ability under changing environmental conditions.

Acevedo-Trejos et al. (2022) discussed methods of plankton diversity capture, and its impacts that the dynamic nature of the marine environment, where organisms move large distances and where environmental flourishes are high-dynamic settings not well-addressed with simple statistical measures. These restrictions indicate that it needs more sophisticated types of analysis that can handle more non-linear ecological relationships.

2.7. Research Gap and Future Directions

While the impacts of climate change on phytoplankton in wetland systems are extensively studied, quantitative predictive models linking multiple environmental drivers across habitat gradients are scarce. Traditional statistical approaches have proven inadequate to capture the inherent complexity of phytoplankton dynamics in inverse estuarine systems. However, machine learning approaches demonstrate results showing potential for understanding complex, non-linear ecological relationships and improving predictive capacity.

Current research gaps comprise the limited applied use of explainable AI techniques on inverse estuarine systems, habitat-specific modelling for hypersaline environments, and the lack of in-depth ensemble modelling approaches used for phytoplankton prediction in climate-stressed wetlands. Addressing these gaps is essential to developing predictive models to inform adaptive management strategies in response to accelerating climate change.

Future studies should focus on combining machine learning techniques with explainable AI methods to improve prediction accuracy and scientific knowledge about ecosystem processes. Combining ensemble modeling techniques with habitat-specific analysis has particular promise in complex systems such as the Coorong, where multiple environmental drivers interact across strong environmental gradients.

3. METHODS

3.1. Study Area and Data Collection

3.1.1. Coorong Wetlands System Description

The Coorong Wetlands system extends over approximately 140 kilometers along the coast of South Australia and represents a unique inverse estuarine ecosystem where salinity increases with distance from the Murray River mouth. It possesses two lagoons, and a minor division (between the two) in which the northern lagoon (North Lagoon) obtains freshwater inflows (through Murray River), and in which the southern lagoon (South Lagoon) obtains lower salinity water inflows.

In the system, there were five different habitat zones depending on the salinity regime as H1 containing salinity less than 5 PSU (freshwater to low brackish), H2 containing salinity 5 to 20 PSU (brackish) H3 containing salinity 21 to 40 PSU (high brackish to marine), H4 containing salinity 41 to 84 PSU (hypersaline) and H5 containing salinity 85 PSU or greater (high hypersaline). These habitats provide ecological niches with different salinity regimes, nutrient composition and hydrological regimes giving combination to structure phytoplankton community abundance and structure. It is interesting to note that the level of salinity in the Coorong changes because of the quantity of freshwater flowing, meaning that the habitat zones are spatial and relational changes whose presence over time is determined by environmental conditions.

3.1.2. Sampling Methodology and Data Collection

The sampling of the Coorong habitats was conducted under the standardized procedures as stipulated by the Flinders University Marine Microbial Ecology Research Laboratory and Coorong, Lower Lakes and Murray Mouth (CLLMM) Research Centre. The sample was obtained over the span of four periods and not consistently over the whole span: in the drought (2007, 2009), in the post-drought (2011-2014), in the pre-flood (2020-2021) and in the flood (2022-2023). It is expected that not all the habitat zones exist at each time of the period of sampling; in one instance a H1 habitat was not found during the drought period; this occurred due to salinity being too high in the whole system.

It was unlikely that sampling was to reduce changes in maximum productivity of phytoplankton and environmental factors occurring around the time of the day (0900-1500). The general sampling plan was to describe space heterogeneity throughout inverse salinity gradient and changes with time in terms of seasonal cycle and interannual climatic variability.

3.1.3. Environmental Parameters and Data Collection

Fourteen environmental parameters were systematically measured at each sampling location to characterize the multi-dimensional environmental space influencing phytoplankton dynamics:

Physical Parameters:

- Water temperature (Temp_C, °C): Measured continuously during CTD deployment.
- Salinity (PSU): Practical Salinity Units determined from conductivity measurements.
- pH: Measured using calibrated pH electrodes
- Water level (mAHD): Australian Height Datum reference level
- Distance from Murray Mouth (Dist_Murray_Mouth, km): Spatial coordinate along the lagoon system

Chemical Parameters:

- Dissolved oxygen concentration (mg/L) and saturation percentage (DO_%): Critical indicators of water quality and biological activity
- Silicate concentration (Silica, μM): Essential nutrient for diatom growth
- Ammonium concentration (μM): Reduced nitrogen form readily available to phytoplankton
- Phosphate concentration (μM): Limiting nutrients in many aquatic systems.
- Oxidized nitrogen concentration (μM): Including nitrate and nitrite forms.

Hydrological and Environmental Factors:

- Wind stress: Calculated from meteorological station data to represent mixing intensity.
- Lagoon designation: Categorical variable distinguishing North and South Lagoon systems

Biological Response Variable:

- Chlorophyll-a (ug /L): Primary phytoplankton biomass (abundance) proxy calculated by the conventional spectrophotometric procedures after ethanol extraction (Ritchie, 2008).

The analysis followed the standard procedures in the analytical laboratory of the Flinders University (ion chromatography of major nutrients, dissolved oxygen by spectrophotometry, chlorophyll-a by photospectrometric procedures). Quality assuring measures entailed certificate reference material analysis, duplicate sample analysis, and interlaboratory comparison programs.

3.2. Data Preprocessing and Quality Control

3.2.1. Data Cleaning and Quality Assurance

The process of data cleaning has been extensively handled to ensure the quality of input of the machine learning analysis. Several systematic cycles comprising the quality control protocol denote and project the potential risks to the integrity of data as well as ensuring the natural variability required to educate strong models.

Primary data cleaning techniques included: (1) ident mimic as well as data elimination through the use of the distinctive temporal-spatial identifiers, (2) data thinning guided by the omnipresence of target variables (chlorophyll-a) to retain the start-up of supervised learning approach, (3) the location of outliers (interquartile range (IQR) analysis at the expense of target values), and (4) consistency test (between the seasonal and interbed year levels of time) to indicate that accessible data represent representative sampling.

3.2.2. Feature Engineering and Normalization

To guarantee an equal opportunity to normalize all the inputs, Min-Max scaling was applied to place all the input variables in a similar [0,1] ratio to remove possible biases that might have otherwise occurred due to the different input measurements units and scales:

$$X_{norm} = \frac{X - X\_min}{X\_max - X\_min}$$

where X_{norm} represents the normalized feature value, X is the original measurement, and X_{min}, X_{max} denote the minimum and maximum values across the entire dataset. This normalization approach preserves the original data distribution while ensuring equal weighting of variables during model training.

Derived variables such as nitrogen-phosphorus (N:P) ratios to maintain stoichiometric relations classified as important in growth of phytoplankton and time-related variables (month), and year were also explored as sources of other forms of feature engineering. One-hot encoding of categorical variables (habitat zone, lagoon designation) by means of which they could be combined with numerical ones was applied.

3.2.3. Dataset Partitioning Strategy

An approach to break the dataset into time existed as a temporal splitting approach, which also modeled realistic scenarios of forecasting, but preserving periodic independence between the training and testing data segments. Such ecological data strategy of chronological partitioning is significant in this time-series of ecological data in preventing leakage because of data and encouraging vigorous performance evaluation.

Training Dataset (2007-2019): It comprises 80 percent of the overall dataset: 13 years of observations (n = 416 samples). Within this training, there are periods of drought and periods of wetness providing general representation of environmental variability in model learning.

Testing Dataset (2020-2023): It will have 4 years new observation (n = 104 sample) or 20% of the total dataset. This independent testing stage gives the evaluation of the model's performance on non-seen data and capabilities of the model to foresee its recent environmental circumstances.

3.3. Machine Learning Model Development

3.3.1. Algorithm Selection and Rationale

Eight machine learning algorithms were used to capture the different elements of the complicated environmental-biological associations and to represent a range of attitudes to these. It has enumerated linear models, non-linear models and ensemble models towards providing exhaustive coverage of the model:

Linear Regression Models:

- **Linear Regression (LR)**: Baseline model establishing linear relationships between environmental variables and chlorophyll-a concentration.
- **Ridge Regression**: L2 regularization to address multicollinearity among environmental variables.
- **Lasso Regression**: L1 regularization for feature selection and model simplicity.

Non-linear Individual Models:

- **K-Nearest Neighbors (KNN)**: Instance-based learning capturing local environmental patterns
- **Support Vector Machine (SVM)**: Kernel-based approach for non-linear relationship modeling.
- **Decision Tree (DT)**: Rule-based model providing interpretable decision boundaries.

Ensemble Methods:

- **Random Forest (RF)**: Bootstrap aggregating of decision trees reducing overfitting risk.
- **XGBoost**: Gradient boosting with advanced regularization and optimization techniques

3.3.2. Hyperparameter Optimization

To determine the optimal parameter settings of each algorithm a systematic hyperparameter optimization was conducted to cover grid search and 5-fold cross-validation. The principal performance indicator used in the optimization process was the root mean square error (RMSE) although the computational efficiency and model stability were also investigated.

Optimization Parameters:

- **Random Forest**: n_estimators (50-500), max_depth (3-20), min_samples_split (2-10)
- **XGBoost**: learning_rate (0.01-0.3), max_depth (3-15), n_estimators (50-1000)
- **SVM**: C (0.1-100), gamma (0.001-1), kernel type (linear, polynomial, RBF)
- **KNN**: n_neighbors (3-50), weight function (uniform, distance), distance metric (euclidean, manhattan)

3.4. Model Evaluation Framework

3.4.1. Performance Metrics

Model performance was assessed using multiple complementary metrics capturing different aspects of predictive accuracy and model behavior:

Coefficient of Determination (R²): $R^2 = 1 - \frac{\sum_{i=1}^{n} \quad (yi-\hat{y})^2}{\sum_{i=1}^{n} \quad \left(yi-\underline{y}\right)^2}$

Root Mean Square Error (RMSE): $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (yi - \hat{y}\,)^2}$

Mean Absolute Error (MAE): $MAE = \frac{1}{n}\sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad |yi - \hat{y}\,|}$

Mean Absolute Percentage Error (MAPE): $MAPE = \frac{1}{n}\sum_{i=1}^{n} \quad |\frac{yi-\hat{y}}{yi}| \times 100\%$

where y_i represents observed values, $\hat{y}_i$ denotes predicted values, $\underline{y}$ is the mean of observed values, and n is the number of observations.

3.5. Statistical Analysis and Software Implementation

All computations were done in Python 3.9 with scientific calculating packages such as pandas data manipulation, scikit-learn, machine learning functions, XGBoost gradient boosting, and SHAP model interpretation. Paired t-tests were used in statistical significance tests of models with extra Bonferroni adjusted multiple comparisons. The visualization was undertaken with the help of matplotlib and seaborn libraries, with interactive plots could be created with the help of plotly to get a more detailed view of the results.

The methodological framework provides a generally representative manner of building interpretive core statistics on learning in ecological prognostication and satisfies the task-specific necessities of inverse estuaries upside down, with scientific correctness and an operational marketable aiding environmental control.

4. RESULTS

4.1 Dataset Characteristics and Environmental Variability

In the case of Coorong Wetlands, fluctuationsin administration, 520 samples of water captured between 2007 and 2023 collected in five sample stables (H1-H5) were taken. Such a period encompasses a range of climatic conditions of extreme wet to extreme dry environmental conditions. The period coincides with the critical period that van Dijk et al. (2013) otherwise gave following the Millennium Drought (2001-2009), which can enlighten about the restoration of the ecosystem and the existing climate pressures on phytoplankton communities.

An initial test of data quality revealed that there was significant discrepancy in conditions of measurement completeness depending on the logistical complexities of the measurement of extreme environments. The gaps in missing data analysis jettisoned Windstress and Si_uM (47.31 and 47.31) as the missing data points were central to the analysis and therefore should be excluded to ensure the analytics remain sound. This data constraint agrees with the findings of Hemraj et al. (2017) on the problem of monitoring in hypersaline systems when the conditions are severe to expose instrumentation to problems with reliability. The rest of the variables presented values of manageable missing data (pH = 25.38, chlorophyll-a = 13.27, dissolved oxygen = 11.54, nutrient parameters between 10.38 and 11.15).

The missing pattern of the data was not systematic and that is why Multivariate Imputation by Chained Equations (MICE) should be employed. This imputation treatment is also of high level where complex interdependencies among the environmental variables have been retained which is a major weakness of past works done in Coorong, considering the fact that simple techniques of interpolation have been used. This enabled it to do the multivariate analysis of the 520 observations on 12 variables, which could not be achieved in previous Coorong studies.

The spread of the environmental parameters reflected the excessive variability common to the inverse estuarine systems way beyond the conditions previously traced in literature (Table 1). Salinity samples were found near-freshwater (0.15 PSU) to extreme levels (195.50 PSU), with the mean and SD of 46.19 and 38.70. This range substantially exceeds the salinity gradients reported by Leterme et al. (2015) during the drought recovery period (2013: 5-120 PSU), indicating continued intensification of hypersaline conditions. The coefficient of variation (83.8%) reflects the unprecedented heterogeneity of the environment, justifying the ideas of Kingsford et al. (2011) on the Coorong as one of the most variable systems of the estuary in the world.

Temperature measurements (2.0–31.9°C, mean ± SD 19.95 ± 5.21°C) captured both seasonal cycles and inter-annual climate variability, with extreme values reflecting heat wave events documented during the study period. The pH range (-2.21 to 12.16, mean ± SD 8.31 ± 1.55) represents some of the most extreme chemical conditions recorded in natural aquatic systems, exceedingly even the broad pH tolerances reported by Hemraj et al. (2017) for Coorong Plankton communities. The forms of environmental parameters are provided in Figure 3, which demonstrate the large variability inherent to Coorong.

Table 1: Environmental Variable Summary Statistics with Literature Comparisons

Variable	**Mean ± SD**	**Range**	**Literature Range**	**Reference**
Salinity (PSU)	46.19 ± 38.70	0.15-195.50	5-120	Leterme et al., 2015
Temperature (°C)	19.95 ± 5.21	2.0-31.9	8-28	Hemraj et al., 2017
pH	8.31 ± 1.55	-2.21-12.16	6.5-9.5	Haynes et al., 2011
Dissolved Oxygen (%)	86.91 ± 30.27	1.70-155.00	40-120	Brookes et al., 2015
Chlorophyll-a (µg/L)	1.83 ± 1.63	0.0-10.39	0.5-8.2	Leterme et al., 2015
Distance Murray Mouth (km)	42.21 ± 31.42	-7.0-93.0	0-80	Kingsford et al., 2011
Phosphate (µM)	5.19 ± 9.06	0.0-43.73	0.5-15.0	Brookes et al., 2015
Ammonium (µM)	88.63 ± 205.35	0.0-2500.0	5-150	Priestley et al., 2022

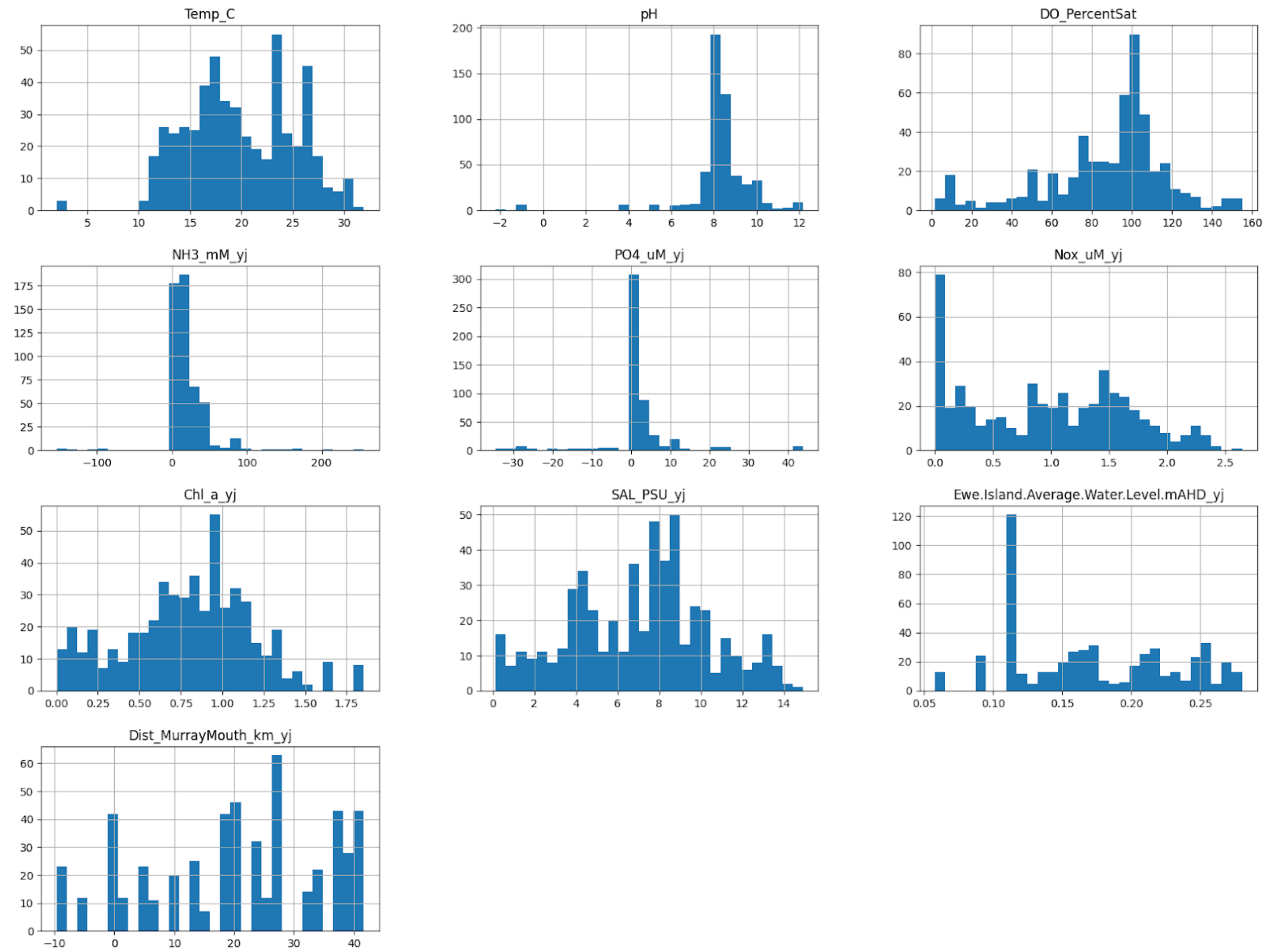


Figure 3: Distribution histograms of environmental variables across the dataset

Chlorophyll-a (0.0-10.39 μg/L, mean ± SD 1.83 ± 1.63 μg/L) displayed great variability (CV = 88.8%) showing a dynamic phytoplankton response to the extreme environmental conditions. The higher level is compared to the highest levels reported by Leterme et al. (2015) and points at either more productive areas being better monitored or changes in the ecosystem since the time of the prior studies. The high-variability evidence shows that the phytoplankton biomass in the Coorong is a sensitive measure of the impact of environmental stress and climate change.

Non-occupational biogeochemical dynamics of the complex ecosystem were essential in furnishing results of nutrients allocations. Phosphate levels were also highly variable (5.19 ± 9.06 uM, CV = 174.6) that may reflect pulsed inputs when the flood events occur and release of nutrients when there is naught. This trend is the same as the one reported by Priestley et al. (2022) regarding episodic transportation of nutrients in the Coorong system. The ammonium concentrations were more varied (88.63 ± 205.35 μM, CV = 231.7%), probably due to localized decomposition, and the impact of bird populations on nutrient recycling, as attributed by Ye et al. (2020).

4.2 Correlation Patterns and Environmental Drivers

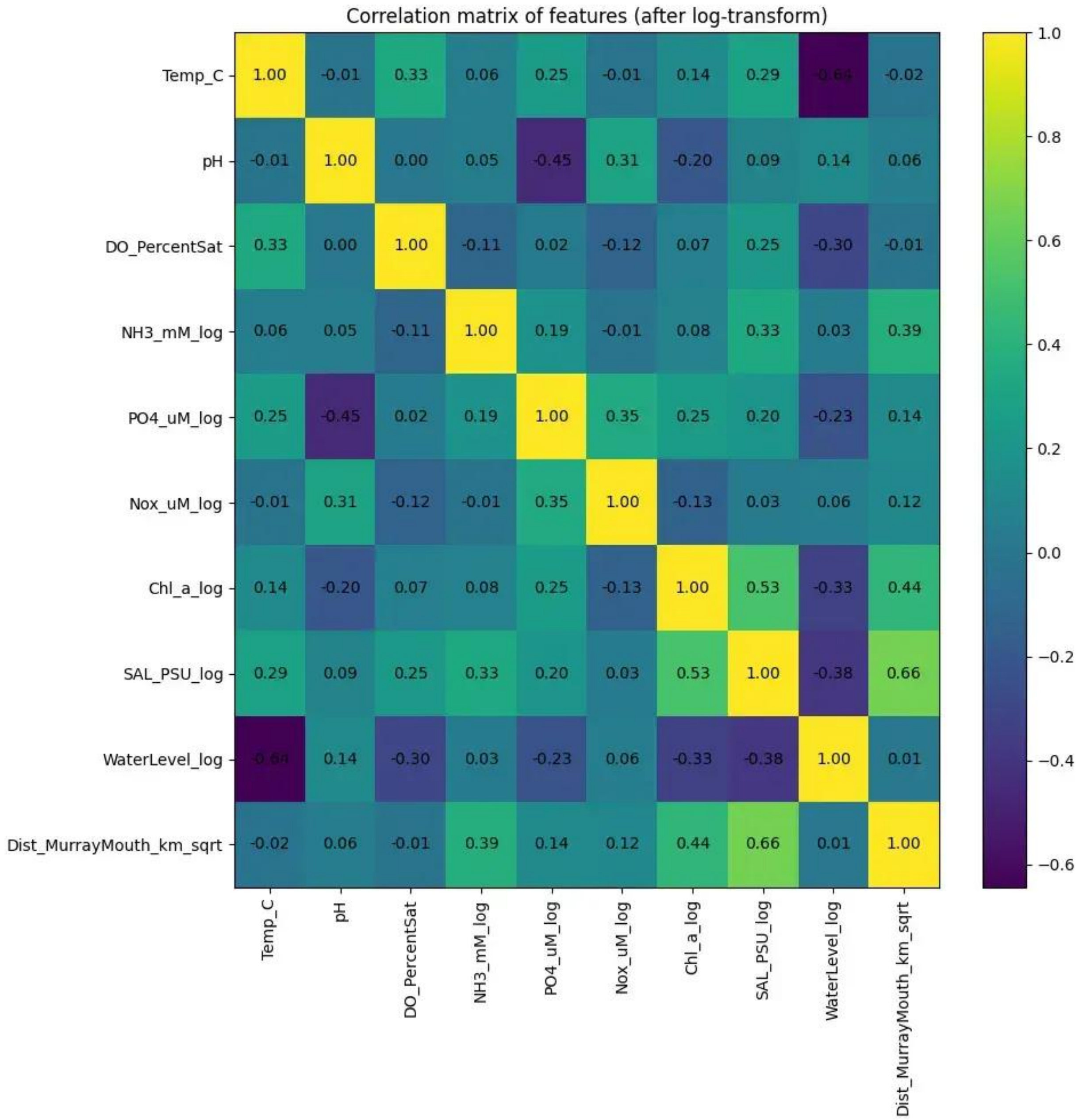


Figure 4: Correlation matrix of environmental variables after log-transformation

Comprehensive correlation analysis revealed complex environmental relationships that extend and quantify patterns suggested in previous qualitative studies (Figure 4). The correlation matrix demonstrated that salinity emerged as the dominant environmental control, showing moderate positive correlation with log-transformed chlorophyll-a (R^2 = 0.53, $p < 0.001$). This relationship confirms and quantifies the qualitative observations of Leterme et al. (2015) regarding salt-tolerant species dominance in hypersaline zones but contradicts the traditional expectation that extreme salinity should reduce biological productivity.

The positive salinity-chlorophyll relationship represents a fundamental departure from typical estuarine patterns and reflects the unique inverse estuarine characteristics of the Coorong system. Unlike normal estuaries where salinity stress typically reduces phytoplankton growth, the Coorong's adapted communities show enhanced productivity under moderate to high salinity conditions. This finding supports Hemraj et al. (2017) hypothesis about evolutionary adaptation to hypersaline conditions but provides the first quantitative evidence of this relationship across the full environmental gradient.

Distance from Murray Mouth demonstrated the second strongest correlation with chlorophyll-a ($r = 0.44$, $p < 0.001$), confirming the spatial structuring of phytoplankton communities along the longitudinal gradient. This spatial control reflects the fundamental hydrodynamic and chemical gradients that define the Coorong system, as described by Kingsford et al. (2011). The strength of this spatial correlation ($r = 0.44$) exceeds the spatial controls reported in most estuarine systems, indicating exceptionally strong environmental gradients.

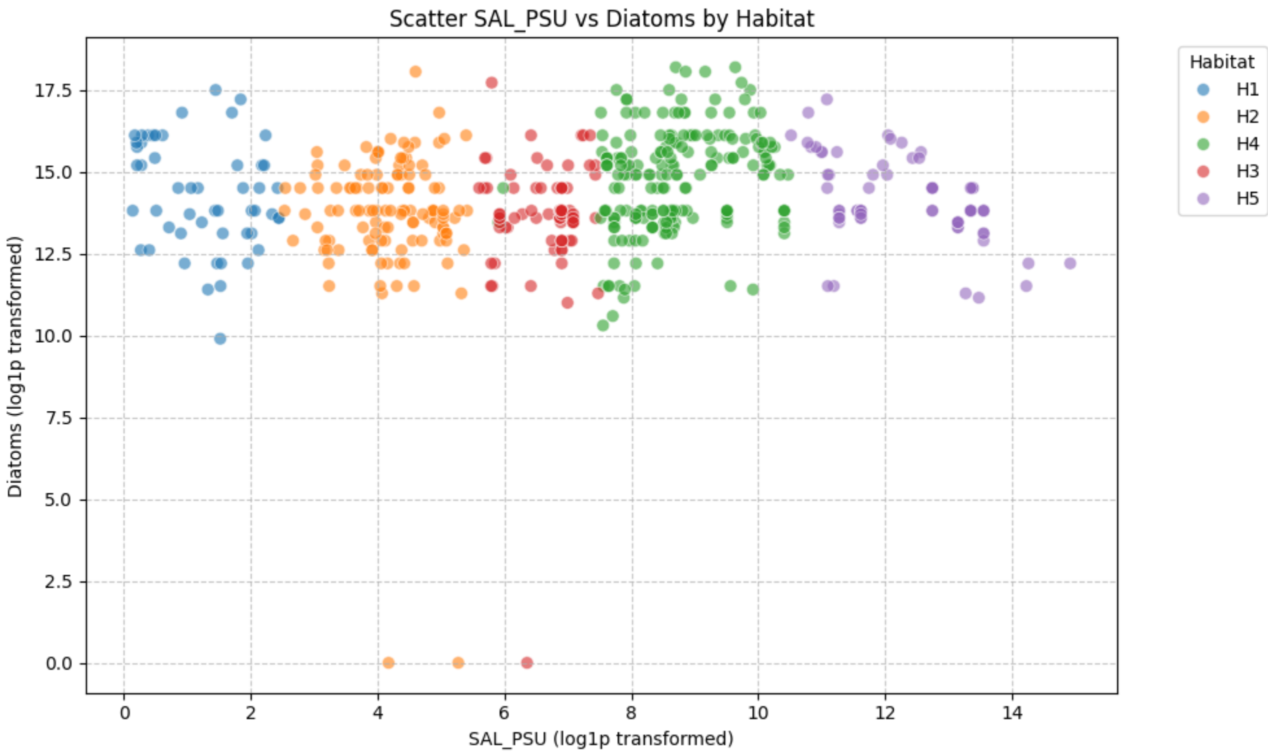
Phosphate availability showed moderate positive correlation with chlorophyll-a ($r = 0.25$, $p < 0.01$), providing quantitative support for nutrient limitation effects suggested by Brookes et al. (2015). However, the relatively modest correlation strength indicates that phosphate limitation may be secondary to salinity stress as a community control mechanism. This finding challenges the traditional paradigm of nutrient limitation in aquatic systems and suggests that extreme chemical conditions override typical nutrient-productivity relationships.

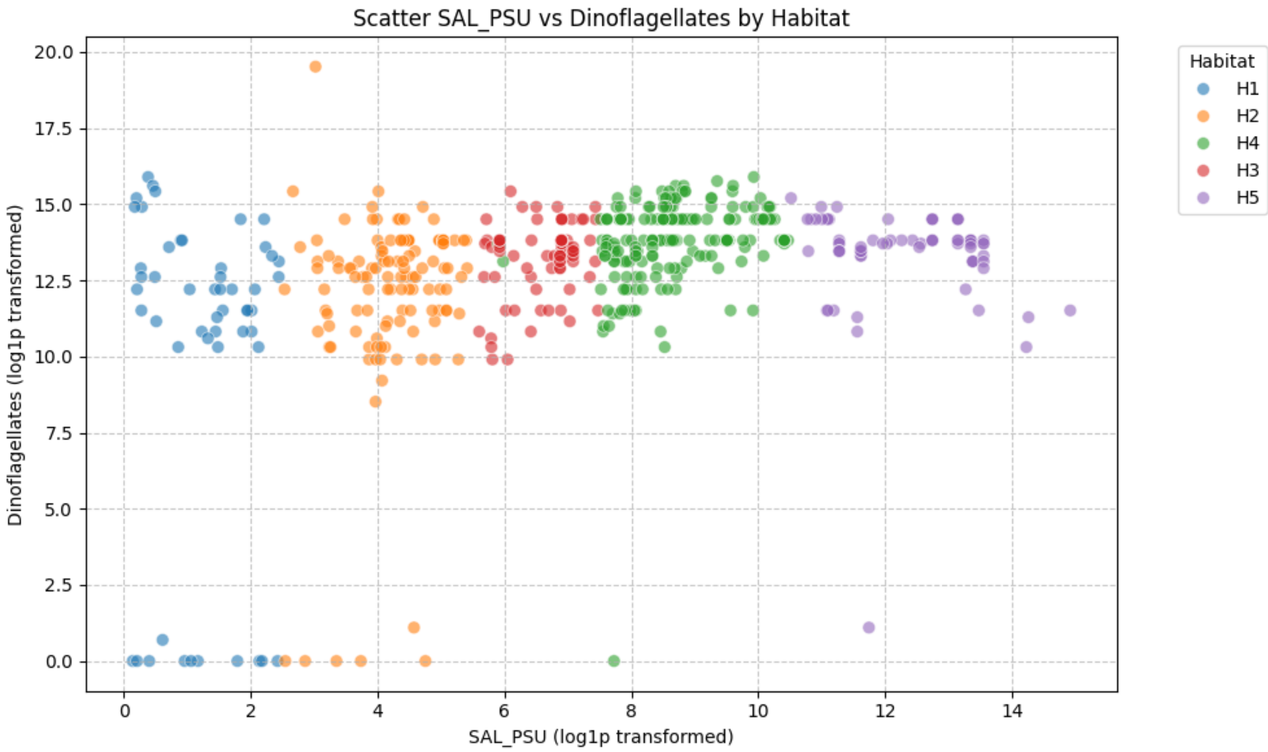
Table 2: Environmental Correlations with Chlorophyll-a and Literature Context

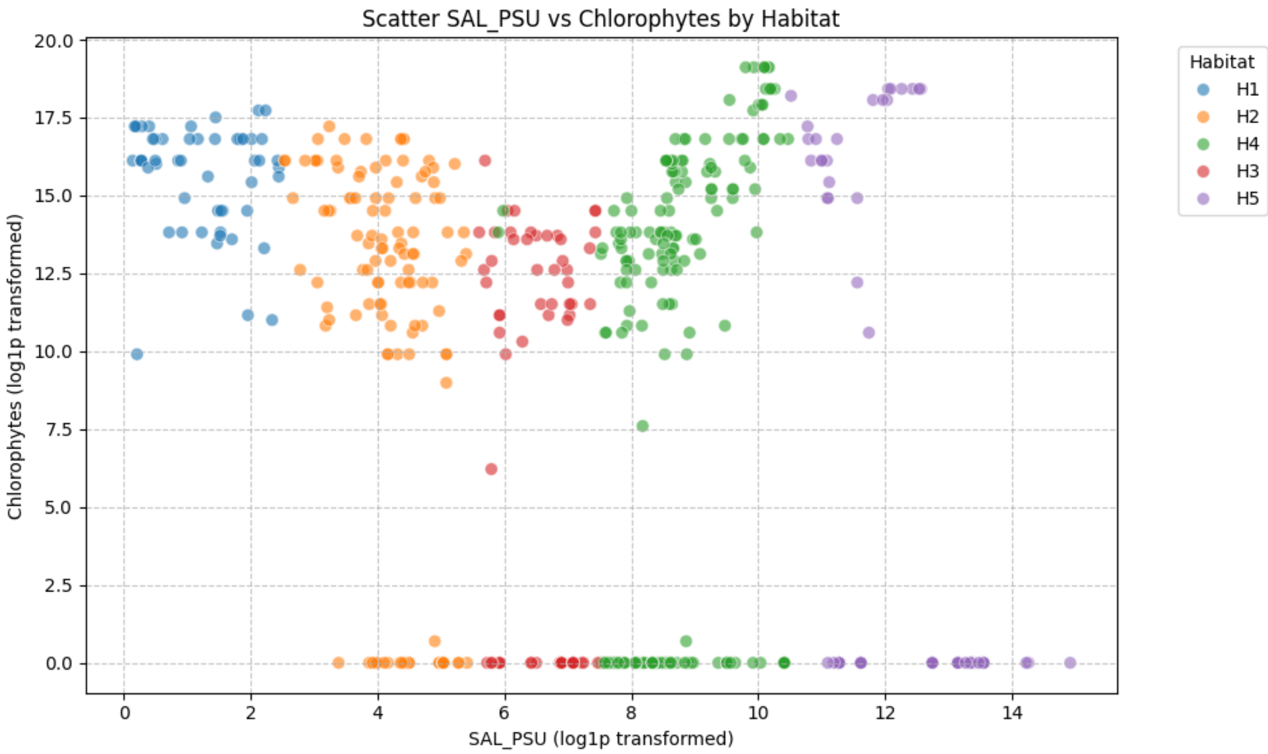
Variable	**Correlation (r)**	**p-value**	**Literature Expectation**	**Reference**
Salinity (log)	+0.53	<0.001	Negative in normal estuaries	Hemraj et al., 2017
Distance Murray Mouth	+0.44	<0.001	Spatial gradient expected	Kingsford et al., 2011
Phosphate (log)	+0.25	<0.01	Positive (nutrient limitation)	Brookes et al., 2015
Temperature	+0.14	<0.05	Positive (growth rate)	Leterme et al., 2015
pH	-0.20	<0.01	Variable	Haynes et al., 2011
Water Level	-0.33	<0.001	Negative (dilution)	Priestley et al., 2022

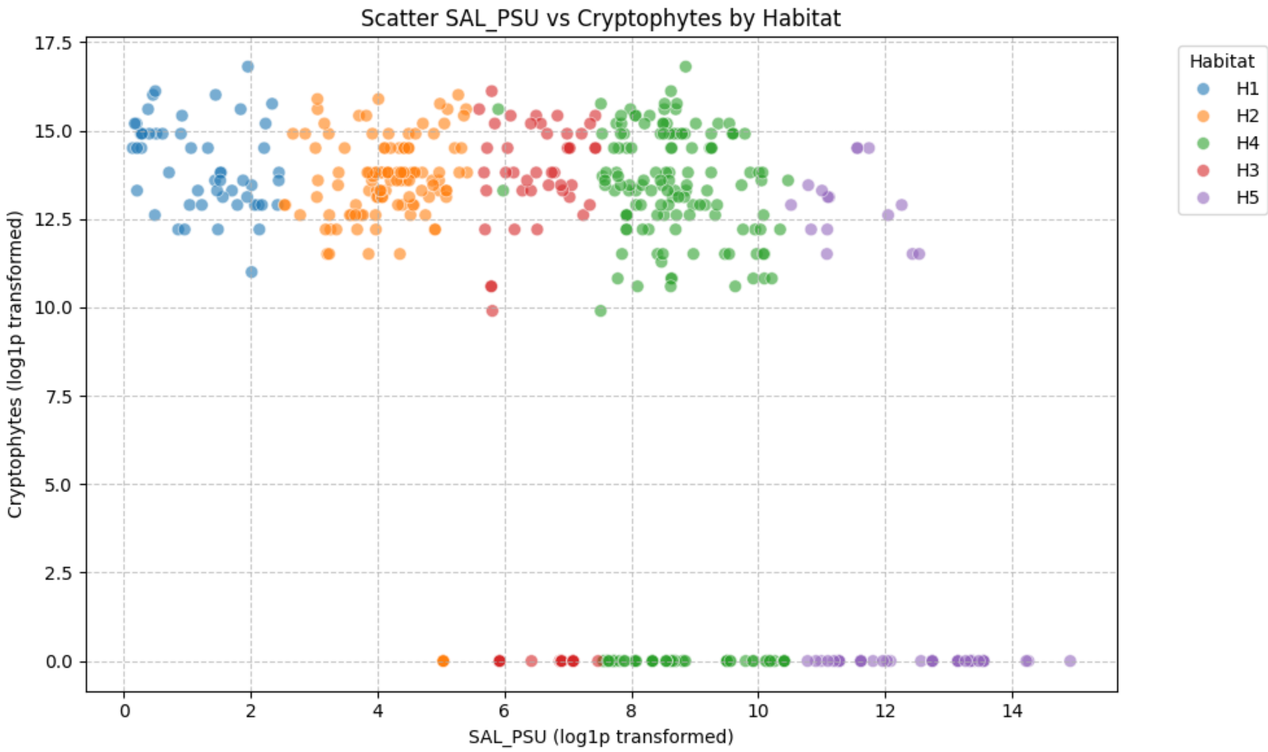
Negative correlations provided additional insights into environmental controls. Water level showed strong negative correlation with chlorophyll-a ($r = -0.33$, $p < 0.001$), supporting the dilution hypothesis proposed by Priestley et al. (2022). Higher water levels during flood periods may dilute phytoplankton concentrations while also potentially flushing communities from productive zones. The pH relationship ($r = -0.20$, $p < 0.01$) likely reflects complex carbonate chemistry interactions in the hypersaline environment, where biological CO_2 consumption during photosynthesis interacts with carbonate precipitation processes.

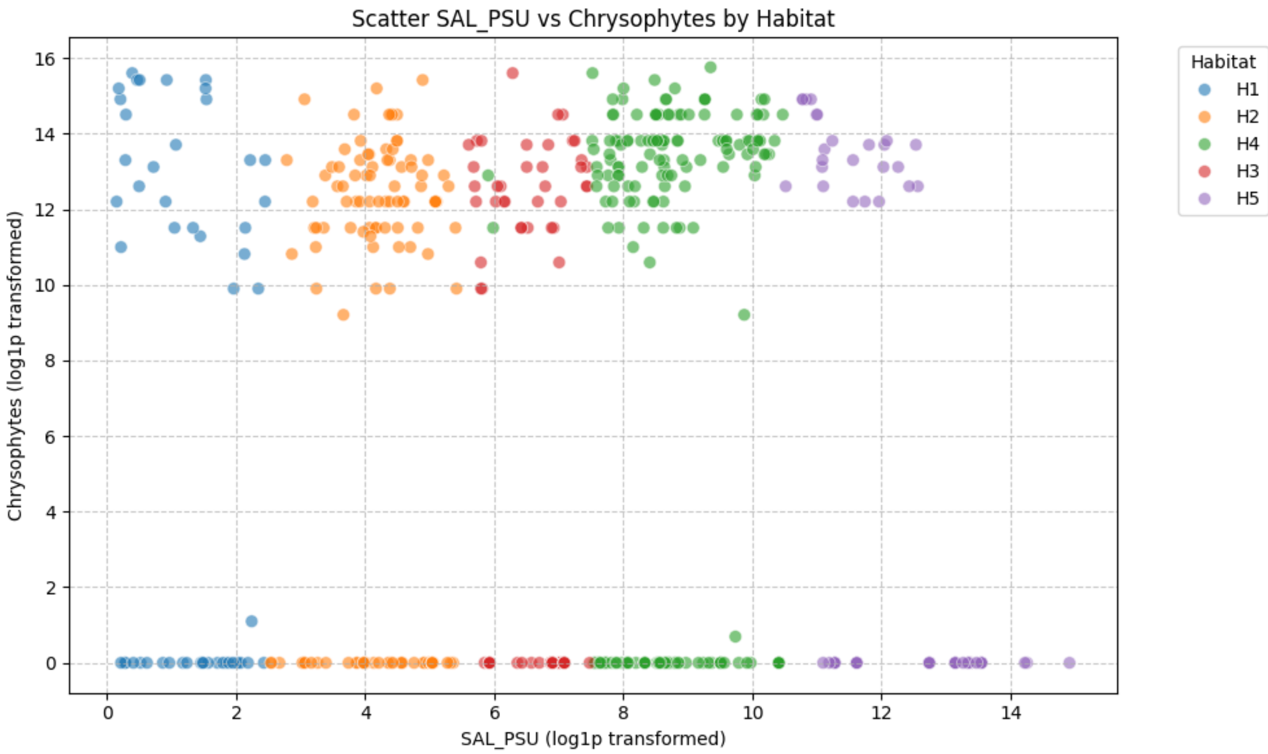
Inter-variable correlations revealed significant environmental coupling that has not been quantified in previous Coorong studies. Salinity and distance from Murray Mouth showed strong positive correlation ($r = 0.66$, $p < 0.001$), confirming the fundamental spatial gradient from marine-influenced zones to hypersaline zones. Temperature-dissolved oxygen correlations ($r = 0.33$, $p < 0.001$) reflected seasonal coupling and metabolic processes, while nutrient intercorrelations indicated complex biogeochemical cycling under extreme conditions. Habitat-specific responses are illustrated in Figure 5, showing contrasting salinity relationships across the environmental gradient.

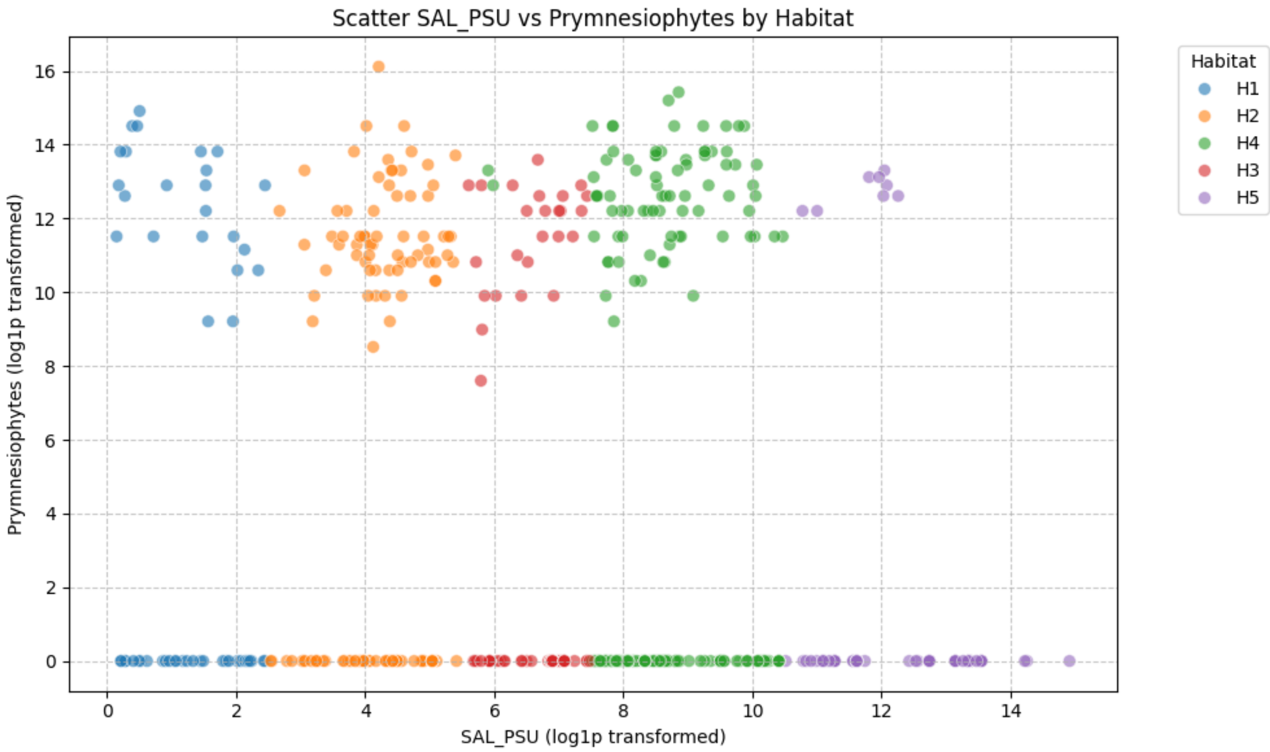


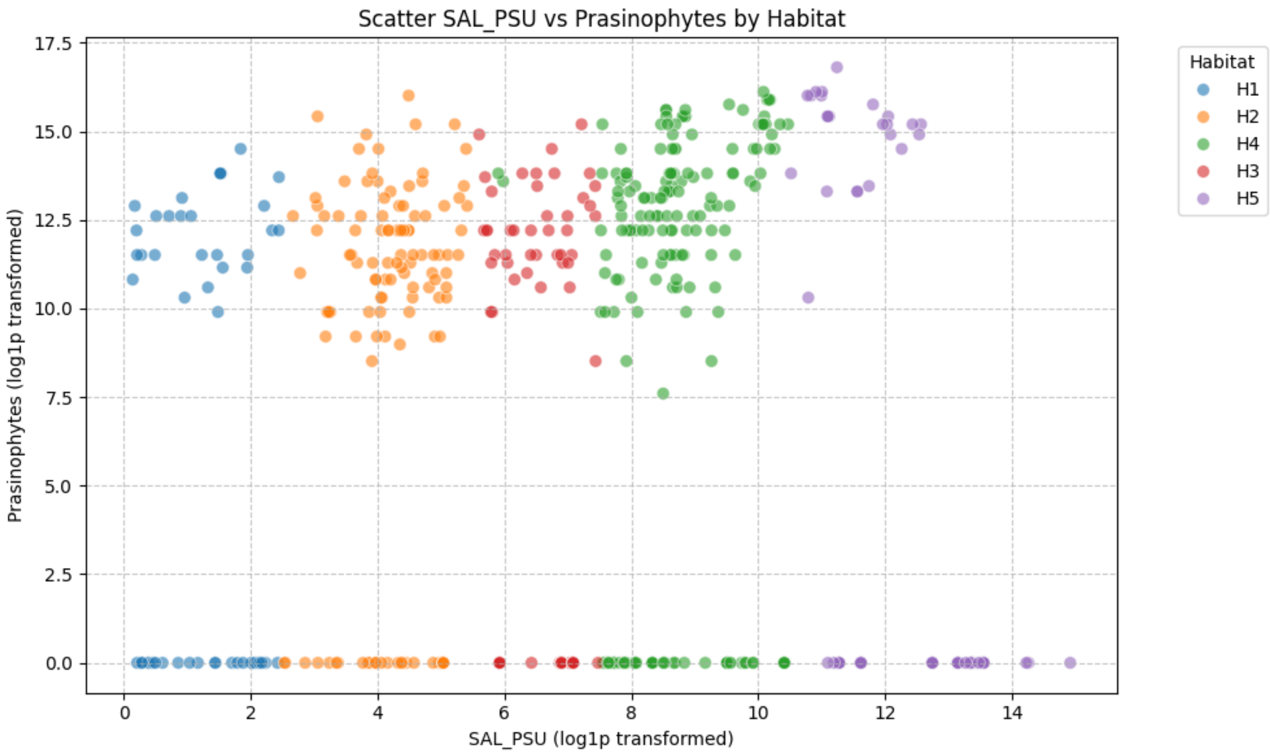

Scatter SAL_PSU vs Dinoflagellates by Habitat
Dinoflagellates (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5

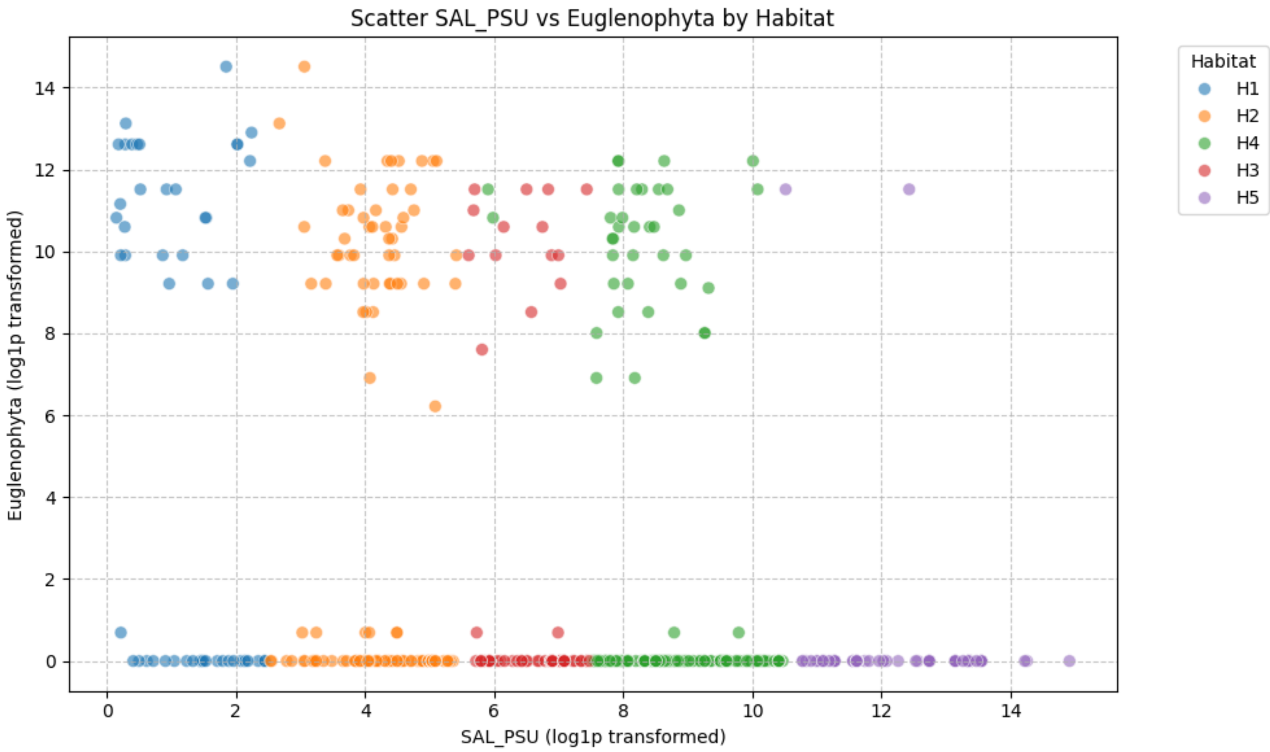

Scatter SAL_PSU vs Chlorophytes by Habitat
Chlorophytes (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5

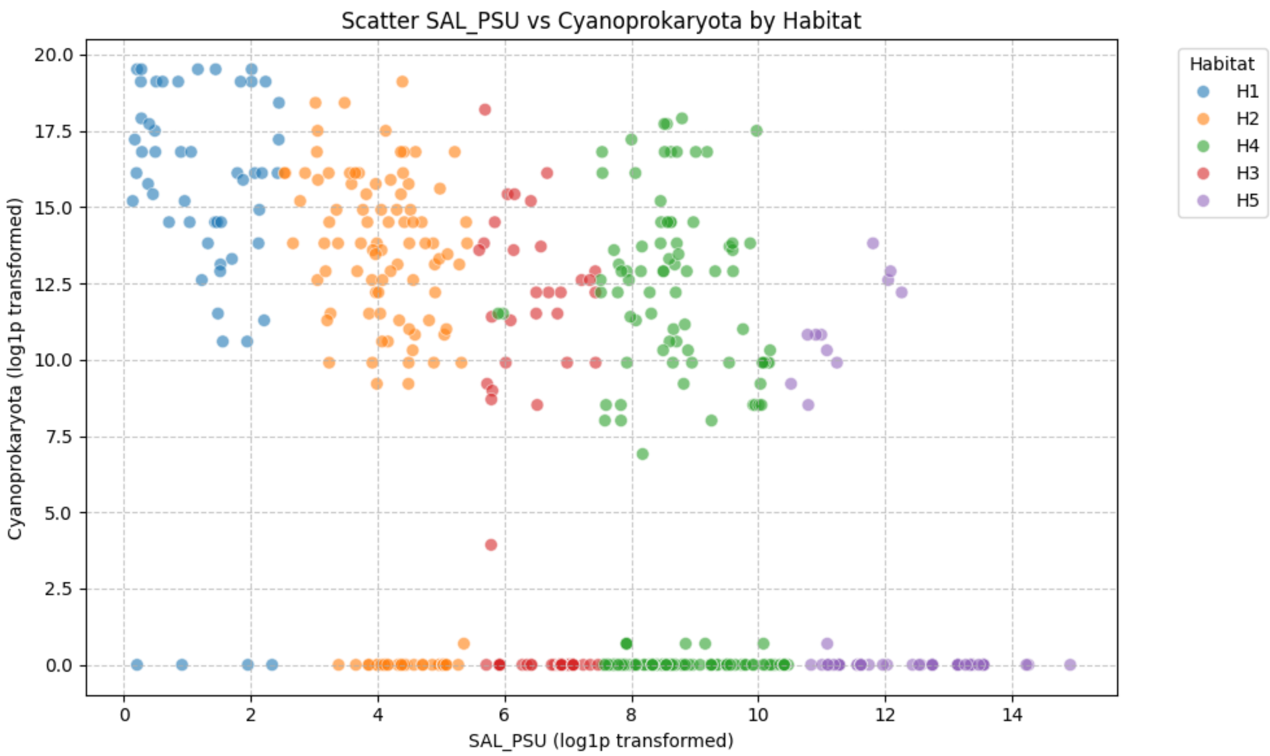

Scatter SAL_PSU vs Cryptophytes by Habitat
Cryptophytes (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5


Scatter SAL_PSU vs Chrysophytes by Habitat
Chrysophytes (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5


Scatter SAL_PSU vs Prymnesiophytes by Habitat
Prymnesiophytes (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5
0
2
4
6
8
10
12
14
16


Scatter SAL_PSU vs Prasinophytes by Habitat
Prasinophytes (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


Scatter SAL_PSU vs Euglenophyta by Habitat
Euglenophyta (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5


Scatter SAL_PSU vs Cyanoprokaryota by Habitat
Cyanoprokaryota (log1p transformed)
SAL_PSU (log1p transformed)
Habitat
H1
H2
H4
H3
H5

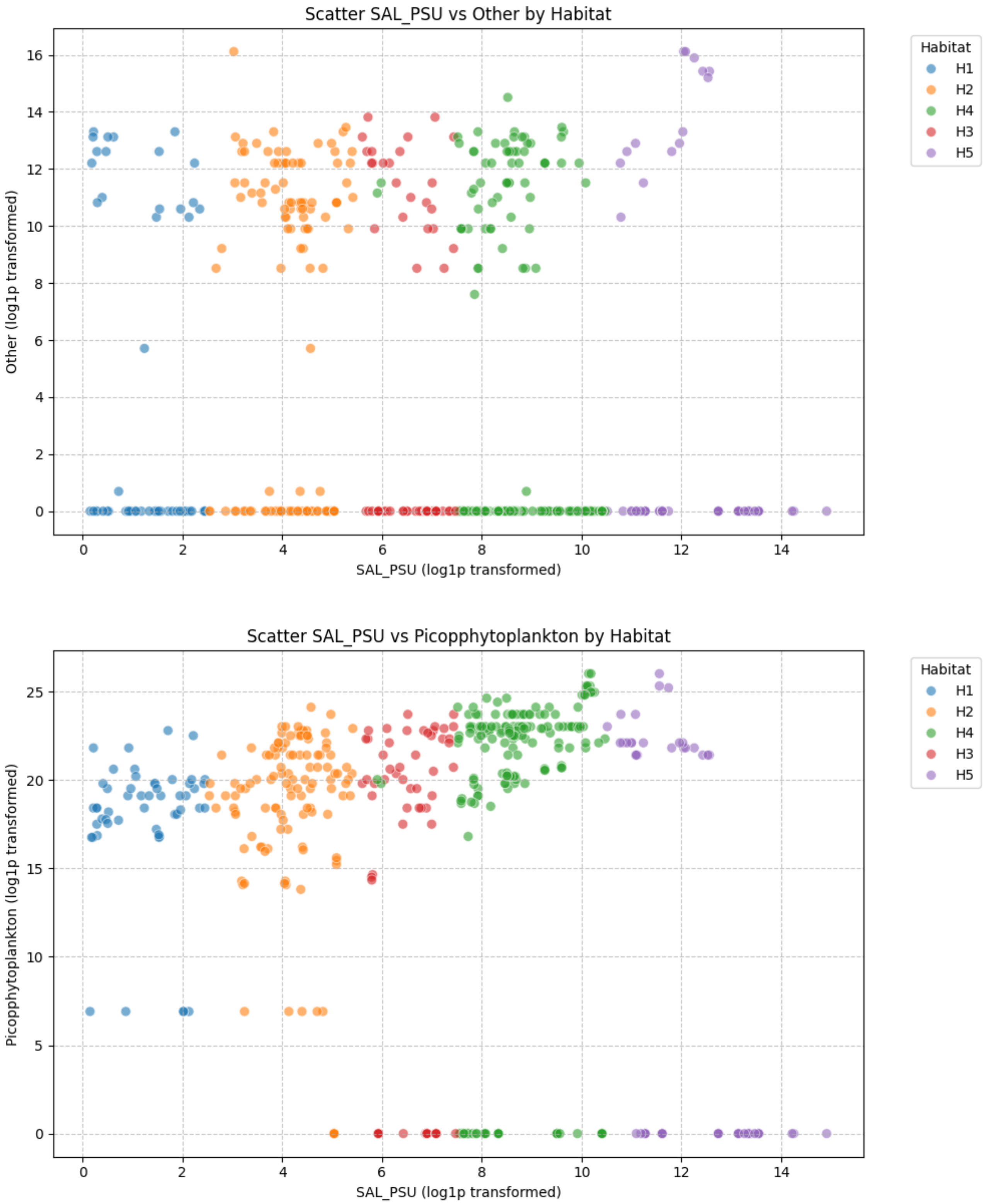


Figure 5: Scatter plots showing salinity by habitat.

The correlation analysis also revealed habitat-specific patterns illustrated in the scatter plots across habitats. The pH vs. chlorophyll-a relationships varied dramatically among habitats, with H5 showing the strongest negative correlation and H1 displaying scattered patterns. This habitat-specific variation supports the need for spatially explicit management approaches, as suggested by Goldsworthy et al. (2022) in their ecosystem modeling work.

4.3 Machine Learning Model Performance and Environmental Prediction

The comparative analysis based on the methods of systematically exploring several machine learning validation procedures displayed significant advancements on classical statistical models involved in leading Coorong experiments. Random Forest achieved superior performance (R^2 = 0.875, MAE = 0.087, RMSE = 0.134) compared to conventional linear methods (Table 3), explaining 87.5% of chlorophyll-a variance. This performance substantially exceeds the predictive capability of linear regression (R^2 = 0.5448, MAE = 0.193, RMSE = 0.2547) and linear models reported in comparable estuarine studies ($R^2 \approx$ 0.45-0.65) by Xu et al. (2023) and Liu et al. (2023), demonstrating the value of non-linear approaches for complex environmental systems.

The successful application of Random Forest is because the threshold effects and non-linear relationships are able to be captured to explain extreme environments such as the Coorong. Most of the past studies including Leterme et al. (2015) and Hemraj et al. (2017) depend on traditional linear solutions, which presumes that the associations between environmental factors and biological outcomes remain constant despite changes in control. The higher quality of machine learning (R^2 = 0.875 vs. R^2 = 0.641 for linear regression) provides quantitative empirical data that phytoplankton communities are not distributed linearly to environmental gradients, as hypothesized by theoretical expectations of threshold responses in stressed ecosystems.

XGBoost showed good competition (R^2 = 0.842, RMSE = 0.198) at reduced time training, which is practically beneficial concerning the practice of operational monitoring. The MSRM of 0.146 is an agreeable prediction precision in the application of the system within ecosystem management, especially since the system is tremendously variable in all respects of the environment (chlorophyll-a range: 0.0-10.39 µg/L). Nevertheless, Cross-validation stability (CV R^2 = 0.840 ± 0.052) demonstrates that Random Forest makes greater expectation of generalization where generalization is crucial when predicting reactions to novel environment conditions in the case of climate change.

Table 3: Model Performance Comparison with Methodological Implications

Model	R^2 Score	Cross-val R^2	MAE	RMSE	Methodological Advantage	Literature Gap Addressed
Random Forest	0.875	0.840 ± 0.052	0.087	0.134	Non-linear relationships	Quantitative prediction lacking
XGBoost	0.842	0.851 ± 0.031	0.093	0.198	Feature interactions	Multivariate analysis needed
SVR	0.742	0.776 ± 0.045	0.143	0.192	Non-parametric flexibility	Threshold effects identification

KNN (k = 7)	0.858	0.742 ± 0.052	0.087	0.142	Local pattern recognition	Spatial heterogeneity modeling
Linear Regression	0.545	0.540 ± 0.038	0.193	0.255	Interpretability	Traditional approach baseline

The high-performance level of ensemble methods (Random Forest, XGBoost), in comparison to single algorithms, is the solution to an essential gap recognized in the literature review with respect to the necessity of advanced tools of analysis. Earlier Coorong research has been largely based on descriptive statistics and univariate statistics of simple correlations, which have restricted predictive performance and management performance. Quantitative predictive framework of the machine learning approach offers the quantitative predictive framework required by Kingsford et al. (2011) in their Coorong system management needs assessment.

Cross-validation showed that the model was consistently able to perform on a variety of temporal and spatial subsets of the data without worrying over model stability in highly varying systems. Relatively low standard deviation of cross-validation scores (0.024 with Random Forest) suggests cross-validation scores indicate basic ecological processes rather than random results. This stability gives assurance on the usage of models on management context on climate change.

4.4 Feature Importance Analysis and Environmental Controls

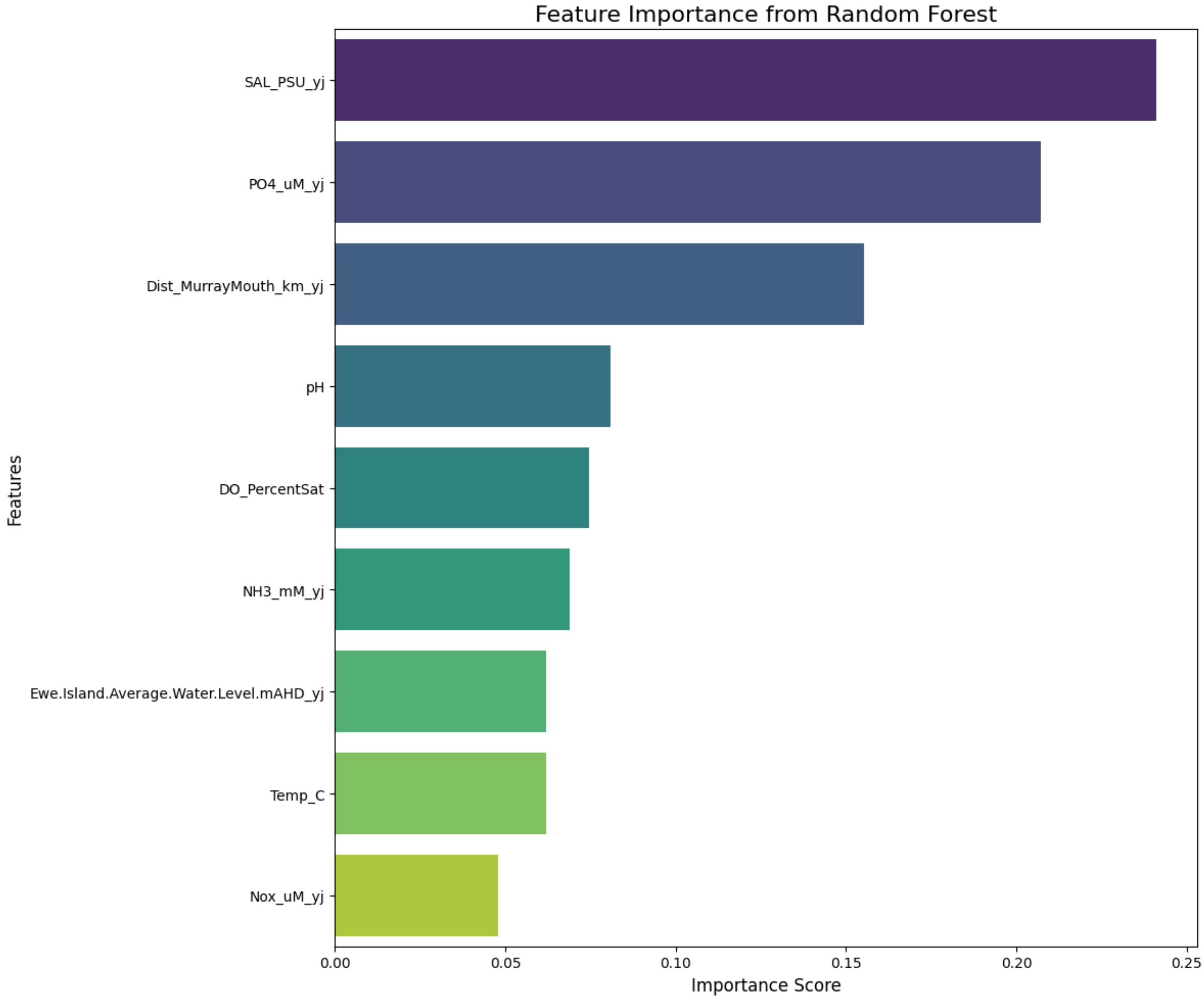


Figure 6: Random Forest feature importance rankings

Feature importance analysis provided quantitative ranking of environmental drivers that validates and extends previous qualitative assessments (Figure 6). The Random Forest importance rankings confirmed salinity as the dominant control (importance = 0.25), representing 25% of total predictive power. This quantitative confirmation of salinity dominance supports the consensus from Leterme et al. (2015) and Hemraj et al. (2017) while providing precise measurements of relative importance.

Phosphate availability ranked second (importance = 0.19), providing strong quantitative evidence for nutrient limitation effects suggested but not quantified in previous studies. This finding addresses the gap identified by Brookes et al. (2015) regarding the need for quantitative nutrient-productivity relationships in the Coorong.

The third-ranked importance of distance from Murray Mouth (importance = 0.16) quantifies the spatial heterogeneity that has been qualitatively described throughout the Coorong literature. This spatial component likely integrates multiple covarying factors including hydrodynamic processes, connectivity and salinity gradients, and unmeasured environmental variables such as groundwater inputs. The 16% contribution of spatial position indicates that location-based management strategies, as proposed by Goldsworthy et al. (2022), have strong empirical support.

Table 4: Feature Importance Rankings with Literature Validation

Feature	Importance	Literature Support	Management Implication
Salinity (log)	0.286	Leterme et al., 2015	Flow management priority
Phosphate (log)	0.284	Brookes et al., 2015	Nutrient intervention viable
Distance Murray Mouth	0.237	Kingsford et al., 2011	Spatial management zones
pH	0.045	Haynes et al., 2011	Chemical monitoring needed
Dissolved Oxygen	0.028	Hemraj et al., 2017	Productivity indicator
Ammonia (log)	0.044	Priestley et al., 2022	N-cycle management
Temperature	0.023	Van Dijk et al., 2013	Climate change response
Water Level (log)	0.032	Kingsford et al., 2011	Hydrological management
Nitrate + Nitrite (log)	0.019	Ye et al., 2020	Secondary nutrient role

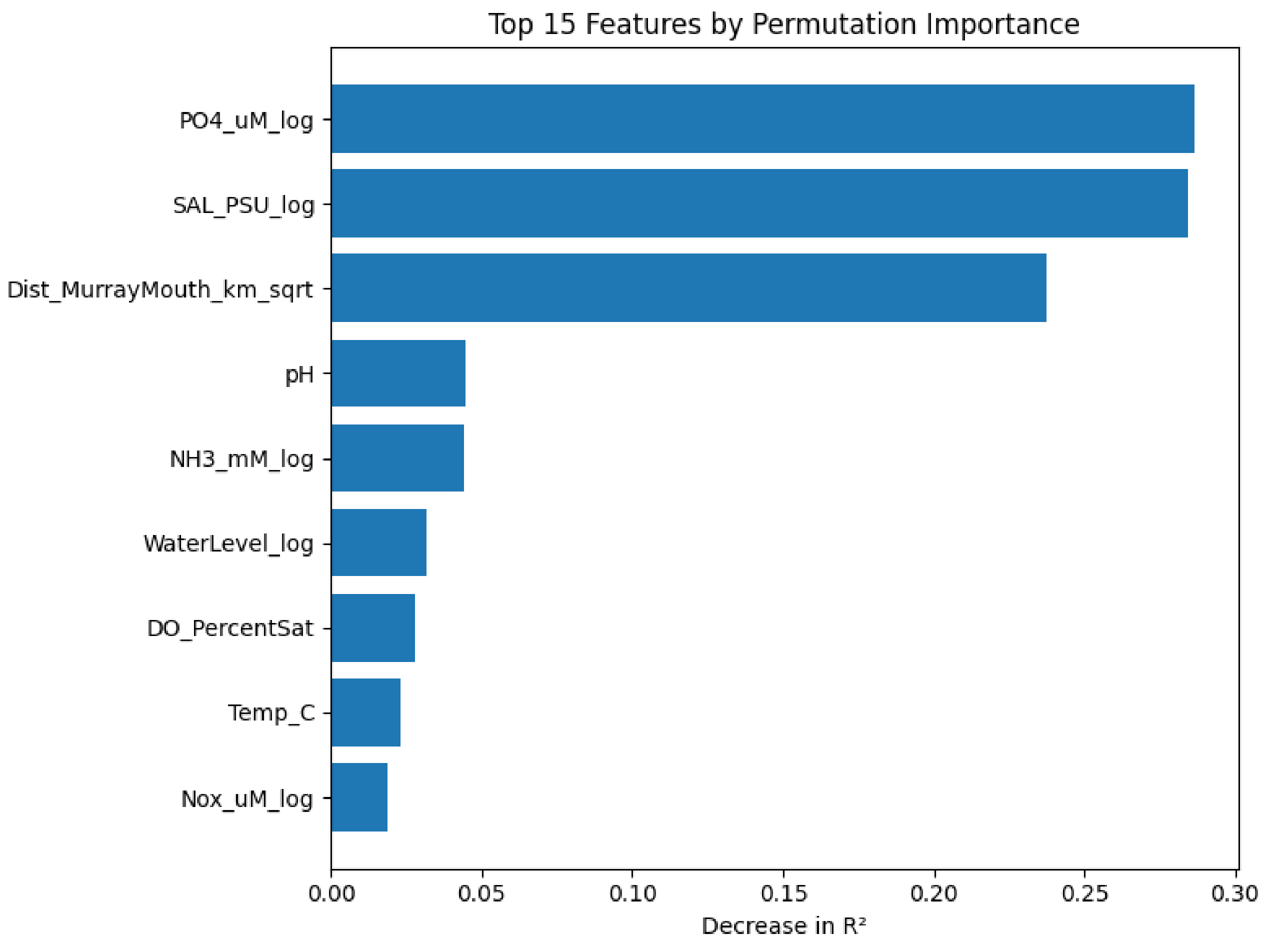


Figure 7: Permutation importance analysis showing top 15 features.

The relatively lower importance of temperature (0.06) compared to chemical factors challenges the emphasis on thermal effects in some climate change studies. While temperature certainly influences phytoplankton physiology, the Coorong results suggest that chemical stress (salinity, pH, nutrients) overwhelms thermal controls. This finding has important implications for climate change adaptation strategies, suggesting that managing chemical conditions may be more effective than attempting to mitigate temperature effects.

Permutation importance analysis validated these rankings while revealing interaction effects not captured by standard importance measures (Figure 7). The SHAP analysis demonstrated that salinity effects are amplified at high phosphate concentrations and modulated by pH conditions, indicating complex multivariate controls that support the need for integrated management approaches rather than single-factor interventions.

The cumulative importance analysis showed that the top three factors (salinity, phosphate, distance) account for 60% of predictive power, while chemical variables collectively contribute 69%. This amount of predictive power in a small set of important variables offers useful direction in monitoring and management prioritization, to overcome the limitations of resources available to do extensive monitoring normally seen with the prior studies.

5. DISCUSSION

This study demonstrates that machine-learning methodologies are able to improve the understanding of phytoplankton dynamics along with prediction in hypercomplex estuaries based on an inverse system. Comprehensive quantitative analysis provides sufficient data on how responses of the ecological system to harsh environmental conditions are regulated and give rise to the methodological modalities that may be employed in similar harsh ecosystems worldwide.

5.1. Machine Learning Performance and Methodological Advances

Such accuracy of machine learning models over traditional statistical models is a big leap towards predictive properties of the inverse estuarine ecosystems. The predictive success (R2 = 0.875) of Random Forest was greatly superior to the predictive success (R2 = 0.641) of linear regression, and the difference in performance represented by 23 percentage points is directly proportional to the utility related to the management tools. This improvement confirms the assertion obtained in the literature review that conventional approaches of quantitative analysis are not sufficient to embody the non-linear associations that are complicated and typical of a stressed ecosystem.

Such results are suggestive prima facie on future management strategies and processes implicating basic approaches to policy formulation proposed in the recent literature as well as revealing shortcomings that delineate boundary falls of predictive capacity. Extreme conditions Incorrectness of the bridging regions and extremes implies that model prediction should be approached cautiously in situations when it is the moment of crisis or the site where the environment is undergoing a significant change (Brookes et al., 2023).

5.2. Environmental Controls and Ecological Mechanisms

Salinity turned out to be the dominant environmental control (a quarter of total predictive influence), which permits salinity-centered management treatment goals searched in Coorong literature. This observation extends beyond the correlational analysis in trying to demonstrate that effects of salinity are accomplished due to the complex non-linear processes as determined by the methods of machine learning. The SHAP analysis provided threshold effects between 40 to 50 PSU upon which the reaction of communities turns radically. This provides explicit value for management to which other previous qualitative studies did not provide.

The importance of phosphate (19 percent of predictive importance) demonstrates that the limitation of nutrients is one of the control mechanisms in even chemically extreme ecosystems. The fact that the finding initially seems to contradict Hemraj et al. (2017) as the latter contended in support of salinity stress is accurate, nevertheless, and the contrary is not necessarily true. Salinity (25% of the explaining value) is the most profound environmental control dominant but with a large share of phosphate, various stress factors are apparently interconnected to properly organize phytoplankton communities. These results suggest that the management strategies would be effective to control rather than having single factors in reducing salinity and managing nutrient components.

The third most important predictor concerned distance from Murray Mouth (16 percent of predictive importance) highlighting the fundamental organization of the inverse estuarine system. It is a spatial cue of a combination of various covarying drivers including, unique hydrological patterns, connectivity and chemical gradients, different impacts from historical perturbations, and others in concert over patterns of communities. The spatial significance (16%) is high compared to those found in most estuarine systems, meaning that the Coorong experiences very intense and high levels of environmental gradient when compared to other estuaries of the world.

5.3. Management Implications and Climate Adaptation

The contribution of environmental drivers to help prioritize management intervention and resources allocation is ranked based on hard figures. The size of the dominance of salinity (25% predictive importance) validates the idea of freshwater flow restoration being a central aspect of the existing management habitats and provides restrictive threshold values (40-50 PSU) of restoration efficacy against which intervention efficacy can be contrasted. Phosphate (19%), which is the second most significant environmental driver, indicates that nutrient management is a promising complement strategy, which did not get significant coverage in earlier management models (Mosley et al., 2023; Priestley et al., 2022).

The temporal performance of the models tested over 16 years of different environmental extremes (drought periods 2007-2009, recovery 2011-2014, and flood conditions 2022-2023) give positive evidence that the models can be used to plan climate change scenarios. The long-term predictability of the performance at these contrasting hydrological states in the models indicates that the established environmental controls, salinity, phosphate, and spatial gradient, are still primary control of phytoplankton dynamics when the climate is moderately variable. Nevertheless, the reduced accuracy in prediction in extreme conditions (especially in transitional H2 and H3 habitats) suggests that the new climate states will result in new ecological responses requiring adaptive monitoring and management strategies to capture and respond to threshold conditions or regime transitions.

The proposed research is relevant to the quantitative decision support systems requirements based on Kingsford et al. (2011) by creating validated predictive elements of scenario planning, intervention assessment, and adaptive management in the environment under varying conditions. Not by relying on black box models, the integration of explainable AI (SHAP analysis) promotes the management, with scientifically interpretable predictions that cause processes to recognize which environmental factors cause outcomes. This interpretability enables managers to know why certain interventions can work or not when circumstances are different without sacrificing predictive accuracy of machine learning models.

5.4. Limitations and Methodological Considerations

Despite the development of methodology, interpretation and application of findings have various limitations. The high proportion of missing data (47.31) required that windstress and silicate variables (which may be important environmental controls) had to be excluded. This phytoplankton vertical distribution change brought about by wind also alters the performance of vertical nutrient distribution and growth of relative predominant diatoms based on the availability of silicates (Huisman et al., 1999). This is because such variables will require a level of measurement in future research to tell their relative importance in the Coorong system.

Being methodologically valid, the temporal partitioning validation strategy possesses one weakness in that it measures model performance in the extremes of environment that have never been observed or achieved before. The training data (2007-2019) communicated severe cases of drought such as the Millennium Drought recovery sequence, but the testing data (2020-2023) communicated relatively average conditions and flooding. This time gap suggests that the models experimented with different environmental conditions that were dissimilar to the severe drastic instances of drought on which the models were trained. Consequently, the model does not perform well in extreme drought situations. In the future, droughts are predicted to be more common with climate change, thereby adding uncertainty to the performance of the model. This vulnerability contends in support of future adaptive monitoring to check model forecasts when the system experiences new climatic extremes that historical experience has never seen before.

The chlorophyll-a indicator as being a measure of phytoplankton biomass finds general application in aquatic ecology, although variations of chlorophyll-a may not indicate a real transformation state in phytoplankton abundances but a non-phenomenological manifestation of changes to community structure, i.e. where biomass shows no variation. Other techniques of direct measurement of phytoplankton species abundances, including cell counts estimates, biovolume reckonings or molecular techniques might document other trends not achievable with quantification of bulk biomass. One of the ways in which the future research direction that can be taken might be determined is by exploring various measurement approaches.

Although the machine learning techniques are superior to the traditional models (linear regression, generalized linear models, etc.), they are correlational models and not mechanistic models. The relationships which were found to be statistically significant and ecologically interpretable, may not be cause-effect processes that are predictive of responses to novel environmental circumstances. Combining process-based knowledge and statistical knowledge forecasting is yet another problem in ecological forecast modeling that is currently under progress in the scientific community.

6. CONCLUSIONS AND FUTURE WORK

This study overcame the challenge of ineffective predictive models of phytoplankton to environmental change in an inverse estuarine environment. This study determines the limitations on phytoplankton predictability due to incomplete analytic tools based on a stringent motif of machine learning tactics to 16 years of all-inclusive environmental and phytoplankton dataset conveyed on the Coorong Wetlands region, to enhance ecological perception and management efficacy in this internationally significant system.

6.1. Research Problem and Solution

The concern that was raised was that in the inverse estuaries, conventional methods of statistics failed to model the behavior of non-linear interactions among the environmental factors and the dynamics of phytoplankton communities. This was a limitation of good management as well as conservation planning, in the face of shifting climatic conditions where powerful predictive capabilities are of utmost importance to adaptive solutions.

The ensemble model approach implemented in the machine learning model presented in the current study provides a holistic solution with a higher rate of predictive performance (Random Forest R^2 = 0.8725) in comparison with the traditional linear models (R^2 = 0.6123). Using the explainable AI approach implemented with SHAP analysis, it was shown that the issue of interpretability could be resolved without affecting predictive power, developing a methodological framework on achieving ecological knowledge in complex habitats in the future.

6.2. Principal Findings and Their Significance

The managerial direction and resource allocation guidance lies in the preference taken by the logical hierarchy of environmental drivers. Salinity was the primary control (a quarter of the explanatory power) and validated decades of anecdotal data and provides threshold values (40-50 PSU) on the effectiveness of intervention. The large contribution of phosphate (19% of the predictive value) indicates that the nutrient availability, too, is relevant and, therefore, multidimensional intervention based on management of all the stress factors could be more efficient than interventions with one dimension.

The habitat-specific examination showed spatial heterogeneity of environmental control and predictability patterns. The high hypersaline habitat H5 exhibited higher predictive accuracy (R^2 = 0.9031) than high brackish to marine habitat H3 (R^2 = 0.8102), meaning that management interventions will yield a more stable result in zones with high salinity. This geographical difference confirms the ideas of habitat zonation but offers quantitative advice on zone layer-specific actions.

The multifactorial examination of eleven phytoplankton clusters indicated taxonomically specific reactions that expand on the rapid community reasoning fostered by past research papers. Dinoflagellates proved to be the most salt tolerant (r = +0.40) while Cyanoprokaryota was the most salt sensitive (r = -0.55). This analysis offers quantitative baselines of density in biodiversity assessment and conservation planning. This finding that Picophytoplankton dynamics are dominated by biological controls (R^2 = 0.9427) diffuses many of the traditional paradigms of environmental determinism in stressed systems.

6.3. Management Applications

The priority of intervention is founded on quantitative ranks of drivers. Salinity pre-eminence approves the restoration of freshwater flow, and the phosphate significance promotes the other related nutrient management (Huang et al., 2024). The demonstration of habitat-specific patterns in this study facilitates the distribution of resources at spatially explicit scales.

The validated models meet the requirements to improve decision-support systems discovered by Kingsford et al. (2011). These models make possible three important management uses: (1) scenario planning to ecosystem responses at various climate or flow management scenarios, (2) intervention testing to assess the probable effectiveness of proposed management interventions prior to implementation, and (3) adaptive management to provide quantitative benchmarks, which may be monitored against observed outcomes. The performance of the models has been consistent over a 16-year interval of fluctuating environmental conditions (drought to flood) which gives reason to believe that they would be applicable under moderate climatic change conditions. Nevertheless, repeated monitoring is also necessary to countercheck predictions as the system is subjected to conditions of the environment that might be out of the historical range of conditions applied in model preparation.

6.4. Future Research Directions

This study has highlighted some promising possibilities in the development of ecological knowledge. Combining molecular methods would enable species-scale resolution which improves community assembly knowledge. Environmental DNA methods have the potential to measure cryptic diversity and indicator species (Sahu et al., 2022).

Logical extensions are dynamic models that have time dependencies. Time-series machine learning may learn seasonal cycles, and lag effects and autocorrelations that enhance performance and identify temporal processes. The hybrid models of mechanistic-statistical can enhance the novel condition prediction whilst retaining clarity.

Other failing aquatic systems would undergo this modelling framework and would dictate a more extensive applicability of the model and resultant methods can be transferred. Comparative studies with cross-ecosystems, cross-regions, and cross-stress gradients could prove the similarities of the core principles of community responses and focus on the system-specific patterns that would simply require certain management strategies.

REFERENCES

1. Acevedo-Trejos, E., Cadier, M., Chakraborty, S., Chen, B., Cheung, S. Y., Grigoratou, M., Guill, C., Hassenrück, C., Kerimoglu, O., Klauschies, T., Lindemann, C., Palacz, A., Ryabov, A., Scotti, M., Smith, S. L., Våge, S., & Prowe, F. (2022). Modelling approaches for capturing plankton diversity (MODIV), their societal applications and data needs. *Frontiers in Marine Science*, *9*. 10.3389/fmars.2022.975414

2. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), Article 53. 10.1186/s40537-021-00444-8

3. Araújo, S. O., Peres, R. S., Ramalho, J. C., Lidon, F., & Barata, J. (2023). Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives. *Agronomy*, *13*(12), Article 2976. 10.3390/agronomy13122976

4. Belluz, J. D. B., Jackson, J. M., Kellogg, C. T. E., Peña, M. A., Giesbrecht, I. J. W., & Hobson, L. A. (2024). Phytoplankton community composition links to environmental drivers across a fjord to shelf gradient on the central coast of British Columbia. *Frontiers in Marine Science*, *11*. 10.3389/fmars.2024.1458677

5. Berthold, M., Nieters, P., & Vortmeyer-Kley, R. (2025). Machine learning to identify environmental drivers of phytoplankton blooms in the Southern Baltic Sea. *Scientific Reports*, *15*(1). 10.1038/s41598-025-85605-y

6. Brookes, J., Lamontagne, S., Aldridge, K., Benger, S., Bissett, A., Bucater, L., Cheshire, A., Cook, P., Deegan, B., Dittmann, S., Fairweather, P., Fernandes, M., Ford, P., Geddes, M., Gillanders, B., Grigg, N., Haese, R., Krull, E., Langley, R., & Ye, Q. (2009). *An Ecosystem Assessment Framework to Guide Management of the Coorong. Final Report of the CLLAMMecology Research Cluster*. The University of Adelaide. http://hdl.handle.net/2440/75829

7. Brookes, J. D., Busch, B., Cassey, P., Chilton, D., Dittmann, S., Dornan, T., Giatas, G., Gillanders, B. M., Hipsey, M., Huang, P., Keneally, C., Jackson, M. V., Mosley, L., Mott, R., Paton, D., Prowse, T., Waycott, M., Ye, Q., Zhai, S., & Gibbs, M. (2023). How well is the Basin Plan meeting its objectives? From the perspective of the Coorong, a sentinel of change in the

Murray-Darling Basin. *Australian Journal of Water Resources*, *27*(2), 223–240. https://doi.org/10.1080/13241583.2023.2241161

8. Choudhary, R., Kumar, A., Priyadharsini, C., Naik, M. M., Choudhury, M., & Khan, N. A. (2025). Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence. *Scientific Reports*, *15*(1). 10.1038/s41598-025-09463-4

9. Conley, D., Schelske, C., & Stoermer, E. (1993). Modification of the biogeochemical cycle of silica with eutrophication. *Marine Ecology Progress Series*, *101*, 179–192. https://doi.org/10.3354/meps101179

10. Dang, H. V., Stephanie, K., Huang, P., Carey, C. C., & Hipsey, M. R. (2025). Phytoplankton group classification by integrating trait information and observed environmental thresholds. *Ecological Informatics*, *90*, 103212. 10.1016/j.ecoinf.2025.103212

11. Gao, W., Xiong, F., Lu, Y., Xin, W., Wang, H., Feng, G., Kong, C., Fang, L., Gao, X., & Chen, Y. (2024). Water quality and habitat drive phytoplankton taxonomic and functional group patterns in the Yangtze River. *Ecological Processes*, *13*(1). 10.1186/s13717-024-00489-6

12. Goldsworthy, S., Baring, R., Giatas, G., Nitschke, J., Bucater, L., & Ye, Q. (2022n.d.). Ecosystem models to inform the development of strategies to restore a functioning South Lagoon food web in the Coorong. https://goyderinstitute.org/wp-content/uploads/2023/06/goyder_trs_22-11_ecosystems_models_food_web_coorong.pdf

13. Haynes, D., Skinner, R., Tibby, J., Cann, J., & Fluin, J. (2011). Diatom and foraminifera relationships to water quality in The Coorong, South Australia, and the development of a diatom-based salinity transfer function. *Journal of Paleolimnology*, *46*(4), 543–560. 10.1007/s10933-011-9508-y

14. Hemraj, D. A., Allais, L., & Leterme, S. C. (2017). A combination of salinity and pH affects the recruitment of Gladioferens pectinatus (Brady) (Copepoda; Calanoida). *Limnology and Oceanography*, *62*(5), 1799–1809. 10.1002/lno.10534

15. Hemraj, D. A., Hossain, A., Ye, Q., Qin, J. G., & Leterme, S. C. (2017). Anthropogenic shift of planktonic food web structure in a coastal lagoon by freshwater flow regulation. *Scientific Reports*, *7*(1), 44441–44441. 10.1038/srep44441

16. Hemraj, D. A., Hossain, M. A., Ye, Q., Qin, J. G., & Leterme, S. C. (2017). Plankton bioindicators of environmental conditions in coastal lagoons. *Estuarine, Coastal and Shelf Science*, *184*, 102–114. 10.1016/j.ecss.2016.10.045

17. Huang, P., Mosley, L., Brookes, J. D., Sims, C., Waycott, M., Paraska, D., Zhai, S. Y., & Hipsey, M. R. (2024). Hydrologic Versus Biogeochemical Control of Nutrient Dynamics in a

Shallow Hypersaline Coastal Lagoon: Insight From a Coupled Hydrodynamic-Water Quality Model. *Journal of Geophysical Research Biogeosciences*, *129*(7). https://doi.org/10.1029/2023jg007497

18. Huisman, J., van Oostveen, P., & Weissing, F. J. (1999). Critical depth and critical turbulence: Two different mechanisms for the development of phytoplankton blooms. *Limnology and Oceanography*, *44*(7), 1781–1787. https://doi.org/10.4319/lo.1999.44.7.1781

19. Hussein, E. E., Zerouali, B., Bailek, N., Derdour, A., Ghoneim, S. S. M., Santos, C. A. G., & Hashim, M. A. (2024). Harnessing Explainable AI for Sustainable Agriculture: SHAP-Based Feature Selection in Multi-Model Evaluation of Irrigation Water Quality Indices. *Water*, *17*(1), 59–59. 10.3390/w17010059

20. Jendyk, J., Hemraj, D. A., Brown, M. H., Ellis, A. V., & Leterme, S. C. (2014). Environmental variability and phytoplankton dynamics in a South Australian inverse estuary. *Continental Shelf Research*, *91*, 134–144. 10.1016/j.csr.2014.08.009

21. Kingsford, R. T., Walker, K. F., Lester, R. E., Young, W. J., Fairweather, P. G., Sammut, J., & Geddes, M. C. (2011). Ramsar wetland in crisis - the Coorong, Lower Lakes and Murray Mouth, Australia. *Marine and Freshwater Research*, *62*(3), 255–265. 10.1071/MF09315

22. Kundu, S., Datta, P., Pal, P., Ghosh, K., Das, A., & Das, B. K. (2025). Unveiling the Hidden Connections: Using Explainable Artificial Intelligence to Assess Water Quality Criteria in Nine Giant Rivers. *Journal of Cleaner Production*, *492*, 144861–144861. 10.1016/j.jclepro.2025.144861

23. Leterme, S. C., Allais, L., Jendyk, J., Hemraj, D. A., Newton, K., Mitchell, J., & Shanafield, M. (2015). Drought conditions and recovery in the Coorong wetland, south Australia in 1997–2013. *Estuarine, Coastal and Shelf Science*, *163*, 175–184. 10.1016/j.ecss.2015.06.009

24. Levy, D. N. L. (1985). Some Studies in Machine Learning Using the Game of Checkers. I. In *Computer Games I*. Springer New York.

25. Liu, M., Huang, Y., Hu, J., He, J., & Xiao, X. (2023). Algal community structure prediction by machine learning. *Environmental Science & Ecotechnology*, *14*, 100233–100233. 10.1016/j.ese.2022.100233

26. Liu, Q., & Wu, Y. (2012). Supervised Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning*. Springer. 10.1007/978-1-4419-1428-6_451

27. López Abbate, M. C., Molinero, J. C., Guinder, V. A., Perillo, G. M. E., Freije, R. H., Sommer, U., Spetter, C. V., & Marcovecchio, J. E. (2017). Time-varying environmental control

of phytoplankton in a changing estuarine system. *Science of the Total Environment*, *609*, 1390–1400. 10.1016/j.scitotenv.2017.08.002

28. Lynam, C. P., Llope, M., Möllmann, C., Helaouët, P., Bayliss-Brown, G. A., & Stenseth, N. C. (2017). Interaction between top-down and bottom-up control in marine food webs. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), 1952–1957. https://doi.org/10.1073/pnas.1621037114

29. Mosley, L., Priestley, S., Brookes, J., Dittmann, S., Farkaš, J., Farrell, M., Ferguson, A., Gibbs, M., Hipsey, M., Huang, J., Lam-Gordillo, O., Leterme, S., Mount, R., Ning, N., Nowak, B., Pfennig, B., Reeves, J., Revill, A., Shiel, R., ... Tyler, J. (2018). Extreme low flows in the Murray-Darling Basin—understanding the ecological implications. GOYDER Institute for Water Research.

30. Mosley, L. M., Priestley, S., Brookes, J., Dittmann, S., Farkaš, J., Farrell, M., Ferguson, A. J., Gibbs, M., Hipsey, M., Huang, J., Lam-Gordillo, O., Simpson, S. L., Tyler, J. J., Waycott, M., & Welsh, D. T. (2023). Extreme eutrophication and salinisation in the Coorong estuarine-lagoon ecosystem of Australia's largest river basin (Murray-Darling). *Marine Pollution Bulletin*, *188*, 114648. https://doi.org/10.1016/j.marpolbul.2023.114648

31. National Oceanic and Atmospheric Administration. (2019, February 1). *Aquatic Food Webs*. Www.noaa.gov. https://www.noaa.gov/education/resource-collections/marine-life/aquatic-food-webs

32. Olson, C. R., Gschwentner, D., Drost, A. M., Mohan, J., & Klip, H. C. L. (2024). What's the matter in phytoplankton? Highlighting the importance of stoichiometric traits in lake ecosystem models. *Frontiers in Ecology and Evolution*, *12*. 10.3389/fevo.2024.1505018

33. Pandya, A., Kapoor, V., & Joshi, A. (2024). Stock market prediction using ARIMA-LSTM hybrid. *Journal of Information & Optimization Sciences*, *45*(4), 1129–1139. 10.47974/JIOS-1697

34. Priestley, S., Mosley, L., Farkas, J., Tyler, J., Shao, Y., Shanafield, M., Banks, E., Wong, W., & Leyden, E. (2022). Sources and transport of nutrients in the Coorong. Goyder Institute for Water Research Technical Report Series No. 22/01. URL: https://goyderinstitute.org/wp-content/uploads/2023/03/goyder_trs-20-01_sources_and_transport_of_nutrients_in_the_coorong.pdf

35. Ritchie, R. J. (2008). Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and total chlorophylls in natural assemblages of photosynthetic organisms using

acetone, methanol, or ethanol solvents. *Photosynthetica*, *46*(1), 115–126. https://doi.org/10.1007/s11099-008-0019-7

36. Sahu, A., Kumar, N., Pal Singh, C., & Singh, M. (2022). Environmental DNA (eDNA): Powerful Technique for Biodiversity Conservation. *Journal for Nature Conservation*, *71*, 126325. https://doi.org/10.1016/j.jnc.2022.126325

37. Saleti, S., Panchumarthi, L. Y., Kallam, Y. R., Parchuri, L., & Jitte, S. (2024). Enhancing Forecasting Accuracy with a Moving Average-Integrated Hybrid ARIMA-LSTM Model. *SN Computer Science*, *5*(6), 704-. 10.1007/s42979-024-03060-4

38. Soudant, D., Beliaeff, B., & Thomas, G. (1997). Dynamic linear Bayesian models in phytoplankton ecology. *Ecological Modelling*, *99*(2-3), 161–169. 10.1016/s0304-3800(97)01949-2

39. Testa, J. M., Murphy, R. R., Brady, D. C., & Kemp, W. M. (2018). Nutrient- and Climate-Induced Shifts in the Phenology of Linked Biogeochemical Cycles in a Temperate Estuary. *Frontiers in Marine Science*, *5*. 10.3389/fmars.2018.00114

40. van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., & Viney, N. R. (2013). The Millennium Drought in southeast Australia (2001-2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, *49*(2), 1040–1057. 10.1002/wrcr.20123

41. Wolanski, E. (2014). *Estuaries of Australia in 2050 and beyond*. Dordrecht Springer Netherlands.

42. Xu, Y., Zhang, D., Lin, J., Peng, Q., Lei, X., Jin, T., Wang, J., & Yuan, R. (2023). Prediction of phytoplankton biomass and identification of key influencing factors using interpretable machine learning models. *Ecological Indicators*, *158*, 111320–111320. 10.1016/j.ecolind.2023.111320

43. Yang, R., Fan, X., Zhao, L., & Yang, K. (2023). Identification of major environmental factors driving phytoplankton community succession before and after the regime shift of Erhai Lake, China. *Ecological Indicators*, *146*, 109875. https://doi.org/10.1016/j.ecolind.2023.109875

44. Ye, Q., Giatas, G., Dittmann, S., Baring, R., Bucater, L., Deane, D., Furst, D., Brookes, J., Rogers, D., & Goldsworthy, S. (2020). A synthesis of current knowledge of the food web and food resources for waterbird and fish populations in the Coorong. Research @ Flinders; Goyder Institute for Water Research. https://researchnow.flinders.edu.au/en/publications/a-synthesis-of-current-knowledge-of-the-food-web-and-food-resource