

# Wavelet-Integrated Machine Learning Models for Predicting Marine Chlorophyll-a Concentration along the California Coast

Blake E. Matz

*La Jolla High School, 750 Nautilus Street, La Jolla, CA 92037, United States*

## ABSTRACT

In recent years, algal blooms have occurred with greater frequency and intensity along the southern and central California coast. Accurate forecasting of blooms is challenging due to the numerous environmental factors that can influence algal growth. Marine chlorophyll-a concentration is one of the key indicators that can be used in monitoring and predicting algal blooms. Previous research efforts that used machine learning models to predict chlorophyll-a concentration in the southern to central California coastal region were mostly targeted at individual locations and used datasets covering fewer than eight years before 2019. In this study, wavelet analysis (WA) was used to pre-process chlorophyll-a marine long-duration time-series data to increase its suitability for machine learning by removing noise while retaining short-term spikes. SVR, Random Forest, XGBoost, ANN and LSTM machine learning models were then applied to the WA-integrated data pipeline along with water quality and meteorological inputs to predict chlorophyll-a concentration at three locations along the southern to central California coast. Additionally, datasets spanning from 2008 to 2025 were employed to address the shorter durations of the previous studies. The WA-ANN model achieved the overall best performance (Scripps Pier  $R^2 = 0.88$ , Cal Poly Pier  $R^2 = 0.79$ , Stearns Wharf  $R^2 = 0.75$ ) for the three locations, accurately capturing the spikes indicative of algal blooms.

**Keywords:** Chlorophyll-a; Harmful Algal Blooms; Machine Learning; Wavelet Analysis; Artificial Neural Networks; California Coast; Time-series Forecasting

## INTRODUCTION

Microorganisms capable of photosynthesis are important primary producers of the energy-containing compounds that are the basis of aquatic food chains (1). Periodic algal blooms are a natural feature of ocean ecosystems, and the increase in phytoplankton during

blooms contributes to the nourishment of higher marine life (2). However, some algal blooms are caused by toxin-producing phytoplankton, and these harmful algal blooms (HABs) are of great concern because of their potential to harm marine mammals and human populations (3, 4). HABs have been increasing in frequency in recent decades, possibly due to human-induced climate change (5). It has therefore become critically important to develop methods for predicting HABs. As little as a few days to one week of advance warning can greatly reduce harm to the human population from toxin exposure through recreational activities or ingestion of contaminated seafood (6, 7). The effectiveness of strategies for mitigating or controlling HABs, such as

---

**Corresponding author:** Blake E. Matz, E-mail: [blakesports29@gmail.com](mailto:blakesports29@gmail.com).

**Copyright:** © 2026 Blake E. Matz. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Accepted** May 14, 2026

<https://doi.org/10.70251/HYJR2348.43124132>

physical scavenging or the introduction of chemical or biological agents targeting the relevant phytoplankton species, also strongly depends on the ability to predict blooms in advance (8).

Chlorophyll-a concentration is commonly used as a proxy for phytoplankton in the assessment of HABs because it is relatively easy to measure (5, 9). Several previous research efforts have used machine learning techniques in an attempt to predict temporal variation in chlorophyll-a concentration from data obtained by direct point measurements of chlorophyll-a, ocean nutrient content, and environmental variables in California coastal regions. Yu *et al.* (10) focused on the prediction of phytoplankton at the Scripps Pier location in La Jolla, CA, using an ocean water quality dataset over a relatively short timeframe (2008–2015), with the gradient boosted decision tree (GBDT) model achieving the highest performance ( $R^2 = 0.997$ ). Cheng *et al.* (11) used a Random Forest approach to assess feature importance and interaction scores using ocean water quality datasets (2011–2018) from the Santa Cruz Wharf in Monterey Bay. Lin *et al.* (12) used both ocean water quality and wind direction and speed datasets (2008–2015) from the Scripps Pier location and achieved the best performance with a wavelet analysis-back propagation neural network (WA-BPNN) model ( $R^2 = 0.73$ ).

In the ocean environment, marine data such as chlorophyll-a concentration is notoriously non-stationary, which means that its mean and variance change over time due to tides, seasons, weather, and climate change (13, 14). Wavelet analysis has been used to address these challenges by decomposing the signal to capture short-lived transients (e.g., rapid increases in chlorophyll-a due to sudden nutrient surges) (15–17). With this pre-processing technique, a single input signal (chlorophyll-a concentration in this study) is shifted and scaled and the data is smoothed by removing noise (18). Despite the smoothing, the sharpness of spikes is maintained, which is needed in the current application to capture the abrupt increases in chlorophyll-a during algal blooms.

There is scant existing research combining machine learning models with wavelet analysis to predict chlorophyll-a concentration in coastal regions, especially for California. Prior investigations focused on relatively localized California coastal regions, and the datasets did not extend beyond 2018 (10–12). Lin *et al.* (12) laid the foundation for future research, but this work was limited in scope, with a dataset spanning only seven years at a single California coastal location. The present study aims to compare the predictive capability of various

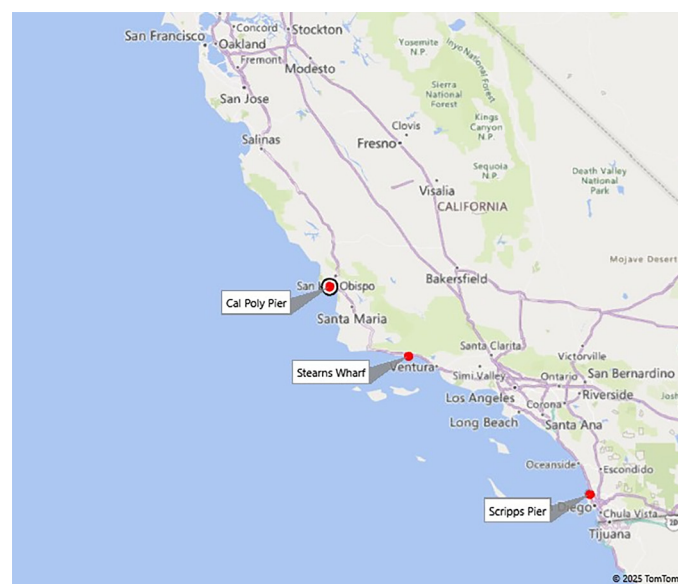
hybrid machine learning models augmented with wavelet analysis for chlorophyll-a concentration at multiple sites along the California coast. Additionally, the datasets used in this research have increased relevance for near-future prediction efforts as they span approximately 17 years and extend nearly to the present.

## METHODS AND MATERIALS

### Dataset and Preprocessing

The California Harmful Algal Bloom Monitoring and Alert Program (CalHABMAP) provided the biological and ocean water quality data for this investigation (19). Data were collected weekly from June 2008 to March 2025 along the central and southern California coast. A number of chemical compounds and the water temperature were measured in each weekly sample. Additionally, cell counts of the *Akashiwo sanguinea*, *Lingulodinium polyedra*, *Pseudo-nitzschia delicatissima*, and *Pseudo-nitzschia seriata* algae species were also included in the datasets.

Three locations from the CalHABMAP datasets (19) were investigated: Scripps Pier (SLP, latitude degrees north 32.867, longitude degrees west 117.257) in La Jolla, Stearns Wharf (SW, latitude degrees north 34.408, longitude degrees west 119.685) in Santa Barbara, and Cal Poly Pier (CPP, latitude degrees north 35.17, longitude degrees west 120.741) in San Luis Obispo (Figure 1).



**Figure 1.** Map of the southern to central California coastline, showing the three locations investigated.

Wind speed and direction datasets for the three locations were obtained from the National Data Buoy Center (NDBC) over the same June 2008–March 2025 time period as the ocean water quality data (20–22). Measurements were recorded hourly as numeric data.

For the CalHABMAP dataset, a previous study (12) examined the Scripps Pier location from 2008 to 2015 at weekly intervals. The features in the dataset were evaluated for relevance using a chi-squared test, and the following ocean water quality features were used in their machine learning models: water temperature (°C), chlorophyll-a ( $\text{mg}/\text{m}^3$ ), phaeophytin ( $\text{mg}/\text{m}^3$ ), phosphate ( $\mu\text{M}$ ), silicate ( $\mu\text{M}$ ), nitrite ( $\mu\text{M}$ ), ammonia ( $\mu\text{M}$ ), and nitrate ( $\mu\text{M}$ ). Both wind speed and direction were included as features in the previously investigated machine learning models. The current study used the same feature list. In addition, this study expanded the scope to three locations along the California coast and extended the time period to 2008–2025.

Each dataset for the various locations along the California coast contained missing weekly measurements that created time interval gaps that would adversely affect the operation of the machine learning models. Large spans (greater than one month) of missing data were removed entirely. Smaller missing-data gaps were imputed individually for each water-quality feature using a quantile-mapping technique with a monthly percentile grid.

For the wind speed and direction datasets, only hourly measurements were available. To match the specific days in the CalHABMAP dataset, which had weekly sampling intervals, a meteorological convention that takes the vector average was used to calculate daily values. The dates in the CalHABMAP dataset were then matched with dates in the calculated daily wind dataset.

After data preprocessing and imputation, Scripps Pier had 796, Cal Poly Pier had 799, and Stearns Wharf had 831 weekly measurements of selected features used in the current study from 2008 to 2025.

### **Feature Engineering**

Time-lagged feature engineering has been widely adopted in chlorophyll-a forecasting studies, where incorporating historical observations significantly improves predictive performance compared to static feature representations (23–25). Therefore, autoregressive lagged features were constructed by incorporating prior observations of chlorophyll-a and associated environmental variables (e.g., water temperature, nutrients, and wind) across multiple previous time steps. This approach captures temporal dependence,

seasonality, and short-term variability inherent in coastal ecosystems. Ultimately, the training data for each week used current-week  $t$  features and previous-week  $t-1$  lag features to predict next-week  $t+1$  chlorophyll-a values, giving a two-week look-back window for a one-week-ahead prediction. To avoid temporal leakage, all engineered features were constructed only using information available at or before the forecast issue time.

### **Baseline Models**

Seven models were implemented and evaluated, comprising two statistical baselines and five machine learning models. Linear Regression served as the primary benchmark, implemented using an ordinary least squares (OLS) approach in scikit-learn to establish a lower-bound reference for non-linear model comparisons. The second statistical baseline, AutoRegressive Integrated Moving Average (ARIMA), is a classical time-series forecasting model that captures temporal autocorrelation and trend; an ARIMA(1,0,0) specification with exogenous regressors (ARIMAX) was implemented using the Python statsmodels package.

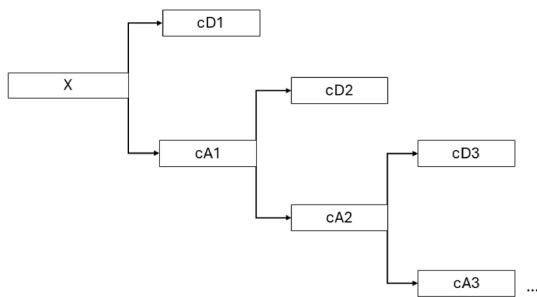
Among the machine learning models, Support Vector Regression (SVR) uses a kernel function to map inputs into a higher-dimensional feature space and fits a regression hyperplane with a margin of tolerance. An RBF kernel was used, with the regularization parameter  $C$ , the margin width  $\epsilon$ , and the kernel coefficient  $\gamma$  tuned during training. Random Forest (RF) is an ensemble method that aggregates predictions from multiple decision trees trained on bootstrapped subsets of the data; the number of estimators, maximum tree depth, minimum samples per leaf, and number of features considered at each split were treated as tunable hyperparameters. XGBoost is a gradient-boosted ensemble method that builds trees sequentially to minimise a regularised loss function, with tuned hyperparameters including the number of estimators, learning rate, maximum tree depth, minimum child weight, subsampling ratio, column subsampling ratio, and L1 and L2 regularization terms.

The two deep learning models differed in architecture and temporal handling. The Artificial Neural Network (ANN), following the architecture of Lin et al. (12), was implemented as a shallow feed-forward regression network in PyTorch with one hidden layer, where the number of input features and hidden layer neurons were tuned hyperparameters and early stopping with a patience of seven epochs was applied to prevent overfitting. The Long Short-Term Memory (LSTM) network is

a recurrent architecture designed for sequential data, capable of learning long-range temporal dependencies through gated memory cells; a two-layer stacked LSTM was implemented in Keras with a four-step rolling input window, with the number of first-layer neurons tuned and subsequent layer sizes derived proportionally, and early stopping applied with a patience of twelve epochs.

### Hybrid Wavelet-Integrated Machine Learning Models

Chlorophyll-a concentration was decomposed using a Daubechies-4 wavelet transform with three decomposition levels. The resulting detail coefficients and final approximation coefficient were used as chlorophyll-a-derived predictors and combined with scaled water quality and wind variables for model training (Figure 2).



**Figure 2.** Wavelet Decomposition Diagram. *X* represents the input signal. The high-pass filter generates high-frequency *D*-level signals. The *cD<sub>n</sub>* Detail coefficients for the *D*-level signals contain the high-frequency details of the signal at each level of decomposition. *cA<sub>n</sub>* represents the Approximation coefficients of the low-frequency *A*-level signals, which are the trend of the signal at each level.

Chronological splitting without shuffling was applied to the dataset in order to preserve temporal structure. Sixty percent of the dataset was used to train the models, twenty percent for validation, and the remaining twenty percent for testing. The test split was reserved throughout model development and used only once for evaluating predictive effectiveness. Feature selection was conducted on the training split using 5-fold Time-Series Split cross-validation. In each fold, a Random Forest Regressor was applied exclusively to the fold-specific training window to calculate feature importance in order to rank them. The input feature subset was selected based on the rankings, and the optimal number of features (12-24) was determined based on mean cross-validated performance across folds. Depending on the model, either a grid or random search was used to tune the hyperparameters. The models'

predictive capability for chlorophyll-a concentration was assessed using coefficient of determination ( $R^2$ ) and Root Mean Squared Error (RMSE) with mean, standard deviation, and 95% confidence intervals (CIs) reported across the ten randomly seeded runs. Furthermore, the models' complexity was intentionally kept to a minimum to avoid overfitting, a common issue in machine learning applications to moderate-size coastal time-series datasets (26, 27).

## RESULTS

The performance of the seven baseline models varied substantially across locations. At SLP, the LSTM baseline achieved the best performance by a large margin ( $R^2 = 0.43$ , 95% CI: 0.37–0.48; RMSE = 8.42, 95% CI: 8.02–8.82), Random Forest and SVR were the only other models with clearly positive  $R^2$ . At CPP, Random Forest was the strongest model ( $R^2 = 0.24$ , 95% CI: 0.24–0.25; RMSE = 12.82, 95% CI: 12.79–12.84), followed by SVR and XGBoost, whereas Linear Regression and ARIMA performed poorly and exhibited wide uncertainty intervals. In contrast, all baseline models performed poorly at SW, where every model produced negative  $R^2$  values. These findings indicate that the baseline models' predictability is site-specific, so that even the LSTM model with the best overall performance at SLP cannot be reliably applied to other coastal locations.

When augmented with wavelet analysis, the machine learning models (WA-SVR, WA-RF, WA-XGBoost, WA-ANN and WA-LSTM) all demonstrated significant performance improvements compared with their baseline models, as shown in Table 1.

WA-ANN was the best-performing model at all three locations, achieving the highest  $R^2$  and lowest RMSE. Figure 3 compares the chlorophyll-a concentration predictions from the WA-ANN machine learning model with measured values. Figure 3 shows the comparisons for Scripps Pier, Cal Poly Pier, and Stearns Wharf, respectively.

WA-SVR was the next strongest model at SLP and CPP, whereas WA-LSTM and WA-XGBoost showed intermediate, location-dependent performance. WA-RF was generally the weakest WA model, particularly at SLP and SW. Overall, the results indicate that the WA-ANN architecture provided the most robust and accurate predictive skill among the WA variants tested. Additionally, WA models demonstrated stability of validation and test metrics across ten random seeds, suggesting that validation-set overfitting was not a

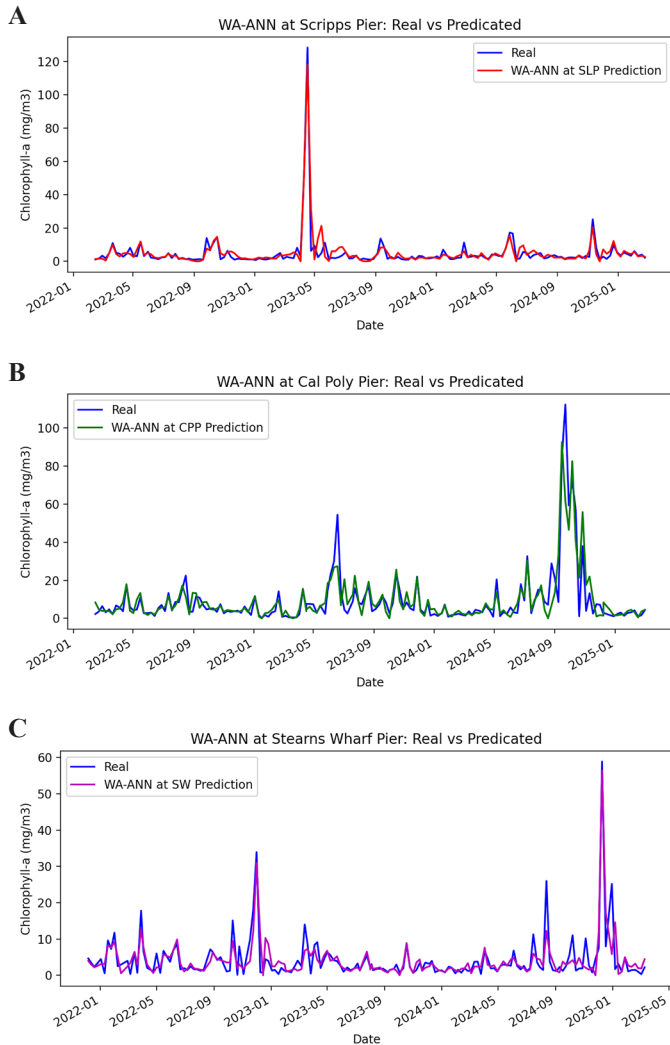
concern.

A paired comparison was carried out on the top-performing WA-integrated hybrid machine learning models; the tests were performed on the aligned held-out test set using seed-matched loss differentials across 10

evaluations. Diebold–Mariano (DM) tests on the absolute and squared errors showed no statistically significant differences across seeds for WA-ANN versus WA-SVR or for WA-ANN versus WA-LSTM. The DM test results are shown in Table 2.

**Table 1.** Performance of five wavelet-integrated hybrid machine learning models at three California coastal locations (SLP, CPP, SW), reported as mean  $R^2$  and RMSE with standard deviation and 95% confidence intervals across ten randomly seeded runs.

Hybrid Model	Location	Metric	Mean	Std	CI 95% Low	CI 95% High
WA-SVR	SLP	$R^2$	0.77	0.08	0.71	0.82
		RMSE	5.32	0.81	4.74	5.90
	CPP	$R^2$	0.67	0.06	0.63	0.71
		RMSE	8.45	0.71	7.94	8.96
	SW	$R^2$	0.64	0.09	0.57	0.71
		RMSE	3.74	0.47	3.40	4.08
WA-RF	SLP	$R^2$	0.34	0.01	0.32	0.35
		RMSE	9.05	0.12	8.96	9.14
	CPP	$R^2$	0.60	0.02	0.58	0.62
		RMSE	9.26	0.25	9.08	9.45
	SW	$R^2$	0.25	0.01	0.24	0.26
		RMSE	5.45	0.06	5.40	5.49
WA-XGBoost	SLP	$R^2$	0.48	0.02	0.47	0.50
		RMSE	7.95	0.18	7.81	8.08
	CPP	$R^2$	0.26	0.20	0.11	0.40
		RMSE	12.58	1.80	11.29	13.87
	SW	$R^2$	0.44	0.02	0.41	0.45
		RMSE	4.73	0.10	4.65	4.80
WA-ANN	SLP	$R^2$	0.88	0.02	0.86	0.89
		RMSE	3.81	0.37	3.54	4.08
	CPP	$R^2$	0.79	0.01	0.77	0.79
		RMSE	6.79	0.20	6.65	6.94
	SW	$R^2$	0.75	0.01	0.74	0.75
		RMSE	3.15	0.07	3.09	3.20
WA-LSTM	SLP	$R^2$	0.46	0.09	0.39	0.52
		RMSE	8.17	0.69	7.67	8.67
	CPP	$R^2$	0.56	0.07	0.50	0.61
		RMSE	9.74	0.90	9.09	10.38
	SW	$R^2$	0.51	0.04	0.47	0.54
		RMSE	4.40	0.20	4.25	4.54



**Figure 3.** WA-ANN model chlorophyll-a concentration predictions compared to real measurements at the Scripps Pier (A), CalPoly Pier (B), and Stearns Wharf (C) locations.

**Table 2.** Diebold–Mariano paired comparison test results for the three best-performing wavelet-integrated models, reporting mean test statistic and mean p-value across ten seed-matched evaluations using absolute and squared error loss functions.

Comparison	Loss	Mean Statistic	Mean p-value
WA-ANN vs WA-SVR	Absolute Error	-0.06	0.76
	Squared Error	-0.89	0.37
WA-ANN vs WA-LSTM	Absolute Error	-0.32	0.75
	Squared Error	-1.01	0.31

## DISCUSSION

From 2021 to 2025, the Southern California coast has experienced four consecutive years of toxic algal blooms (28). The most recent, in January 2025, was the largest in recorded history, spanning from Baja California, Mexico, to central California (3). It was also among the longest-lasting and most lethal algal blooms in recent memory (29). This bloom had a dramatic impact on marine ecosystems, primarily from domoic acid, a neurotoxin produced during these algal blooms that can poison marine life (30, 31). Accurate prediction of algal blooms from chlorophyll-a concentration can provide the timely warnings needed for preemptive action to protect human and animal health, reduce economic losses, and preserve marine ecosystems (32).

For the 17 years of data, even though LSTM is theoretically better at learning time-series sequences than SVR or ANN, the poor quality of the input data (gaps in crucial exogenous variables like nutrient concentration) limits its superiority and also makes the memory component of LSTM less advantageous (33). Since the more complex WA-LSTM model did not significantly outperform simpler WA-ANN and WA-SVR hybrid models, there is no need to give up predictive performance to reduce model complexity and computational time (LSTM is generally slower). If there is no statistically significant difference in the models' performance, then the choice of model can be made based on ease of deployment, interpretation, and training (34). The similar performance for all three models suggests they may have reached a plateau in accuracy given deficiencies in the quality of the current data, indicating a need for higher-resolution or more diverse explanatory variables (e.g., satellite ocean color data and turbidity), or extension beyond 17 years of data to achieve a statistically significant difference in performance (35, 36).

The wavelet-integrated data pipeline applied to California coastal data in the present investigation is highly transferable to other coastal regions, such as the Atlantic or Gulf coasts. Coastal environmental data everywhere is naturally non-stationary, which means that its statistical properties change over time. The wavelet transform effectively handles this non-stationarity better than traditional Fourier methods (37, 38). Additionally, the wavelet decomposition is extremely useful for regions with high environmental variability or frequent extreme events (e.g., hurricanes in the Atlantic or monsoons in Asia), where separating short-term disturbances from

long-term trends is crucial for accurate analysis (39). Once a clean and uncluttered signal is obtained, the choice of AI model becomes secondary (40).

## CONCLUSION

In this study, several machine learning models with and without wavelet decomposition preprocessing were evaluated for their ability to predict ocean water chlorophyll-a concentration, a key indicator of harmful algal bloom activity, along the southern to central California coast. Previous studies in this region were limited by short time spans and single-location datasets and did not investigate the effect of wavelet analysis preprocessing on machine learning model predictive performance. To address these gaps, the present study used 17 years of time-series data from three locations in the southern to central California coastal region and applied wavelet analysis preprocessing. The predictive ability of WA-SVR, WA-RF, WA-XGBoost, WA-ANN, and WA-LSTM machine learning models for marine chlorophyll-a concentration was then evaluated against the corresponding baseline models without wavelet decomposition. The models were trained and tuned by splitting the data into training, validation, and test datasets.

The WA-ANN machine learning model achieved the best performance at all examined locations, although DM tests did not show a statistically significant difference between WA-ANN and WA-SVR, and between WA-ANN and WA-LSTM. In the future, the same methodology can be applied to other coastal regions that have sufficient data with minor modifications, e.g., incorporating site-specific environment drivers operating in the new regions and re-calculating time lag effects to fit the faster or slower biological responses to the input variables in the new locations. In summary, future research at different coastal regions should focus on acquiring more physical and satellite data from more sites while maintaining a similar model structure.

## ACKNOWLEDGEMENTS

I would like to thank my mentor Dr. Joe Xiao from Veritas AI for his guidance on this research.

## CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

## REFERENCES

- Berdalet E, Fleming LE, Gowen R, Davidson K, Hess P, Backer LC, *et al.* Marine harmful algal blooms, human health and wellbeing: Challenges and opportunities in the 21st century. *J Mar Biol Assoc U K.* 2016; 96 (1): 61–91. doi:10.1017/S0025315415001733
- Christensen VG, Crawford CJ, Dusek RJ, Focazio MJ, Fogarty LR, Graham JL, *et al.* Interdisciplinary science approach for harmful algal blooms (HABs) and algal toxins—A strategic science vision for the U.S. Geological Survey. U.S. Geological Survey Circular No. 1520. 2024; Available from: <https://pubs.usgs.gov/publication/cir1520/full>, doi:10.3133/cir1520 (accessed on 2025-12-31)
- Purtill C. Lethal algae bloom is over, but sickened marine mammals aren't safe yet. Los Angeles Times [Internet]. 2025-06-05. Available from: <https://www.latimes.com/science/story/2025-06-05/one-of-worst-harmful-algal-blooms-in-southern-californias-history-is-officially-over> (accessed on 2025-12-31)
- Lewitus AJ, Horner RA, Caron DA, Garcia-Mendoza E, Hickey BM, Hunter M, *et al.* Harmful algal blooms along the North American west coast region: History, trends, causes, and impacts. *Harmful Algae.* 2012; 19: 133–159. doi:10.1016/j.hal.2012.06.009
- Nezlin NP, McLaughlin K, Booth JAT, Cash CL, Diehl DW, Davis KA, *et al.* Spatial and temporal patterns of chlorophyll concentration in the Southern California Bight. *J Geophys Res Oceans.* 2018; 123 (4): 231–245. doi:10.1002/2017JC013324
- Newbold SC, Lindley S, Albeke S, Viers J, Johnston RJ. Valuing satellite data for harmful algal bloom early warning systems [Internet]. Washington (DC): Resources for the Future Working Paper WP 22-23. 2022; Available from: <https://www.rff.org/publications/working-papers/valuing-satellite-data-for-harmful-algal-bloom-early-warning-systems/> (accessed on 2026-04-26)
- Jin D and Hoagland P. The value of harmful algal bloom predictions to the nearshore commercial shellfish fishery in the Gulf of Maine. *Harmful Algae.* 2008; 7 (6): 772–781. doi:10.1016/j.hal.2008.03.002
- Anderson DM, Wells ML, Trainer VL, Suddleson M, Claridge K, Coyne KJ, *et al.* Controlling harmful algal blooms (HABs) in marine waters: Review of current status and future prospects. *Harmful Algae.* 2025; 150: 102989. doi:10.1016/j.hal.2025.102989
- Wang R, Li X, Song J, Wang Z, Zhong G, Yuan H, *et al.* Surface seawater chlorophyll-a variability in the South China Sea: Influence of pCO<sub>2</sub> and co-varying environmental factors. *Environ Res.* 2025; 279 (Pt 1): 121808. doi:10.1016/j.envres.2025.121808
- Yu P, Gao R, Zhang D, and Liu ZP. Predicting coastal

- algal blooms with environmental factors by machine learning methods. *Ecol Indic.* 2021; 123: 107334. doi:10.1016/j.ecolind.2020.107334
11. Cheng Y, Bhoot VN, Kumbier K, Sison-Mangus MP, Brown JB, Kudela R, *et al.* A novel random forest approach to revealing interactions and controls on chlorophyll concentration and bacterial communities during coastal phytoplankton blooms. *Sci Rep.* 2021; 11 (1): 19944. doi:10.1038/s41598-021-98110-9
  12. Lin J, Liu Q, Song Y, Liu J, Yin Y, Hall NS, *et al.* Temporal prediction of coastal water quality based on environmental factors with machine learning. *J Mar Sci Eng.* 2023; 11 (8): 1608. doi:10.3390/jmse11081608
  13. Ma P and Zhang Y. A time-varying copula approach for describing seasonality in multivariate ocean data. *Mar Struct.* 2024; 94: 103567. doi:10.1016/j.marstruc.2023.103567
  14. Simmer RA, Jansen EJ, Patterson KJ, and Schnoor JL. Climate change and the sea: A major disruption in steady state and the master variables. *ACS Environ Au.* 2023; 3 (4): 195–208. doi:10.1021/acsenvironau.2c00061
  15. Zhang K, Gençay R, and Yazgan ME. Application of wavelet decomposition in time-series forecasting. *Econ Lett.* 2017; 158: 41–46. doi:10.1016/j.econlet.2017.06.010
  16. Dawn A. Wavelet analysis to describe biological cycles and signals of non-stationarity. Oregon State University Geospatial Ecology of Marine Megafauna Laboratory Blog [Internet]. 2022-02-28; Available from: <https://blogs.oregonstate.edu/gemmlab/2022/02/28/wavelet-analysis-to-describe-biological-cycles-and-signals-of-non-stationarity/> (accessed on 2026-04-25).
  17. Rajae T and Boroumand A. Forecasting of chlorophyll-a concentrations in South San Francisco Bay using five different models. *Appl Ocean Res.* 2015; 53: 208–217. doi:10.1016/j.apor.2015.09.001
  18. Daubechies I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans Inf Theory.* 1990; 36 (5): 961–1005. doi:10.1109/18.57199
  19. California HABMAP. Data – California HABMAP [Internet]. Available from: <https://calhabmap.org/datasites> (accessed on 2026-01-19)
  20. National Data Buoy Center, National Oceanic and Atmospheric Administration. Scripps Pier, La Jolla, CA [Internet]. Available from: [https://www.ndbc.noaa.gov/station\\_page.php?station=ljpc1](https://www.ndbc.noaa.gov/station_page.php?station=ljpc1) (accessed on 2026-01-31)
  21. National Data Buoy Center, National Oceanic and Atmospheric Administration. Stearns Wharf, Santa Barbara, CA [Internet]. Available from: [https://www.ndbc.noaa.gov/station\\_page.php?station=ntbc1](https://www.ndbc.noaa.gov/station_page.php?station=ntbc1) (accessed on 2026-01-31)
  22. National Data Buoy Center, National Oceanic and Atmospheric Administration. Cal Poly Pier, San Luis Obispo, CA [Internet]. Available from: [https://www.ndbc.noaa.gov/station\\_history.php?station=cpxc1](https://www.ndbc.noaa.gov/station_history.php?station=cpxc1) (accessed on 2026-01-31)
  23. Kramar V and Alchakov V. Time-series forecasting of seasonal data using machine learning methods. *Algorithms.* 2023; 16 (5): 248. doi:10.3390/a16050248
  24. Wells ML, Trainer VL, Smayda TJ, Karlson BSO, Trick CG, Kudela RM, *et al.* Harmful algal blooms and climate change: Learning from the past and present to forecast the future. *Harmful Algae.* 2015; 49: 68–93. doi:10.1016/j.hal.2015.07.009
  25. Chen C, Hu M, Chen Q, Zhang J, Feng T, and Cui Z. Long-term trend forecast of chlorophyll-a concentration over eutrophic lakes based on time series decomposition and deep learning algorithm. *Sci Total Environ.* 2024; 951: 175451. doi:10.1016/j.scitotenv.2024.175451
  26. Nieves V, Radin C, and Camps-Valls G. Predicting regional coastal sea level changes with machine learning. *Sci Rep.* 2021; 11: 7650. doi:10.1038/s41598-021-87460-z
  27. Alvarez-Cuesta M, Toimil A, and Losada I. Reduced-complexity modeling as a valuable tool for studying the coastal impacts of climate change. EGU General Assembly 2022, Vienna, Austria. 2022; doi:10.5194/egusphere-egu22-11246. Available from: <https://meetingorganizer.copernicus.org/EGU22/EGU22-11246.html> (accessed on 2026-01-19)
  28. Berk D. Understanding harmful algal blooms in California [Internet]. San Diego Coastkeeper Blog. Available from: <https://www.sdcoastkeeper.org/blog/understanding-harmful-algal-blooms-in-california/> (accessed on 2026-01-24)
  29. Kenitz K and Anderson C. California HAB bulletin: March–April 2025. Southern California Coastal Ocean Observing System [Internet]. 2025; Available from: <https://sccoos.org/california-hab-bulletin/march-april-2025/> (accessed on 2026-01-24)
  30. NOAA Staff. Early bloom of toxic algae off Southern California sickens hundreds of sea lions and dolphins. U.S. National Oceanic and Atmospheric Administration Fisheries Feature Story [Internet]. 2025; Available from: <https://www.fisheries.noaa.gov/feature-story/early-bloom-toxic-algae-southern-california-sickens-hundreds-sea-lions-and-dolphins> (accessed on 2026-01-24)
  31. NPR Staff. Sea lions are released after toxic algae bloom in California. KPBS Public Media [Internet]. 2025-07-08. Available from: <https://www.kpbs.org/news/science-technology/2025/07/08/sea-lions-are-released-after-toxic-algae-bloom-in-california> (accessed on 2026-01-24)
  32. Jin D, Kourantidou M, Weir MJ, and Horstmann I.

- Assessing the value of harmful algal bloom forecasts in the Pacific Northwest. *ICES J Mar Sci.* 2024; 81 (9): 1796–1816. doi:10.1093/icesjms/fsae126
33. Yang J, Zheng Y, Zhang W, Zhou Y, and Zhang Y. Comparative analysis of machine learning methods for prediction of chlorophyll-a in a river with different hydrology characteristics: A case study in Fuchun River, China. *J Environ Mgmt.* 2024; 364: 121386. doi:10.1016/j.jenvman.2024.121386
34. Rajaei T, Khani S, and Ravansalar M. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Lab Syst.* 2020; 200: 103978. doi:10.1016/j.chemolab.2020.103978
35. Qin M, Xu R, Xu L, Zhao B, Chen X, and Du Z. Multi-step forecasting of chlorophyll-a concentration in coastal waters through Wavelet Dense Attention Transformer model. *Environ Res.* 2026; 292: 123627. doi:10.1016/j.envres.2025.123627
36. González-Enrique J, Rodríguez-García MI, Ruiz-Aguilar JJ, Carrasco-García MG, Enguix IF, and Turias JJ. Chlorophyll- $\alpha$  forecasting using LSTM, bidirectional LSTM and GRU networks in El Mar Menor (Spain). *Logic J IGPL.* 2025; 33 (5): jzae046. doi:10.1093/jigpal/jzae046
37. Chen X, Banno M, and Mori N. Less Is More: Short-Term Window Calibration Improves Seasonal Shoreline Prediction in Modeling. *Geophys Res Ltrs.* 2025; 52 (17): e2025GL117764. doi:10.1029/2025GL117764
38. Flinchem EP and Jay DA. An Introduction to Wavelet Transform Tidal Analysis Methods. *Estuarine, Coastal and Shelf Sci.* 2000; 51 (2): 177–200. doi:10.1006/ecss.2000.0586
39. Liu J, Wang X. A Wavelet-Based Deep Learning Framework for Predicting Peak Intensity of Hurricanes in the Atlantic Ocean. *Atmo and Clim Sci.* 2026; 13 (04): 587–606. doi:10.4236/acs.2023.134033
40. Zhou T, Wang F, and Yang Z. Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. *Water.* 2017; 9 (10): 781. doi:10.3390/w9100781