Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Modeling Dynamics of Kelp Forests Along the West Coast of North America: Integrating Environmental and Anthropogenic Drivers

Jingjing Tao

tao121@uw.edu

University of Washington

Caleb Zheng

University of Washington

Sergey Rabotyagov

University of Washington

Chris Anderson

University of Washington

Research Article

Keywords: Kelp forest, time series analysis, coupled human–environment systems, spatiotemporal modeling, recurrent neural network

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11380710/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Modeling Dynamics of Kelp Forests Along the West Coast of North America: Integrating Environmental and Anthropogenic Drivers

Jingjing Tao[*1], Caleb Zheng[2], Sergey Rabotyagov[1], Chris Anderson[3]

[1]*School of Environmental and Forest Sciences, University of Washington, Seattle, United States*

[2]*Department of Electrical & Computer Engineering, University of Washington, Seattle, United States*

[3]*School of Aquatic and Fishery Sciences, University of Washington, Seattle, United States*

[*]Correspondence: tao121@uw.edu

Abstract

Kelp forests are vital marine ecosystems facing escalating threats from climate change and human activities. Using 41 years (1984–2024) of quarterly kelp canopy data from 585,408 sites, we apply time series models with environmental and anthropogenic covariates to characterize spatiotemporal biomass dynamics and compare forecasting performance. Among univariate models, an autoregressive state-space model has the lowest in-sample error and a seasonal autoregressive integrated moving-average model with covariates (SARIMAX) gives the best short-term forecasts. In the multivariate setting, multivariate autoregressive state-space with covariates (MARSSX) model reveals two ecologically distinct latent regimes separated by Point Conception: southern kelp clusters show high-amplitude, disturbance-prone volatility, while northern clusters exhibit smoother, inertia-driven cycles. Marine heatwave intensity is associated with lower latent kelp canopy whereas nutrient-rich upwelling with higher latent kelp canopy across models; the association with recreational fishing effort is small and inconsistent in sign. Long short-term memory (LSTM) neural network improves short-term forecasting accuracy over the linear state-space model, indicating that nonlinear temporal dependencies and cross-region interactions add predictive value. This complementary framework reveals a tradeoff between interpretability and predictability: MARSSX model provides ecologically meaningful regime structure and covariate associations, while LSTM offers stronger predictive performance. Our results provide decision-relevant information for early warning of kelp decline, spatial prioritization of monitoring and restoration, and region-specific decision support for coastal managers and conservation planners.

Keywords

Kelp forest, time series analysis, coupled human–environment systems, spatiotemporal modeling, recurrent neural network

Introduction

Kelp forests are among the most extensive and productive coastal habitats and provide critical ecosystem services, including carbon sequestration, nutrient cycling, coastline protection from storms, water quality improvement, and cultural benefits to coastal communities, contributing substantial ecological and economic value[1,2]. Characterized by high primary productivity and structurally complex habitats, kelp forests support a wide variety of ecologically and commercially important marine species[3] and predominantly occur in temperate and polar coastal regions, where cold, nutrient-rich waters foster rapid growth. However, kelp abundance can fluctuate dramatically due to their sensitivity to environmental burdens[4], and oceanographic conditions, biological interactions, and human activities influence their stability and resilience[5].

Kelp forests are increasingly threatened by climate change, which manifests through ocean warming, acidification, and intensified extreme weather[6–8]. Human disturbances, including commercial and recreational fishing, coastal development, and pollution, further degrade kelp habitats and alter population dynamics[9], undermining both the ecological integrity and the socio-economic systems that depend on them[10–12]. Yet our understanding of how kelp forests dynamically respond to these coupled human–environmental pressures remains limited. There is a pressing need to investigate these interactions in an integrated framework to inform conservation and adaptive management.

Despite a growing body of kelp forest research, a literature gap remains in the dynamic integration of co-occurring environmental and anthropogenic stressors[18,25]: most existing studies

isolate individual drivers or focus on localized case studies, limiting understanding of how these systems respond to interacting pressures over broad spatial and temporal scales. Existing regional studies have characterized canopy trends and heatwave responses region by region[26], we model the regions jointly, evaluating whether regions share dynamics, and testing a number of models on out-of-sample forecast skill. The west coast of North America, home to some of the world's largest and most ecologically important kelp forests[27], has received limited attention in long-term, empirical, time series-based modeling efforts, though this region is essential for studying kelp forest resilience due to its oceanographical, ecological variability, and exposure to a wide range of stressors[28,29]. A key open question is whether model structures differing in interpretability and flexibility reveal different, ecologically relevant aspects of kelp dynamics. Autoregressive models with lags allow us to test whether heat waves leave a lasting biological footprint on kelp, or whether responses are largely contemporaneous. Latent state-space models with spatially structured regimes allow us to ask whether the magnitude of forcing differs among ecological regimes. Comparing these frameworks anchors model evaluation in ecological questions, such as the persistence and spatial structure of heat wave effects, rather than in statistical fit alone.

Two canopy-forming kelps occur along this coast, and they differ in ways that matter for how kelp abundance varies in space and time. Giant kelp (*Macrocystis pyrifera*) is a perennial that regrows from a persistent holdfast and continually replaces its fronds, but it is thinned by large storm waves and by warm, nutrient-poor water[3,13]. It dominates from central California southward and reaches its warm (equatorward) range limit in Baja California, Mexico[14]. Bull kelp (*Nereocystis luetkeana*) is an annual (occasionally biennial) alga that must re-establish from microscopic stages every year, so its canopy depends strongly on conditions during the spring–summer growth window; it becomes dominant north of about Santa Cruz in central California and extends to Alaska[15]. The satellite canopy record analyzed here does not distinguish the two species[16,17]; because giant kelp is the dominant surface-canopy species over most of the latitudes represented in our sites (central California to Baja California), we treat the data as a single kelp canopy signal and use species ecology only to interpret it. Both species rely on cold, nutrient-rich water delivered by coastal upwelling, so nutrient supply and temperature are central drivers of their canopy in space and time[15,18].

Kelp canopy also responds to biological interactions that vary along the coast. Sea urchins are the main grazers: where their predators are abundant, urchins stay hidden in crevices and kelp forests persist, but when predators decline urchins can graze forests down to persistent "urchin barrens"[19]. Sea otters (*Enhydra lutris*), whose control of urchins was first demonstrated in Alaska[39], occur in central California but not in southern California, where predatory fishes such as California sheephead and spiny lobsters help regulate urchins and are themselves targeted by fisheries[20]. Two large events during our study period illustrate how physical and biological drivers combine. From late 2013 to 2016, a persistent warm-water anomaly in the northeast Pacific (the Blob), followed by the 2015–2016 El Niño, produced the largest marine heatwave recorded in the region[21]; its effects on giant kelp were immediate and severe near the species' southern limit in Baja California but delayed or smaller farther north in southern California[14]. Beginning in 2013, sea star wasting disease killed more than 20 sea star species from Mexico to Alaska, and the sunflower star (*Pycnopodia helianthoides*), an important urchin predator, declined by 80–100% across its range, with the largest declines coinciding with warm anomalies[22]. Combined with warm water, this predator loss contributed to purple sea urchin (*Strongylocentrotus purpuratus*) outbreaks and, in northern California, to the collapse of bull

kelp forests into urchin barrens[23,24]. That collapse lies mostly north of the sites in our dataset, but it shows the type of multi-driver event (temperature, disease, predation) that a satellite canopy time series records only as a decline in canopy; it is therefore used here as context for interpretation and for the cautious framing of our results, not as a mechanism the models test.

This study addresses literature gaps with an integrated system modeling framework to characterize kelp canopy variability across the West Coast, providing a forecasting and decision-support tool that can inform monitoring and restoration priorities and early-warning signals of decline without requiring causal identification of underlying mechanisms. Utilizing long-term time series data with environmental, biological, and anthropogenic covariates, we characterize spatiotemporal patterns of kelp biomass variability and evaluate the predictive utility of different model structures, including their ability to reveal ecologically meaningful regime differences across biogeographic boundaries.

To make explicit what each model contributes, we select complementary models to address analytical questions. A seasonal autoregressive integrated moving average regression with covariates (SARIMAX) asks how well a coast-wide kelp series can be forecast from its own annual cycle, year-to-year persistence and a few drivers. A univariate autoregressive state-space model (UARSSX) separates real changes in kelp abundance from satellite measurement noise. A generalized additive model (GAM) allows drivers to act non-linearly. A multivariate state-space model (MARSSX) lets stretches of coast share or differ in their unobserved kelp state, which evaluates whether Point Conception, a known biogeographic boundary, separates ecological regimes. Finally, a long short-term memory network (LSTM) serves as a flexible benchmark that shows how much predictive skill the linear models leave unused.

We make three key contributions. First, while prior studies have characterized spatial or temporal trends in kelp canopy[13,23,30-33], few studies have applied multivariate state-space forecasting to the full extent of the West Coast with environmental and anthropogenic covariates over a multi-decade horizon. Second, rather than relying on a single model class or treating all models as competing alternatives, we we use complementary modeling approaches to address distinct objectives. Linear models support ecological inference by estimating latent ecosystem states, quantifying covariate relationships, and identifying regime structure across regions, while nonlinear models as flexible benchmark capture temporal dependencies and interactions. Third, by explicitly modeling latent ecological regimes separated by Point Conception, our interpretable state-space framework characterizes spatially structured dynamics that provide an empirical basis for region-specific monitoring and planning.

Methods

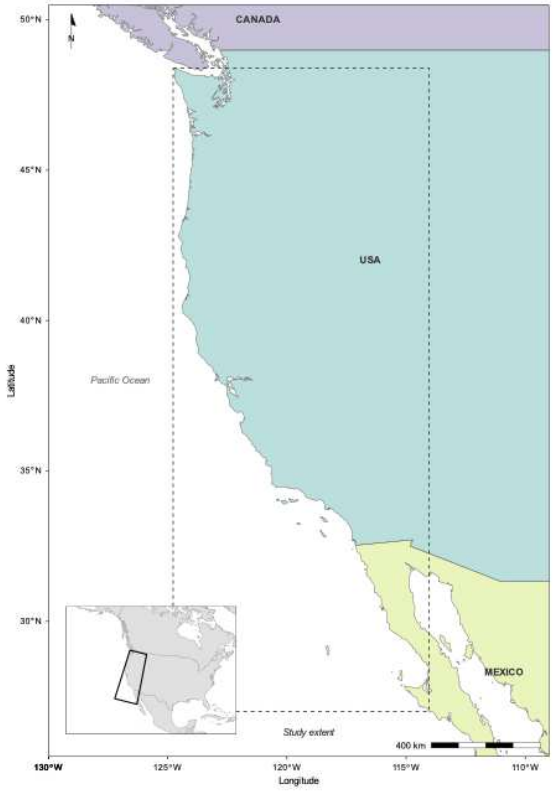
Study Region & Data

This study investigates kelp forest dynamics along the west coast of North America, also referred to as the California Current Region, spanning from the U.S.–Canada border to Baja California, Mexico. The response variable is kelp forest biomass, encompassing two canopy-forming macroalgae species: giant kelp (Macrocystis pyrifera) and bull kelp (Nereocystis luetkeana), which are particularly abundant along the eastern Pacific coastline[34]. Kelp canopy area is used as a proxy for above-ground kelp biomass. The satellite product reports surface canopy and does not distinguish the two species[16,17]; the two differ in life history and geographic dominance, but our analysis treats the data as a single kelp canopy signal.

The kelp biomass dataset covers 585,408 spatial sites across the study region (Fig. 1), with each 30-m Landsat pixel contributing quarterly observations from 1984 to 2024. These data

are derived from multispectral imagery collected by the Landsat satellite program, including the Thematic Mapper (TM; Landsat 4 and 5), Enhanced Thematic Mapper Plus (ETM+; Landsat 7), and Operational Land Imager (OLI; Landsat 8). The raw imagery is processed and standardized into a harmonized kelp canopy time series, providing a long-term, remotely sensed kelp biomass dataset[16,35].

Fig. 1: Spatial coverage of kelp biomass sites used in this study (585,408 sites, 1984–2024).



Note. The dashed rectangle marks the study extent (27.01°N–48.40°N, 124.77°W–114.04°W), which runs from near the U.S.–Canada border to Baja California, Mexico. Coordinates are in decimal degrees (WGS84), and the scale bar is approximate. Basemap data: Natural Earth.

Covariates

Guided by kelp's known sensitivity to external stressors[36,37], the study incorporates eleven covariates: marine heatwave intensity (HEAT), biologically effective upwelling transport index (BEUTI, a proxy for nitrate availability), Pacific Decadal Oscillation (PDO), sea surface temperature (SST), Oceanic Niño Index (NINO), zooplankton biomass (ZOO), angler trips and recreational fishing effort (ANGLER), total fishery harvest (HARVEST), coastal population (POP), beach closures due to contamination or hazards (BEACH), and billion-dollar disaster events (DISASTER). All covariates originally at monthly resolution are aggregated to quarterly values by averaging to match the temporal scale of the response variable. The ecological rationale and the expected direction of effect of each covariate are summarized in Extended Data Table 1.

Missing Value Imputation

Missing values in several covariate time series are addressed by treating imputation as a model selection problem for each covariate individually. Four candidate models are evaluated:

Autoregressive Integrated Moving Average (ARIMA), Generalized Additive Model (GAM), Error-Trend-Seasonality (ETS), and a stochastic level Dynamic Linear Model (DLM). Model performance is evaluated by Akaike Information Criterion (AIC) and diagnostic checks of residual normality and autocorrelation, and the best-balanced model is selected for each covariate (Extended Data Table 2). Selecting the model separately for each covariate allows series with different behavior to be filled in appropriately.

Covariate Selection

Following imputation, covariate selection identifies the most relevant predictors for model training. To detect nonlinear relationships and interactions among covariates, three machine learning models are trained on 70% of the data and compared using a 15% validation holdout, with a further 15% reserved for out-of-sample testing: Random Forest (RF), Lasso Regression, and a single-hidden-layer feedforward Neural Network (NN) with sigmoid activation. Hyperparameters are tuned via 6-fold cross-validation, and predictive performance is assessed using root mean squared error (RMSE) on the held-out test set. The feature set is a multivariate time series of interpolated covariates, and the response variable is the total kelp canopy area across all 30-m Landsat pixels in the study domain (quarterly mean per pixel). RF hyperparameters and the LASSO penalty parameters are tuned via 6-fold cross-validation; NN hidden-unit count and weight-decay regularization are likewise tuned via 6-fold cross-validation. All interpolated and selected covariates are z-score normalized, facilitating model convergence and comparability across variables measured on different scales; the response variable is log-transformed to stabilize variance, which enhances the robustness and accuracy of time series model fitting. Using three methods with different assumptions (tree-based, neural and penalized linear) reduces the chance that a covariate is selected because of the idiosyncrasies of a single method.

The structures and specifications of the machine learning models used for covariate selection and comparison are described as follows:

Random Forest (RF):

$$Y_t = f_{RF}(X_{1t} \dots X_{jt}) + \epsilon_t$$

where f_{RF} is an ensemble of regression trees. The key hyperparameter (the number of predictor variables randomly selected as candidates at each split) is optimized using 6-fold cross-validation.

Lasso Regression (LASSO):

$$L(\beta) = \prod_{j=1}^{j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_t - \boldsymbol{X}_t\beta)^2}{2\sigma^2}\right)$$

$$\min_{\beta} \sum_{j=1}^{j} (Y_t - \boldsymbol{X}_t\beta)^2 + \lambda \sum_{j=1}^{j} |\beta_j|$$

where the penalty parameter λ is tuned over a 6-fold cross-validation.

Neural Network (NN):

$$Y_t = \alpha_0 + \sum_{m=1}^{M} \alpha_m \sigma\left(\sum_{j=1}^{j} w_{mj} X_t + b_m\right) + \epsilon_t$$

where we train a single-hidden-layer feedforward network with M hidden units and w weight-decay regularization, both tuned via 6-fold cross-validation. The activation function $\sigma(\cdot)$ is a sigmoid function.

Non-Spatial Time Series Models

We compare three univariate model frameworks that incorporate covariates while accounting for autocorrelation and temporal structure. The three frameworks are chosen to represent complementary assumptions: SARIMAX emphasizes seasonal and autoregressive persistence, UARSSX separates ecological change from measurement noise, and GAM allows flexible nonlinear driver effects without explicit temporal dynamics.

Seasonal ARIMA model with covariates (SARIMAX):

$$Y_t = \beta_0 + \mu_t + \sum_{j \in \mathcal{J}} \beta_j \mathcal{X}_{j\, t-\ell_j}^{T} + \epsilon_t,\ \epsilon_t \sim N(0, \sigma^2)$$

$$\Phi(\mathcal{B}^s)\phi(\mathcal{B})(1-\mathcal{B})^d(1-\mathcal{B}^s)^D\left(Y_t - \sum_{j \in \mathcal{J}} \beta_j \mathcal{X}_{j\, t-\ell_j}^{T} - \beta_0\right) = \Theta(\mathcal{B}^s)\theta(\mathcal{B})\epsilon_t$$

$$\phi(\mathcal{B}) = 1 - \phi_1\mathcal{B} - \cdots - \phi_p\mathcal{B}^p$$

$$\theta(\mathcal{B}) = 1 + \theta_1\mathcal{B} + \cdots + \theta_q\mathcal{B}^q$$

$$\Phi(\mathcal{B}^s) = 1 - \Phi_1\mathcal{B}^s - \cdots - \Phi_P\mathcal{B}^s$$

$$\Theta(\mathcal{B}^s) = 1 + \Theta_1\mathcal{B}^s + \cdots + \Theta_Q\mathcal{B}^s$$

where Y_t denotes the observed kelp biomass. $\mathcal{X}_{j\, t-\ell_j}$ represent the vector of selected covariates at their optimal number of lags. ϵ_t is modeled with white noise. μ_t captures the seasonal ARIMA process. $\mathcal{B}$ is the backshift operator and $s = 4$ denotes quarterly seasonality. d and D represent the orders of non-seasonal and seasonal differencing orders. Model orders are determined automatically via the Hyndman-Khandakar algorithm, an automatic procedure for selecting ARIMA orders, with stepwise and exact search procedures.

Univariate autoregressive state-space model with external covariates (UARSSX):

$$x_t = x_{t-1} + C\boldsymbol{c}_t + w_t,\ w_t \sim N(0, Q)$$

$$Y_t = x_t + D\boldsymbol{d}_t + a + v_t,\ v_t \sim N(0, R)$$

where x_t is a one-dimensional latent state vector and Y_t is the observed kelp biomass. $\boldsymbol{c}_t$ includes selected covariates and Fourier seasonal assumed to influence the kelp latent state. $\boldsymbol{d}_t$ comprises covariates believed to affect the observed kelp biomass directly. a is an additive

observation intercept, accounting for systematic bias or offset between the latent state and the observed data, as well as helps adjust for average mismatch. Q captures natural variability, unmodeled processes, and environmental stochasticity. R is the observation noise, accounting for measurement error and imprecision not explained by the covariates. The model assumes an unobserved (latent) kelp condition that evolves through time in response to the covariates and random shocks (variance Q), and that each satellite observation equals this condition plus measurement error (variance R).

Generalized Additive Model (GAM):

$$E(Y_t) = \beta_0 + \sum f_j(\mathrm{X}_{jt}) + f(T) + f(s) + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2)$$

where $E(Y_t)$ is the expected value of kelp biomass, and X_{jt} is the vector of selected covariates. T is the trend variable and s represents the seasonal term. The function $f(\cdot)$ represents smooth functions utilized for covariates, allowing for non-linear effects. The error term ε_t is modeled as white noise.

Model evaluation separates in-sample fit from out-of-sample forecast accuracy. In-sample fit is assessed by fitting each model to the full 1984–2024 series and comparing fitted with observed values. Out-of-sample accuracy is assessed with rolling-origin (forward-chaining) one-step-ahead forecasts: each model is first fitted to the initial 30 years of data (1984–2013) and used to forecast the next quarter, then re-fitted with the newly observed quarter added, repeating through 2024Q4 (44 forecasts, 2014Q1–2024Q4). Each forecast uses the observed same-quarter covariate values (lag 0), so these are conditional one-step-ahead forecasts. Kelp canopy area is standardized (z-scored) before model fitting. Performance is assessed via AIC, root mean squared error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE). These criteria collectively evaluate model fit, predictive accuracy, and relative error magnitude.

To ensure the validity of model assumptions and to detect potential misspecifications, residual diagnostics are systematically conducted, including autocorrelation function (ACF), residual histogram and Q-Q plots, Ljung-Box test for independence, and Shapiro–Wilk test for normality.

Spatial Exploratory Analysis & Clustering

We assess spatial structure using Global Moran's I to quantify overall spatial autocorrelation, and Local Indicators of Spatial Association (LISA) to identify local hot spots, cold spots, and spatial outliers. Given the scale of the dataset, we apply Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to aggregate sites into a manageable number of spatial clusters, aiming to capture dominant spatiotemporal patterns while filtering site-level noise. DBSCAN is well suited to the irregular, non-convex spatial features of coastal kelp beds and, unlike centroid-based methods such as k-means, relies on point density to reflect spatial continuity and patchiness. DBSCAN parameters are informed by k-nearest-neighbor distance distributions and the spatial-autocorrelation distance-decay curve, using the elbow method to balance cluster granularity against noise exclusion.

Spatial State-Space Model

We fit a Multivariate Auto-Regressive State-Space with covariates (MARSSX):

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{C}\boldsymbol{c}_t + \boldsymbol{w}_t,\ \boldsymbol{w}_t \sim N(0, \boldsymbol{Q})$$

$$\boldsymbol{Y}_t = \boldsymbol{Z}\boldsymbol{x}_t + D d_t + \boldsymbol{a} + \boldsymbol{v}_t,\ \boldsymbol{v}_t \sim N(0, \boldsymbol{R})$$

where $\boldsymbol{x}_t$ is an m-dimensional latent state vector; $\boldsymbol{Y}_t$ represents the clustered multi-group kelp observations; $\boldsymbol{c}_t$ includes selected covariates and Fourier seasonal terms influencing the kelp latent state over time; $\boldsymbol{d}_t$ consists of covariates affecting the measurement of kelp biomass; a is the drift term accounting for potential sensor degradation when merging data across TM, ETM+, and OLI sensors; $\boldsymbol{Q}$ governs the latent-state evolution uncertainty and it is modeled either as diagonal unequal (each state having its own variance) or unconstrained (fully estimating covariance between trends); $\boldsymbol{R}$ is observation noise variance, assumed to have a diagonal unequal structure; $\mathbf{Z}$ maps each cluster to its corresponding latent state. Each latent state is an unobserved regional kelp condition; the mapping Z assigns each of the 16 clusters to one state, so clusters assigned to the same state are assumed to share the same underlying trend.

We consider $m = \{1,2\}$ latent states, testing whether all kelp clusters share one underlying trend or separate into two trends divided by Point Conception: a well-documented oceanographic transition zone[38-40]. Point Conception was chosen a priori because it marks a transition between cooler, upwelling-dominated waters to the north and the warmer Southern California Bight to the south, and a change in the composition of marine communities[38,39].

Spatial Neural Network Model

To complement the interpretable MARSSX framework, we implemented an LSTM network as a nonlinear spatiotemporal forecasting model[41]. LSTM is well suited for ecological time-series analysis because it can capture long-range temporal dependencies and complex nonlinear relationships. It is a recurrent neural network with "gates" that learn which information from previous quarters to keep, update or forget when predicting the next quarter.

Let $Y_t \in \mathbb{R}^n$ denote the vector of clustered kelp biomass at quarter t, where n is the number of spatial clusters, and let $c_t \in \mathbb{R}^p$ denote the vector of selected covariates. The input at each time step is defined as the concatenated vector $s_t = [Y_{t-1}, c_t] \in \mathbb{R}^{n+p}$. Given a lookback window of length L, the model takes the sequence $S_t = (s_{t-L+1}, \ldots, s_t)$ as input and predicts kelp biomass in the next quarter. The LSTM updates are given by

$$f_\tau = \sigma\left(W_f s_\tau + U_f h_{\tau-1} + b_f\right)$$
$$i_\tau = \sigma(W_i s_\tau + U_i h_{\tau-1} + b_i)$$
$$\tilde{m}_\tau = \tanh(W_m s_\tau + U_m h_{\tau-1} + b_m)$$
$$m_\tau = f_\tau \odot m_{\tau-1} + i_\tau \odot \tilde{m}_\tau$$
$$o_\tau = \sigma(W_o s_\tau + U_o h_{\tau-1} + b_o),$$

and $h_\tau = o_\tau \odot \tanh(m_\tau)$. where $\odot$ is element-wise product, f_τ, i_τ, and o_τ are forget, input, and output gates, m_τ is the cell state, and h_τ is the hidden state. The final hidden state is then mapped to the forecast through a dropout layer and a fully connected output layer,

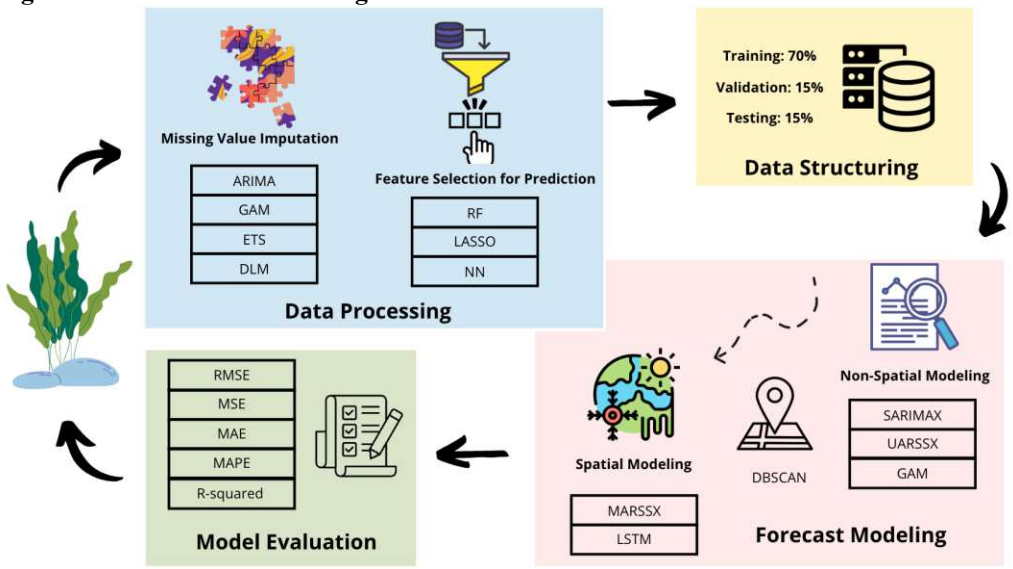
$$\hat{Y}_t = W_y\, Dropout(h_t) + b_y.$$

Hyperparameters are selected on validation-set performance: an 8-quarter lookback window, a 2-layer LSTM with hidden dimension 4, dropout rate 0.2, learning rate 5×10^{-5}, weight decay 10^{-4}, batch size 16, Adam optimization with mean-squared-error loss, and early stopping

with patience 50. Model performance is evaluated on the held-out test set using RMSE, MSE, MAE, and MAPE.

Fig. 2 summarizes the full modeling workflow, from time-aware imputation through predictive feature selection, the 70/15/15 training/validation/test split, and the two complementary modeling branches (non-spatial: SARIMAX, UARSSX, GAM; spatial: MARSSX, LSTM).

Fig. 2: Overview of the modeling workflow.



Note. Colored boxes show the main stages: (blue) data processing, in which gaps in covariates are filled using the best of four time series models (ARIMA, GAM, ETS, DLM) and covariates are screened for predictive value (RF, LASSO, NN); (yellow) data structuring, with a 70/15/15% split into training, validation and test sets; (pink) forecast modeling, with non-spatial models for the coast-wide series (SARIMAX, UARSSX, GAM) and spatial models for the 16 DBSCAN clusters (MARSSX, LSTM); (green) model evaluation using RMSE, MSE, MAE, MAPE and R^2. Arrows indicate the order of steps, and the curved arrow at left indicates iterative refinement. All abbreviations are defined in Extended Data Table 5.

Results

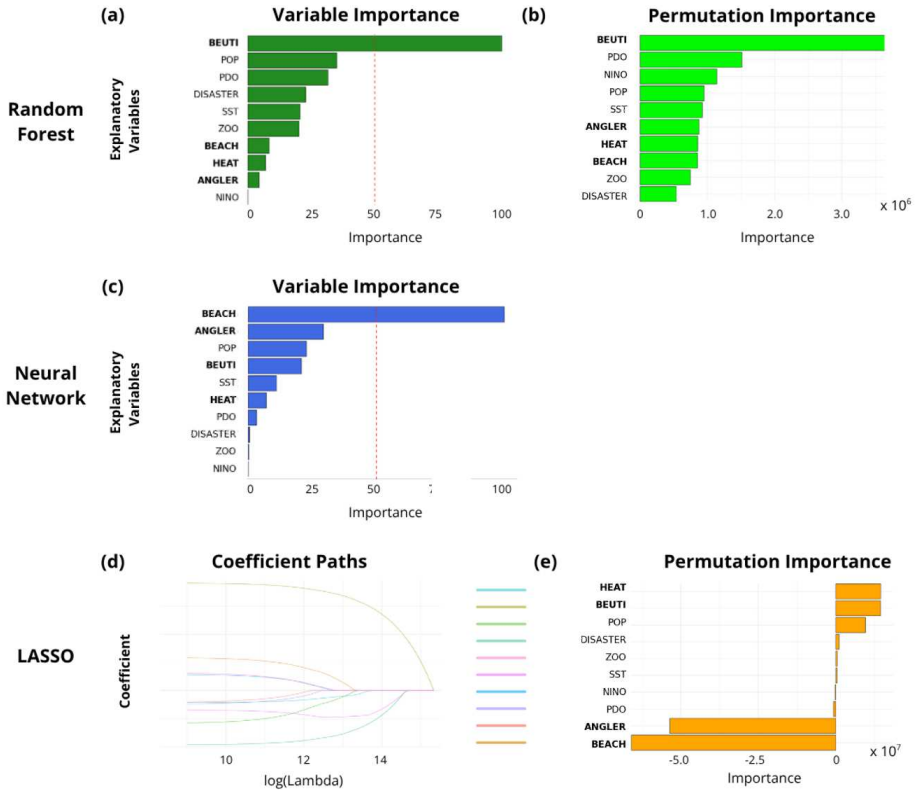
Missing Value Imputation

Seven of the eleven covariates (beach closures, BEACH; fishery harvest, HARVEST; coastal population, POP; billion-dollar disaster events, DISASTER; the biologically effective upwelling transport index, BEUTI, a measure of nitrate supply by coastal upwelling; sea surface temperature, SST; and zooplankton biomass, ZOO) contain gaps that require imputation. We individually fit four candidate models (ARIMA; GAM; error–trend–seasonality exponential smoothing, ETS; and dynamic linear model, DLM) to each covariate's time series and compare them by Akaike information criterion (AIC) and residual diagnostics. ARIMA consistently outperforms the alternative approaches across all seven covariates (Extended Data Table 2), demonstrating superior goodness-of-fit for univariate time series interpolation and robust capacity to capture temporal dependencies and seasonal patterns. Complete imputed time series for each covariate are shown in Supplementary Fig. S1, illustrating the effectiveness of the imputation approach and the continuity of the reconstructed datasets. The imputed values preserve each covariate's seasonal pattern and persistence rather than introducing artificial jumps.

Covariate Selection

We evaluate candidate covariates using three machine learning models: random forest (RF), a feedforward neural network (NN), and LASSO regression. The three models do not yield a single consistent ranking (Fig. 3), so we use them as a screening tool rather than a definitive ordering. BEUTI is the most robust covariate. It ranks first in RF under both mean decrease in impurity and permutation importance (Fig. 3a, b), retains large coefficients across a broad range of penalty strengths in the LASSO path (Fig. 3d), and has positive LASSO permutation importance (Fig. 3e), although it falls below the scaled midpoint in the NN (Fig. 3c). BEACH and ANGLER rank first and second in the NN (Fig. 3c) but rank low in RF (Fig. 3a, b). In LASSO they have the largest permutation-importance magnitudes, but with negative values, meaning that shuffling them did not degrade and even improved predictions (Fig. 3e). HEAT has the highest LASSO permutation importance (Fig. 3e) but ranks low in RF and the NN. We therefore retain BEUTI, BEACH, ANGLER, and HEAT based on a combination of predictive evidence and ecological coverage rather than unanimous rankings. HEAT is the focal exposure of this study, and together the four covariates represent nutrient supply (BEUTI), coastal contamination (BEACH), recreational fishing effort (ANGLER), and marine heatwave intensity (HEAT), spanning bottom-up, human, and climatic pathways. We exclude SST, the Pacific Decadal Oscillation (PDO), and the Oceanic Niño Index (NINO) because they describe ocean thermal conditions that overlap conceptually with HEAT, and including them risks collinearity, even though PDO and NINO rank moderately high in RF permutation importance. We also exclude human population (POP), which ranks moderately high in several models, because it is a coarse proxy for human pressure that overlaps with the more specific ANGLER and BEACH pathways. DISASTER and ZOO ranked low in most models. Importance in these models reflects usefulness for prediction, not causal influence.

Fig. 3: Variable importance rankings from random forest (a, b), neural network (c), and LASSO (d, e) covariate-selection models.



Note. Importance is shown for the eleven candidate covariates (acronyms are defined in Extended Data Table 1). (a) RF importance based on mean decrease in impurity, scaled from 0 to 100. (b) RF permutation importance, i.e., how much prediction accuracy declines when a covariate's values are randomly shuffled (axis $\times 10^6$). (c) NN importance, scaled from 0 to 100. (d) LASSO coefficient paths: each line is one covariate and shows how its coefficient shrinks to zero as the penalty (log λ, increasing to the right) grows; covariates that persist at larger penalties are more important. (e) LASSO permutation importance (axis $\times 10^7$). In (a) and (c) the dashed red line marks the midpoint (50) of the scaled importance axis. Covariates in bold type (BEUTI, BEACH, ANGLER, HEAT) were retained for subsequent modeling.

Exploratory Analysis

Augmented Dickey–Fuller tests, which test whether a series has a stable mean and variance over time, indicate that kelp biomass, BEUTI, and HEAT are stationary ($p = 0.02$, $p = 0.01$, $p=0.03$), while BEACH and ANGLER are non-stationary ($p = 0.24$ and $p = 0.42$), requiring first differencing. Seasonal and Trend decomposition using Loess (STL), which splits a series into seasonal, long-term trend and remainder components, plots exhibit BEUTI has a high seasonal variance ratio (66.76%; Supplementary Fig. S2, Supplementary Table S2), so it is de-seasonalized prior to modeling to mitigate potential confounding effects. No strong pairwise correlations are detected among selected covariates (Supplementary Fig. S3), and cross-correlation and likelihood-based lag selection indicate that contemporaneous (lag-0) covariate values produce the best model fit (Supplementary Table S1). Ecologically, the strong seasonality of upwelling (BEUTI) is expected because the winds that drive upwelling peak in spring and summer, whereas long-term trends dominate the human-pressure indicators (BEACH, ANGLER) and heatwave intensity (HEAT) is dominated by rare, episodic peaks, most notably around 2014–2016 (Supplementary Fig. S2). That same-quarter values fit best suggests that kelp canopy responds to these drivers within a season, although lagged effects cannot be excluded.

Non-Spatial Time Series Model Prediction Results

We compare three univariate frameworks incorporating covariates: SARIMAX, UARSSX, and GAM, each describing a single, coast-wide kelp series and evaluated via 30-year forward-chaining cross-validation. The best-fitting SARIMAX specification is a regression with ARIMA(5,0,0) errors, with all five AR terms significant at the 95% level (AR1–AR5 = 0.52, −0.35, 0.04, 0.51, −0.40). Among covariates, HEAT shows the strongest negative coefficient (−0.15), BEUTI a positive coefficient (0.08), and BEACH and ANGLER small negative coefficients (−0.09 and −0.03). Residual diagnostics (Supplementary Fig. S4), which check whether model errors are uncorrelated (Ljung–Box test) and approximately normal (Shapiro–Wilk test), show no significant autocorrelation (Ljung–Box $p = 0.04$) but moderate deviation from normality (Shapiro–Wilk $W = 0.96$, $p < 0.00$). Overall, the diagnostics suggest a well-fitting model with minor limitations in capturing higher-order distributional features. The negative HEAT coefficient indicates smaller canopy in quarters with stronger marine heatwaves, and the positive BEUTI coefficient indicates larger canopy when more nitrate is delivered by upwelling, in line with our expectations.

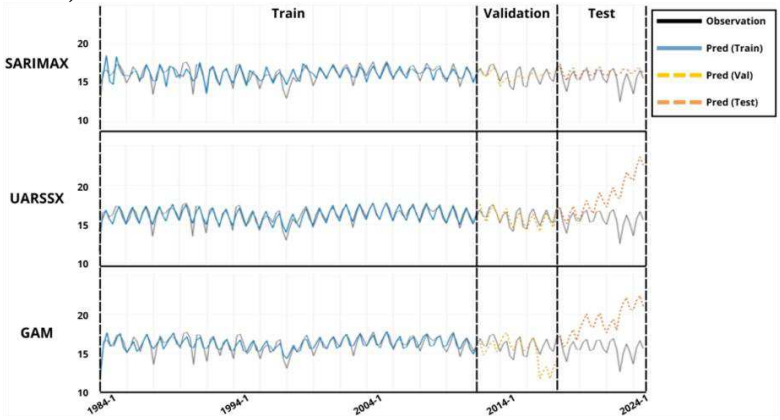
The best UARSSX model has observation scaling $A = 15.00$, observation error variance $R = 0.33$, and process error variance $Q = 0.03$, indicating that most unexplained variation resides in the observation component. Seasonal terms F1 and F2 (Fourier terms: paired sine and cosine curves that represent the annual cycle) have the largest coefficients (−0.85 and −1.32); HEAT is negative (−0.04), while BEUTI (+0.02) and ANGLER (+0.01) are positive. Residual diagnostics (Supplementary Fig. S5) show minor autocorrelation (Ljung–Box $p = 0.01$) and departures from normality in the tails (Shapiro–Wilk $p < 0.001$). While mild non-normality is common in ecological time series and may not greatly impact inference, it could influence uncertainty estimates and prediction intervals. A state-space model separates an unobserved "true" kelp condition, which evolves through time with process error (Q), from the noisy satellite measurements through which it is observed, with observation error (R). The finding that R greatly exceeds Q suggests that quarter-to-quarter variation in satellite canopy estimates, attributable to cloud cover, tide height, submersion of floating canopy by currents, and inter-sensor differences among Landsat platforms, exceeds variation in the underlying kelp state. This may explain why the model fits the observed series well yet has limited skill in forecasting short-term fluctuations in satellite-derived canopy estimates, which may reflect observation noise rather than actual changes in kelp condition.

The final GAM, with smooth terms for trend, seasonality, and each covariate, explains 58.4% of deviance. Seasonality (edf = 1.96, $F = 72.97$, $p < 0.001$) and the long-term trend (edf = 6.36, $F = 3.33$, $p < 0.001$) are highly significant, as is BEUTI (edf = 1.00, $F = 7.08$, $p = 0.01$); HEAT, ANGLER, and BEACH are not significant (all $p > 0.10$). Here edf is the effective degrees of freedom of a smooth term and values near 1 indicate an approximately linear effect. Residual diagnostics (Supplementary Fig. S6) indicate unresolved autocorrelation (Ljung–Box $p < 0.001$) and non-normality (Shapiro–Wilk $W = 0.94$, $p < 0.001$). While GAM captures some nonlinear seasonal and environmental effects, it likely misses key dynamic or autoregressive components inherent in the time series. In practical terms, the GAM recognizes the annual cycle and a long-term trend but not the way each quarter's canopy depends on previous quarters.

Across the three models (Extended Data Table 3; Fig. 4), UARSSX achieves the lowest in-sample AIC (353.23), RMSE (0.51), MAPE (2.60%), and highest R^2 (0.76). SARIMAX shows reasonable but weaker in-sample performance yet tracks short-term forecast changes more closely than the other two models with better out-of-sample performances (RMSE = 0.42; R^2 =

0.63). SARIMAX forecasts remain close to the observed level through the test period (2018–2024), whereas UARSSX and GAM forecasts drift upward over the longer horizon, so the model with the best fit is not the best long-range forecaster. The out-of-sample window spans the 2014–2016 marine heatwave, so model tuning was exposed to an extreme event. Ecologically, these results are consistent with the coast-wide kelp series being dominated by a strong annual cycle and short-term persistence, which autoregressive models capture better than the smooth, non-dynamic GAM.

Fig. 4: Fitted values and forecast performance across the three univariate models (SARIMAX, UARSSX, GAM).

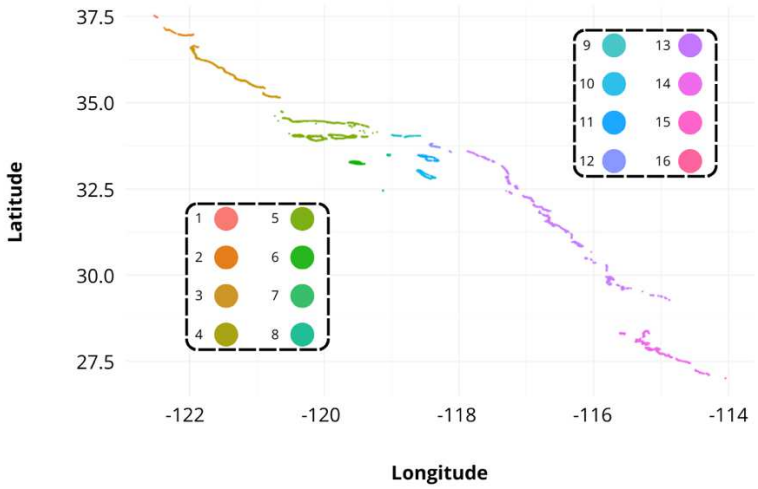


Note. The solid black/grey line is observed kelp canopy area (log-transformed response variable, as defined in Methods); the blue line is the fit to the training period; yellow dotted lines are predictions for the validation period; orange dotted lines are forecasts for the test period. Vertical dashed lines separate training, validation and test periods. X-axis labels give year and quarter.

Spatial Clustering

Global Moran's I and Local Indicators of Spatial Association (LISA) identified significant spatial clustering of kelp biomass, including hot spots, cold spots, and spatial outliers along the coast, as shown for the illustrative quarter of 2024-Q4 (Supplementary Fig. S7). To reduce the 585,408 site-level observations to a tractable number of spatial units, we applied DBSCAN clustering; k-nearest-neighbor distance and spatial-autocorrelation decay curves (Supplementary Fig. S8) indicated an optimal neighborhood radius (ε) between 0.05 and 0.1, yielding 16 distinct clusters from 585,408 sites (Fig. 5). The clusters correspond to contiguous stretches of coastline from central California (about 37.5°N) to central Baja California (about 27°N); sites not assigned to any cluster are not represented in the cluster-level models.

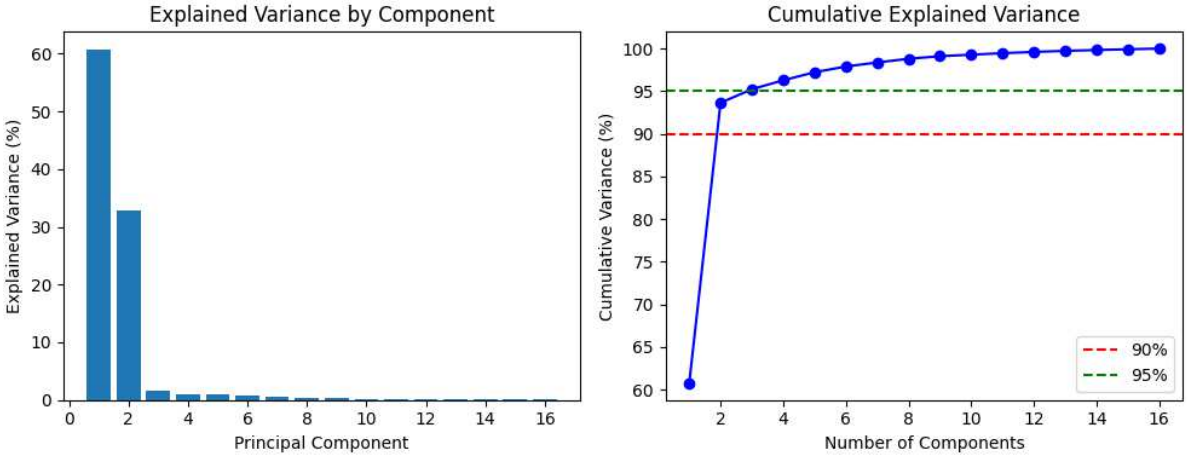
Fig. 5: Spatial distribution of the 16 DBSCAN clusters of kelp biomass in 2024-Q4, partitioned into northern (1–8) and southern (9–16) latent-state groups at Point Conception.



Note. Clusters 1–8 (lower-left legend) form the northern group, assigned to latent state x_1 in the state-space model, and clusters 9–16 (upper-right legend) form the southern group, assigned to x_2; the groups are separated at Point Conception (about 34.45°N, 120.47°W). Axes show longitude (°W) and latitude (°N).

Principal component analysis (PCA) on the 16-cluster spatiotemporal matrix showed that the first component explained 61% of total variance and the first two components 94% (Fig. 6), indicating that most joint variability across clusters lies in a 2-dimensional subspace. In ecological terms, most of the coast-wide variation in kelp canopy can be summarized by two dominant modes of variation, which motivates testing a state-space model with one versus two latent regions.

Fig. 6: Principal component analysis of the kelp biomass training set across the 16 DBSCAN clusters.



Note. The left plot shows the percentage of the variance in the 16 cluster time series explained by each principal component (PC). The right plot shows the cumulative variance explained as components are added; dashed lines mark 90% (red) and 95% (green). PC1 explains about 61% of the variance and PC1 and PC2 together about 94%.

Spatial State-Space Model Prediction Results

Based on lowest AIC and consistent convergence across rolling 30-year folds, the optimal multivariate autoregressive state-space model with covariates (MARSSX) is a two-state model with unconstrained state error variance, with clusters 1–8 (north of Point Conception) assigned to latent state x_1, and clusters 9–16 (south of Point Conception) assigned to x_2. The

state process includes ANGLER, HEAT, BEUTI, and seasonal Fourier terms (F1, F2) for time-varying ecological and anthropogenic drivers; the observation equation includes BEACH, capturing regulatory or episodic effects on measured biomass. The process noise variance is higher in the southern state ($Q_{2,2}$ = 0.14) relative to northern state ($Q_{1,1}$ = 0.07), with modest covariance ($Q_{2,1}$ = 0.07).

Among covariates, ANGLER has a small negative coefficient on latent kelp biomass (–0.01), but the sign of this association is model-dependent. We therefore do not interpret it as evidence of a general relationship between fishing activity and kelp canopy. If a negative association were confirmed, it would be compatible with a food-web pathway in which fishing removes urchin predators, or it could reflect coastal human use more broadly. Distinguishing these explanations would require data on urchin and predator abundance and tests with explicitly lagged or spatially resolved fishing effort. Furthermore, HEAT carries a negative coefficient (–0.01), indicating a predictive association between marine heatwaves and biomass declines, a pattern consistent with thermal stress mechanisms documented in the literature[42,43]. In contrast, BEUTI has a strong positive predictive association (+0.06), underscoring the utility of upwelling indices as forecasting inputs. The seasonal components F1 and F2 (–0.13 and –0.56, respectively) highlight recurring but asymmetric biomass peaks and troughs, potentially reflecting regional differences in growth and senescence timing. The BEACH variable shows a weak and near-zero predictive association (–0.00), suggesting that beach closures may have limited utility as an input to kelp biomass forecasting models, perhaps due to spatial or temporal mismatches between regulatory actions and ecological responses. These coefficients are small in absolute terms because covariates are standardized and kelp canopy is log-transformed, so their signs and relative sizes are more informative than their magnitudes.

Both estimated hidden states reveal rich temporal dynamics, exhibiting multi-year oscillations consistent with oceanographic regimes. The northern latent process displays smoother, more stable cycles, with biomass peaking seasonally and increasing gradually: indicative of greater ecological inertia and more synchronized responses to environmental cues. In contrast, the southern state exhibits sharper declines, episodic peaks, and larger amplitude fluctuations, consistent with a more disturbance-prone and fragmented ecological regime. These regime differences describe the combined behavior of all drivers, including those that are not modeled.

Spatial Neural Network Model Results

To complement the MARSSX regime model, we implement an LSTM to forecast kelp canopy area across all 16 clusters simultaneously, conditioned on BEUTI, BEACH, ANGLER, and HEAT, using an 8-quarter lookback window, a 2-layer architecture with hidden dimension 4, dropout 0.2, and early stopping.

On the held-out test set (Extended Data Table 4), the LSTM outperformed MARSSX across all metrics, reducing RMSE from 0.851 to 0.702, MSE from 0.724 to 0.605, MAE from 0.573 to 0.535, and MAPE from 228.7% to 104.8%. These improvements suggest that the nonlinear LSTM is able to capture temporal structure and cross-region dependence beyond what is represented by the linear MARSSX formulation. MAPE values above 100% arise because percentage errors become very large when observed values are close to zero, as when canopy is nearly absent after storms or forest collapse; the large MAPE reduction therefore suggests that the LSTM handles such low-canopy periods better, while RMSE and MAE provide a more stable

basis for comparison. Taken together, the results indicate a trade-off between interpretability and predictive flexibility: MARSSX provides a more interpretable latent-state description of kelp regime structure, whereas LSTM delivers stronger forecasting accuracy.

Discussion

This study employs a suite of time series modeling approaches to investigate the spatiotemporal dynamics of kelp forest biomass along the west coast of North America, integrating environmental and anthropogenic drivers over 41 years of quarterly data. Kelp forests exhibit strong seasonal rhythms and complex autoregressive memory structures, with clear spatial heterogeneity across the biogeographic boundary at Point Conception.

Among the non-spatial models, UARSSX demonstrates best in-sample fit, consistent with a system in which kelp canopy has relatively stable internal dynamics but is measured with considerable observation noise, while SARIMAX's autoregressive structure has the lowest out-of-sample one-step-ahead error. GAM's weaker performance and unresolved residual autocorrelation suggest that it under-represents the temporal dependence structure of kelp biomass relative to the two autoregressive frameworks. However, none of the non-spatial models can represent coordinated dynamics across regions. Explicitly incorporating spatial structure proved important because the clustered regions do not evolve independently but instead show both shared, coast-wide modes of variation and geographically distinct responses to environmental forcing. Ecologically, the strong one-year autoregressive term in SARIMAX is consistent with the annual cycle of canopy growth, loss and regrowth that characterizes these kelp forests.

MARSSX, with its regional latent states, provides ecologically interpretable and decision-relevant insights, integrating state-space modeling with ecological theory and domain-specific clustering. The greater process-noise variance and sharper, episodic fluctuations in the southern latent state, relative to the smoother, inertia-driven northern state, are consistent with the south's higher exposure to episodic disturbances such as marine heatwaves, pollution events, and habitat fragmentation. This latent-state partitioning, grounded in the well-documented Point Conception transition zone[38,39], describes region-specific dynamics and suggests that monitoring designs could differ by region: northern sites, with smoother trends, may be best suited to long-term planning and reserve-network evaluation, whereas the greater volatility of southern sites suggests value in more responsive, event-triggered monitoring. Coast-wide summaries can mask these regional differences.

These regimes can be interpreted ecologically, within the limits of the data. Because the canopy product does not separate species, and giant kelp is expected to dominate the canopy in most of the clusters, the contrast between the northern and southern states should not be read as a contrast between giant and bull kelp; it more plausibly reflects regional differences in oceanography and food webs. The southern group consists of giant kelp forests of the Southern California Bight and Baja California, which lie near the species' warm range limit and experienced the strongest and most spatially variable effects of the 2014–2016 marine heatwave[14]; the greater process variance and episodic declines of the southern state are consistent with exposure to heat, El Niño-related nutrient loss and storm disturbance. Sea otters are absent from southern California, so urchin grazing there is held in check mainly by fished predators such as sheephead and lobster[20], which may help explain why fishing effort (ANGLER) is of interest as a predictor. The northern group lies north of Point Conception, in the cooler, more upwelling-dominated waters of central California[38,39], where sea otters are present; these

differences may contribute to the smoother cycles of the northern state, but they are hypotheses rather than tests, because the models contain no species, grazer or predator data.

The documented collapses since 2013, in which sea star wasting disease removed sunflower stars and warm water favored urchins[22,24], arose from combined disease, predation and temperature effects that no single covariate in our models captures. In our framework such events would be expected to appear as shocks to the latent state or as increased process variance, not as coefficients on HEAT or ANGLER, so the latent-state trajectories should be read as the integrated outcome of all drivers, not of heat alone. The bull kelp forests where the most severe collapse occurred (northern California, Oregon and Washington) lie largely north of the sites that form our clusters, so our results speak mainly to the kelp canopy of central and southern California and northern Baja California.

The MARSSX results carry several management implications. The negative associations of marine heatwaves (HEAT) and positive associations of upwelling (BEUTI) with kelp biomass across models support the practical value of thermal anomaly monitoring and upwelling-index tracking as early-warning inputs. Next, the relationship between fishing activity (ANGLER) and kelp biomass is model-dependent and suggests that any association between fishing activity and kelp is weak, context-specific, or mediated by lagged or indirect pathways not captured here, and it warrants caution when weighing anthropogenic pressures on this foundational habitat-forming species.

Within this decision-support framing, the two model families offer complementary strengths: MARSSX provides spatially structured, interpretable characterization of regime boundaries and covariate associations useful for spatial prioritization and longer-horizon planning, while LSTM's improved forecasting accuracy, reducing RMSE, MSE, MAE, and MAPE relative to MARSSX, indicates that nonlinear temporal dependencies and cross-region interactions add predictive value beyond the linear state-space structure, and could support near-term early-warning applications. Together, these results indicate that incorporating spatial information is essential for robust kelp forecasting, that the choice between MARSSX and LSTM reflects a tradeoff between ecological interpretability and predictive performance rather than a strict superiority of one approach. This complementarity points toward hybrid approaches that combine interpretability with forecast skill as a direction for future work.

Some limitations remain. Sea urchin, sunflower star and sea otter abundances were not included as covariates because consistent, coast-wide quarterly records do not exist; their effects are therefore absorbed into the latent states and process variance, and the coefficients for HEAT, ANGLER and BEUTI should be read as predictive associations that may partly stand in for these unmeasured processes. The sites also do not cover the northern bull kelp-dominated coast, and DBSCAN sets low-density sites aside as noise, so sparse kelp stands may be under-represented. Integrating regional dive-survey data on urchins, predators and kelp with the satellite record, and testing region- or species-specific models where species-resolved canopy maps exist, are natural next steps. Despite these limitations, this study contributes a robust analytical framework for kelp forest dynamics and offers decision-support information for monitoring, early warning and conservation planning for one of the most productive and threatened coastal ecosystems.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Clinical trial number
Not applicable.

Availability of data and material
The kelp canopy biomass data that support the findings of this study are publicly available from the Santa Barbara Coastal Long Term Ecological Research (SBC LTER) program via the Environmental Data Initiative repository (https://doi.org/10.6073/pasta/5b441f39c0876ef901068709df8f3e50). Covariate data (marine heatwave intensity (HEAT), biologically effective upwelling transport index (BEUTI) - proxy for nitrate availability, pacific decadal oscillation (PDO), sea surface temperature (SST), oceanic Niño index (NINO), zooplankton biomass (ZOO), angler trips and recreational fishing effort (ANGLER), total fishery harvest (HARVEST), coastal population (POP), beach closures due to contamination or hazards (BEACH), and billion-dollar disaster events (DISASTER)) were compiled from publicly available federal and academic sources, as described in the Materials & Methods. Processed and imputed covariate panels and clustered kelp biomass time series generated during this study are available from the corresponding author upon reasonable request and will be deposited in a public repository upon acceptance.

Code availability
Code used for data imputation, covariate selection, time series and state-space model fitting (SARIMAX, UARSSX, MARSSX), spatial clustering (DBSCAN), and the LSTM forecasting model is available from the corresponding author upon reasonable request and will be deposited in a public repository upon acceptance.

Acknowledgements
This research received no specific funding. We thank Mark Scheuerell (School of Aquatic and Fishery Sciences, University of Washington), Eli Holmes (NOAA Northwest Fisheries Science Center and School of Aquatic and Fishery Sciences, University of Washington), and Eric Ward (NOAA Northwest Fisheries Science Center and School of Aquatic and Fishery Sciences, University of Washington) for valuable discussions that inspired this work.

Author’s contributions
J.T.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing. C.Z.: Methodology, Formal analysis, Writing – review & editing. S.R.: Conceptualization, Supervision, Writing – review & editing. C.A.: Supervision, Writing – review & editing. All authors reviewed and approved the final manuscript.

Use of Generative AI and AI-assisted Technologies

During the preparation of this work, the authors used OpenAI's GPT-4 to format the manuscript text, references, and structural layouts to meet the journal's submission guidelines. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Competing interests

The authors declare no competing interests.

Funding

This study received no funding.

References

1. Kelp Forest Alliance. *Kelp Forests*. Kelpforestalliance.com (2025). https://kelpforestalliance.com/kelp-forests

2. Eger, A. M. et al. The value of ecosystem services in global marine kelp forests. *Nature Communications* **14**, 1894 (2023). https://doi.org/10.1038/s41467-023-37385-0

3. Schiel, D. R. & Foster, M. S. *The Biology and Ecology of Giant Kelp Forests* (University of California Press, 2015).

4. Hall, D. *Kelp and Kelp Forests* (Smithsonian Ocean, 2024). https://ocean.si.edu/kelp-and-kelp-forests

5. National Marine Sanctuaries. *Impacts on Kelp Forests*. National Ocean Service | NOAA | Department of Commerce (2025). https://sanctuaries.noaa.gov/visit/ecosystems/kelpimpacts.html

6. Suskiewicz, T. S., Byrnes, J. E. K., Steneck, R. S., Russell, R., Wilson, C. J. & Rasher, D. B. Ocean warming undermines the recovery resilience of New England kelp forests following a fishery-induced trophic cascade. *Ecology* **105**(7) (2024). https://doi.org/10.1002/ecy.4334

7. Schlenger, A. J., Beas-Luna, R. & Ambrose, R. F. Forecasting ocean acidification impacts on kelp forest ecosystems. *PLOS ONE* **16**(4), e0236218 (2021). https://doi.org/10.1371/journal.pone.0236218

8. Davis, T. R., Larkin, M. F., Forbes, A., Veenhof, R. J., Scott, A. & Coleman, M. A. Extreme flooding and reduced salinity causes mass mortality of nearshore kelp forests. *Estuarine, Coastal and Shelf Science* **275**, 107960 (2022). https://doi.org/10.1016/j.ecss.2022.107960

9. National Park Service. *Kelp Forests – Oceans, Coasts & Seashores*. NPS.gov (2016). https://www.nps.gov/subjects/oceans/kelp-forests.htm

10. Smale, D. A. Impacts of ocean warming on kelp forest ecosystems. *New Phytologist* **225**(4), 1447–1454 (2019). https://doi.org/10.1111/nph.16107

11. Smith, J. G., Malone, D. & Carr, M. H. Consequences of kelp forest ecosystem shifts and predictors of persistence through multiple stressors. *Proceedings of the Royal Society B: Biological Sciences* **291**(2016), 20232749 (2024). https://doi.org/10.1098/rspb.2023.2749

12. Piñeiro-Corbeira, C., Barrientos, S., Barreiro, R. & de la Cruz-Modino, R. Assessing the importance of kelp forests for small-scale fisheries under a global change scenario. *Frontiers in Marine Science* **9** (2022). https://doi.org/10.3389/fmars.2022.973251

13. Reed, D. C., Rassweiler, A. R., Miller, R. J., Page, H. M. & Holbrook, S. J. The value of a broad temporal and spatial perspective in understanding dynamics of kelp forest ecosystems. *Marine and Freshwater Research* **67**(1), 14–24 (2016). https://doi.org/10.1071/mf14158

14. Cavanaugh, K. C., Reed, D. C., Bell, T. W., Castorani, M. C. N. & Beas-Luna, R. Spatial variability in the resistance and resilience of giant kelp in southern and Baja California to a multiyear heatwave. *Frontiers in Marine Science* **6**, 413 (2019). https://doi.org/10.3389/fmars.2019.00413

15. Springer, Y. P., Hays, C. G., Carr, M. H. & Mackey, M. R. Toward ecosystem-based management of marine macroalgae—the bull kelp, *Nereocystis luetkeana*. *Oceanography*

766 *and Marine Biology: An Annual Review* **48**, 1–42 (2010).
767 https://doi.org/10.1201/EBK1439821169-c1

768 16. Bell, T., Cavanaugh, K. & Siegel, D. SBC LTER: Time series of quarterly NetCDF files of
769 kelp biomass in the canopy from Landsat 5, 7 and 8, since 1984 (ongoing) ver 27.
770 *Environmental Data Initiative* (2025).
771 https://doi.org/10.6073/pasta/5b441f39c0876ef901068709df8f3e50

772 17. Bell, T. W., Allen, J. G., Cavanaugh, K. C. & Siegel, D. A. Three decades of variability in
773 California's giant kelp forests from the Landsat satellites. *Remote Sensing of Environment*
774 **238**, 110811 (2020). https://doi.org/10.1016/j.rse.2018.06.039

775 18. Castorani, M. C. N., Bell, T. W., Walter, J. A., Reuman, D. C., Cavanaugh, K. C. &
776 Sheppard, L. W. Disturbance and nutrients synchronise kelp forests across scales through
777 interacting Moran effects. *Ecology Letters* **25**(8), 1854–1868 (2022).
778 https://doi.org/10.1111/ele.14066

779 19. Estes, J. A. & Palmisano, J. F. Sea otters: their role in structuring nearshore communities.
780 *Science* **185**(4156), 1058–1060 (1974). https://doi.org/10.1126/science.185.4156.1058

781 20. Hamilton, S. L. & Caselle, J. E. Exploitation and recovery of a sea urchin predator has
782 implications for the resilience of southern California kelp forests. *Proceedings of the Royal*
783 *Society B: Biological Sciences* **282**(1799), 20141817 (2015).
784 https://doi.org/10.1098/rspb.2014.1817

785 21. Di Lorenzo, E. & Mantua, N. Multi-year persistence of the 2014/15 North Pacific marine
786 heatwave. *Nature Climate Change* **6**, 1042–1047 (2016).
787 https://doi.org/10.1038/nclimate3082

788 22. Harvell, C. D. et al. Disease epidemic and a marine heat wave are associated with the
789 continental-scale collapse of a pivotal predator (*Pycnopodia helianthoides*). *Science*
790 *Advances* **5**, eaau7042 (2019). https://doi.org/10.1126/sciadv.aau7042

791 23. McPherson, M. L., Finger, D. J. I., Houskeeper, H. F., Bell, T. W., Carr, M. H., Rogers-
792 Bennett, L. & Kudela, R. M. Large-scale shift in the structure of a kelp forest ecosystem co-
793 occurs with an epizootic and marine heatwave. *Communications Biology* **4**(1), 298 (2021).
794 https://doi.org/10.1038/s42003-021-01827-6

795 24. Rogers-Bennett, L. & Catton, C. A. Marine heat wave and multiple stressors tip bull kelp
796 forest to sea urchin barrens. *Scientific Reports* **9**, 15050 (2019).
797 https://doi.org/10.1038/s41598-019-51114-y

798 25. Wear, B., O'Connor, N. E., Schmid, M. & Jackson, M. C. What does the future look like for
799 kelp when facing multiple stressors? *Ecology and Evolution* **13**(6), e10203 (2023).
800 https://doi.org/10.1002/ece3.10203

801 26. Bell, T. W. et al. Kelpwatch: A new visualization and analysis tool to explore kelp canopy
802 dynamics reveals variable response to and recovery from marine heatwaves. *PLOS ONE*
803 **18**(3), e0271477 (2023). https://doi.org/10.1371/journal.pone.0271477

27. West Coast Regional Office. *Kelp Forest Habitat on the West Coast*. NOAA Fisheries (2024). https://www.fisheries.noaa.gov/west-coast/habitat-conservation/kelp-forest-habitat-west-coast

28. Krumhansl, K. A. et al. Global patterns of kelp forest change over the past half-century. *Proceedings of the National Academy of Sciences* **113**(48), 13785–13790 (2016). https://doi.org/10.1073/pnas.1606102113

29. Selgrath, J. C., Carlton, J. T., Pearse, J., Thomas, T. & Micheli, F. Setting deeper baselines: kelp forest dynamics in California over multiple centuries. *Regional Environmental Change* **24**(3), 104 (2024). https://doi.org/10.1007/s10113-024-02260-1

30. Randell, Z., Kenner, M., Tomoleoni, J., Yee, J. & Novak, M. Kelp-forest dynamics controlled by substrate complexity. *Proceedings of the National Academy of Sciences* **119**(8) (2022). https://doi.org/10.1073/pnas.2103483119

31. Pfister, C. A., Berry, H. D. & Mumford, T. The dynamics of kelp forests in the Northeast Pacific Ocean and the relationship with environmental drivers. *Journal of Ecology* **106**(4), 1520–1533 (2018). https://doi.org/10.1111/1365-2745.12908

32. Assis, J., Fragkopoulou, E., Gouvêa, L., Araújo, M. B. & Serrão, E. A. Kelp forest diversity under projected end-of-century climate change. *Diversity and Distributions* **30**(6), e13837 (2024). https://doi.org/10.1111/ddi.13837

33. Wernberg, T. Marine heatwave drives collapse of kelp forests in Western Australia. In *Ecosystem Collapse and Climate Change* (eds Canadell, J. G. & Jackson, R. B.) 325–343 (Springer, 2021). https://doi.org/10.1007/978-3-030-71330-0_12

34. National Park Service. *Kelp Forests – Glacier Bay National Park & Preserve*. NPS.gov (2016). https://www.nps.gov/glba/learn/nature/kelp-forest.htm

35. Hamilton, S. L., Bell, T. W., Watson, J. R., Grorud-Colvert, K. A. & Menge, B. A. Remote sensing: generation of long-term kelp bed data sets for evaluation of impacts of climatic variation. *Ecology* **101**(7), e03031 (2020). https://doi.org/10.1002/ecy.3031

36. Le, D. M., Desmond, M. J., Pritchard, D. W. & Hepburn, C. D. Effect of temperature on sporulation and spore development of giant kelp (*Macrocystis pyrifera*). *PLOS ONE* **17**(12), e0278268 (2022). https://doi.org/10.1371/journal.pone.0278268

37. Kurman, M. D. & terHorst, C. Individual and population-level variation in susceptibility to temperature in early life history stages of giant kelp. *Marine Ecology* **44**(6), e12770 (2023). https://doi.org/10.1111/maec.12770

38. Miller, E. C. Historical biogeography supports Point Conception as the site of turnover between temperate East Pacific ichthyofaunas. *PLOS ONE* **18**(9), e0291776 (2023). https://doi.org/10.1371/journal.pone.0291776

39. Schroeder, I. D., Santora, J. A., Mantua, N., Field, J. C., Wells, B. K., Hazen, E. L., Jacox, M. & Bograd, S. J. Habitat compression indices for monitoring ocean conditions and ecosystem impacts within coastal upwelling systems. *Ecological Indicators* **144**, 109520 (2022). https://doi.org/10.1016/j.ecolind.2022.109520

40. University of California, Santa Barbara, Bren School of Environmental Science & Management. *Assessing Range Shifts of Coastal Species to Inform Conservation in a Key*

California Biogeographic Transition Zone (2025). https://bren.ucsb.edu/projects/assessing-range-shifts-coastal-species-inform-conservation-key-california-biogeographic

41. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735

42. Starko, S. et al. Microclimate predicts kelp forest extinction in the face of direct and indirect marine heatwave effects. *Ecological Applications* **32**(7) (2022). https://doi.org/10.1002/eap.2673

43. Wernberg, T., Coleman, M. A., Bennett, S., Thomsen, M. S., Tuya, F. & Kelaher, B. P. Genetic diversity and kelp forest vulnerability to climatic stress. *Scientific Reports* **8**(1), 1851 (2018). https://doi.org/10.1038/s41598-018-20009-9

Extended Data

Table 1. Covariates Used, with Definitions & Expected Ecological Roles

Acronym	Full name	What it represents	Expected effect on kelp canopy
HEAT*	Marine heatwave intensity	Magnitude of unusually warm sea-surface-temperature events	Negative: heat stress, reduced nutrient supply and greater disease and grazing pressure during warm events[10,21,42]
BEUTI*	Biologically effective upwelling transport index	Amount of nitrate carried to the surface layer by coastal upwelling	Positive: nutrient supply fuels kelp growth[18]
PDO	Pacific Decadal Oscillation	Decades-long pattern of North Pacific Ocean temperature	Indirect: warm phases generally coincide with warmer, lower-nutrient coastal water
SST	Sea surface temperature	Mean temperature of coastal surface waters	Negative above thermal tolerance, especially for early life stages[36,37]
NINO	Oceanic Niño Index	Tropical Pacific temperature anomaly that defines El Niño and La Niña	Negative during strong El Niño: warm, nutrient-poor water and large storm waves[13,14]
ZOO	Zooplankton biomass	Abundance of small drifting animals; an indicator of ocean productivity and food-web state	Indirect: reflects ocean conditions rather than acting directly on kelp
ANGLER*	Angler trips and recreational fishing effort	Level of recreational fishing	Indirect, negative if fishing removes urchin predators and releases urchin grazing; also indexes coastal human use[20]
HARVEST	Total fishery harvest	Quantity of fish and invertebrates landed	Indirect and ambiguous: harvest of predators can release grazing, whereas harvest of grazers (for example, urchins) can reduce it[20,24]
POP	Coastal population	Number of people living in coastal areas	Negative and indirect: runoff, pollution and coastal development[9]

Acronym	Full name	What it represents	Expected effect on kelp canopy
BEACH*	Beach closures	Days of beach closure due to contamination or hazards	Negative and indirect: indicator of runoff and contamination events
DISASTER	Billion-dollar disaster events	Count of weather and climate disasters causing more than US$1 billion in damage	Negative: storms and floods can remove kelp or reduce salinity[8]

Note. Expected effect summarizes the hypothesized direction of influence on kelp canopy based on the literature; it is a rationale for inclusion, not a result of this study. Covariates marked with an asterisk (*) were retained after covariate selection.

Table 2. AIC Scores Across Candidate Models for Interpolation

	ARIMA	GAM	ETS	DLM	Selection
BEACH	1081.72	1144.18	2067.39	1091.05	ARIMA
HARVEST	4447.31	4661.29	5096.88	4753.20	ARIMA
POP	-295.26	46.64	42.86	-161.93	ARIMA
DISASTER	-69.51	50.22	298.14	-64.92	ARIMA
BEUTI	682.50	800.06	1065.34	803.08	ARIMA
SST	25.57	241.46	392.77	32.53	ARIMA
ZOO	-314.68	-133.22	-14.56	-306.11	ARIMA

Note. Akaike information criterion (AIC; lower values indicate a better balance of fit and simplicity) of the four candidate models fitted to each covariate that contained gaps. The Selection column gives the model used to fill the gaps.

Table 3. Univariate Model Metrics Comparison

	In-sample AIC	In-sample RMSE	In-sample R^2	In-sample MAPE (%)	Out-of-sample RMSE	Out-of-sample R^2	Normality	Auto-correlation
SARIMAX	374.17	0.72	0.55	3.40	0.42	0.63	fair	fair
UARSSX	353.23	0.53	0.76	2.60	0.56	0.62	fair	fair
GAM	375.45	0.70	0.59	3.40	0.58	0.50	poor	poor

Note. Performance of the three non-spatial models under 30-year forward-chaining cross-validation: AIC (lower is better), root mean squared error (RMSE; same units as the modeled kelp series, lower is better), squared correlation between predicted and observed values (R^2; higher is better) and mean absolute percentage error (MAPE; lower is better). Normality and Autocorrelation are qualitative ratings (good, fair, poor) of residual diagnostics based on the Shapiro–Wilk and Ljung–Box tests and residual plots (Supplementary Figs. S4–S6).

Table 4. Multivariate Model Metrics Comparison

	RMSE	MSE	MAE	MAPE (%)
MARSSX	0.851	0.724	0.573	228.71
LSTM	0.702	0.605	0.535	104.84

Note. Held-out test-set performance of the multivariate state-space model (MARSSX) and the long short-term memory network (LSTM) across all 16 clusters. RMSE, mean squared error (MSE), mean absolute error (MAE) and

MAPE are all lower-is-better. MAPE is inflated by observations close to zero and should be read alongside RMSE and MAE.

Table 5. Glossary of Technical Terms

Term	Meaning	In plain terms
ACF	Autocorrelation function	Correlation of a series (model residuals) with its own past values
ADF test	Augmented Dickey–Fuller test	Tests whether a series has a stable mean over time (stationary) or drifts
AIC	Akaike information criterion	Model-selection score; lower values indicate a better balance of fit and simplicity
AR	Autoregressive	Current value depends on previous values; AR1–AR5 refer to lags of 1–5 quarters
ARIMA	Autoregressive integrated moving average	Standard time series model built from past values and past errors
DBSCAN	Density-based spatial clustering of applications with noise	Groups nearby sites into clusters and sets isolated sites aside as noise
DLM	Dynamic linear model	State-space model with a slowly changing level
ETS	Error–trend–seasonality	Exponential smoothing model with error, trend and seasonal components
Fourier terms (F1, F2)	Paired sine and cosine curves	Represent the repeating annual cycle in a model
GAM	Generalized additive model	Regression with flexible smooth (non-linear) terms
LASSO	Least absolute shrinkage and selection operator	Penalized regression that shrinks unimportant coefficients to zero
Latent state	Unobserved underlying condition	Regional kelp condition inferred from noisy satellite observations
LISA	Local indicators of spatial association	Identify local hot spots, cold spots and spatial outliers
Ljung–Box test	Test for residual autocorrelation	Small p-values indicate that errors are still correlated through time
LSTM	Long short-term memory network	Recurrent neural network that learns which past information to keep
MAE; MSE; RMSE	Mean absolute error; mean squared error; root mean squared error	Typical size of prediction errors (lower is better)

Term	Meaning	In plain terms
MAPE	Mean absolute percentage error	Average error as a percentage of observed values; inflated when observations are near zero
MARSSX	Multivariate autoregressive state-space model with covariates	Estimates shared latent regional kelp states from 16 cluster series
Moran's I	Spatial autocorrelation statistic	Indicates whether nearby sites have similar values
NN	Feedforward neural network	Flexible machine-learning regression model
PCA	Principal component analysis	Summarizes the main shared patterns of variation among series
Process noise (Q); observation noise (R)	Variance of true change; variance of measurement error	Separate real changes in kelp state from satellite measurement noise
Q–Q plot	Quantile–quantile plot	Graphical check of whether residuals are normally distributed
R^2	Coefficient of determination	Share of variation explained (higher is better)
RF	Random forest	Ensemble of decision trees used for regression
SARIMAX	Seasonal ARIMA with external covariates	Time series regression with seasonal and autoregressive structure
Shapiro–Wilk test	Test for normality	Small p-values indicate that residuals depart from a normal distribution
STL	Seasonal and trend decomposition using Loess	Splits a series into seasonal, trend and remainder components
TM; ETM+; OLI	Thematic Mapper; Enhanced Thematic Mapper Plus; Operational Land Imager	Landsat sensors that supplied the satellite imagery
UARSSX	Univariate autoregressive state-space model with external covariates	State-space model for a single coast-wide kelp series
z-score	Standardized value	Value rescaled to mean 0 and standard deviation 1

882

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementaryinformation.docx
- sir1.docx