Reach-scale models show heterogeneity of stream benthic invertebrate responses to eutrophication stress

David J. Gillett a,⁎, Raphael D. Mazor a, b, Martha Sutula a, Anne Holt a, 1

a Southern California Coastal Water Research Project, 3535 Harbor Blvd, Suite 110, Costa Mesa, CA, USA
b California Department of Fish and Wildlife, Rancho Cordova, CA, USA

ARTICLE INFO

Keywords:
Comparator Sites
Stressor-Response Modelling
Water Quality Criteria

ABSTRACT

Statistical stressor-response models are a common approach to derive biologically relevant water quality criteria for the management of waterbody health. These types of models are typically derived at state-wide or ecoregional scales and consequently incorporate a large amount of natural and disturbance-related variability that may obscure the relationship that one is interested in quantifying. We demonstrate an alternative approach termed “reach-specific modeling” to identify potential biological response thresholds to eutrophication in the Santa Margarita River watershed (California, U.S.A.). Individual models of benthic invertebrate response to eutrophication stress were created for both bioassessment sampling sites and NHD+ stream-segments in the watershed (46 sites, 832 segments). Each model was built using only data from environmentally similar sites from a state-wide dataset to minimize variation from natural environmental gradients, while allowing eutrophication stress to vary. Thresholds of potential biological impact were extracted from each stressor-response model. Across the whole watershed thresholds varied from location-to-location: total nitrogen (1.14–1.26 mg L−1 TN), total phosphorus (0.12–0.15 mg L−1 TP), benthic algal biomass (29–39 mg m−2 benthic chl-a), and benthic ash-free dry mass (2.5–3.0 mg cm−2 AFDM). Notably, nearly all of the thresholds derived from these reach-specific models were ~10–90 % higher than those from a similar state-wide model. Furthermore, there were a number of spatial groupings of thresholds for each eutrophication indicator across the watershed, suggesting reach-scale natural gradients in hydrogeomorphology and natural land cover type may mediate the stressor-biology interaction. Reach-scale models tended to have better fits than their state-wide counterparts, but had equivalent or slightly worse accuracy. The reach-specific approach to threshold development illustrates that the biological response to stress is likely not uniform within a single system, much less between systems. As a consequence, this approach can allow managers to identify systems that are more sensitive or resistant to a given stressor across diverse landscapes and make better informed decisions on their management accordingly.

1. Introduction

Environmental management frameworks centered around the protection of aquatic life from the effects of human disturbance rely on assessing the health and integrity of the biota, as well as their response to different environmental stressors. From these relationships, thresholds of stressor exposure that correspond to critical changes in biological response can be extracted and used as the basis for establishing water quality criteria (WQC) protective of aquatic life (Yuan et al., 2010). Statistical modeling of relationships between biological condition and environmental stressors is a common approach to derive WQC (Heiskary and Bouchard, 2015, Jessup et al., 2015). Most typical are statistical models that relate spatially/temporally synoptic field measures of biological communities (benthic invertebrates, fish, algae, etc) to indicators of environmental stress measured at landscape or in-stream-scales (Lamon and Qian, 2008; Mazor et al., 2022; Schneider et al., 2020).

The studies that inform WQC development are often conducted at the ecoregional or state-wide scale. For the management and regulatory community, this single model approach is often seen as straightforward and equitable, as it allows for a uniform application of a model or threshold across a state or a region. However, the data sets used for this type of modelling will often contain a large amount of additional

⁎ Corresponding author.
E-mail address: davidg@sccwrp.org (D.J. Gillett).
1 Present address: San Diego State University, San Diego, California.
variability from both natural gradients and non-target disturbances beyond the stressor one is interested in developing criteria for (e.g., pesticides, nutrient concentrations, hydromodification). Consequently, models developed at the scale of broad typologies like ecoregions may be insufficiently sensitive to account for small-scale variability in stressor-response relationships in environmentally diverse regions. In turn, the WQC derived from such studies may be either over stringent or insufficiently protective at a given location in a region. Site-specific numerical models (e.g., Wool et al., 2020; Zehra et al., 2023) have been touted as an alternative approach to derive watershed specific WQC, but data requirements and resources required for mechanistic numerical modeling can be cost-prohibitive to generate and compute for many systems.

To allow for a greater amount of flexibility than traditional state-wide or ecoregional approaches may provide, we have developed an alternative we have termed “reach-specific modeling.” With this reach-specific approach, a unique relationship between stress and response is modeled for discrete sections of a watershed. Small-scale natural variability between different sections is accounted for by developing different models using only data selected from other streams that have a similar environmental setting to the reach of interest. This can be contrasted to the less curated data sets – typically all of the available data in a state, region, or other geographic domain – used in a broader ecoregional approach (e.g., Kim et al., 2021).

We define environmentally similar stream-reaches as locations that would be expected to support similar biotic assemblages to the reach of interest in the absence of disturbance, thereby creating a curated data set with which one can track changes in biological composition under exposure to stress (Gillett et al., 2019). This selection approach creates a constrained set of stream data for the subsequent analyses, where naturally occurring factors (e.g., elevation, rainfall, reproductive populations) are kept relatively constant, while potential causal factors associated with human activity are allowed to vary. This analytical approach is an outgrowth of site-specific stressor identification (i.e., causal assessment) (Gillett et al., 2023, 2019; Norton et al., 2015) and is analogous to control cases in epidemiology (e.g., Rothman et al., 2008). Applying this approach across a watershed has the potential to identify different stressor-response relationships – and therefore response thresholds – for individual sites or segments, which could be of use for reach-scale management actions in the waterbody.

To illustrate the practical implications of reach-specific modelling and threshold development, we have applied the approach to the eutrophication problems impacting a large, ecologically diverse watershed in southern California (USA). The goals of the study were to: 1) develop models of eutrophication stress on benthic invertebrate assemblages for multiple discrete bioassessment sites, as well as National Hydrography Data set stream segments within the watershed, 2) determine the similarity or differences in biotic response to eutrophication across the watershed; 3) compare thresholds protective of biotic integrity derived from the reach-specific models to those derived from state-wide models that used a similar statistical modeling approach.

2. Methods

2.1. Analytical approach

Reach-specific responses to eutrophication stress were modeled with two different schemes within the Santa Margarita River watershed, U.S.A (Fig. 1): 1) At previously sampled bioassessment monitoring sites; and 2) At every National Hydrography Data Plus (NHD + ) stream segment. As depicted in Fig. 2, the process began when environmentally similar stream sites were identified from California’s bioassessment database for each distinct reach (i.e., site or segment) in the watershed. The intent of this process is to create a curated data set for each reach within the watershed that minimizes the “noise” (e.g., community variation due to natural gradients among different locations) without dampening the “signal” (e.g., community change due to stressor gradients) one is interested in understanding. Benthic invertebrate and eutrophication data from these environmentally similar sites were then used to create reach-specific logistic regression models that predicted the probability of the reach supporting reference condition biota across a gradient of eutrophication stress. From the logistic models, the level of eutrophication stress with a 90% probability of supporting reference condition biota was then set as a threshold of potential impact for a given reach. Once thresholds were identified, the relative heterogeneity of stream invertebrate-eutrophication relationships across the watershed was compared among all sites or segments within the watershed, as well as to thresholds developed from a California-wide data set.

---

**Fig. 1.** An illustration of the Santa Margarita River watershed in south coast of California, U.S.A. Streamlines are based upon NHD + information and are colored to delineate the upper (dark green), middle (blue), and lower (eggplant) sections of the watershed, in following with how the watershed is typically delineated by local agencies. The points depict the locations of the previously sampled bioassessment sites that fall within the watershed. The color of the points indicate the their observed condition based upon benthic macroinvertebrate community composition (SCSI, Mazor et al., 2016). The inset shows the approximate location of the watershed along the Pacific coast of California. Base map obtained from [www.maps.stamen.com](http://www.maps.stamen.com) with the ggmap package (v3.0.0) in R (Kahle and Wickham, 2013).
and Fallbrook creeks, is located along the southern Pacific coast of California, U.S.A. (Delong et al., 2023). It is represented by 832 unique sites, represented by the dashed circles and solid arrows. These sub-data sets are then used to create stressor-response models, represented by the logistic curve, from which a threshold of biological response specific to that reach can be extracted. For the state-wide approach, all data within the state-wide data set are selected. These data are then used to create a model from which a threshold applicable to all reaches across the state can be extracted.

2.2. Study area and data set

The Santa Margarita River and its tributary streams, including Murrieta, Temecula, Pechanga, Devils, Stone, Rainbow, Sandia, De Luz, and Fallbrook creeks, is located along the southern Pacific coast of California, U.S.A. (Delong et al., 2023). It is represented by 832 unique stream segments in the medium-resolution NHD data set that comprise 1,790 km of stream-length. The watershed drains nearly 1,927 square kilometers of urban, agricultural, and undeveloped uplands. The region has a Mediterranean climate with a cool, winter wet season (November–April), dry hot summers, and an average annual rainfall of 241 mm (Yoon and Stein, 2008). The Santa Margarita River is one of the few remaining free flowing, undammed rivers in the region and the lower 43 km of the river to its estuarine headwaters are not channelized. Like many wadeable streams in heterogenous landscapes like California, the whole of the Santa Margarita River watershed contains a wide diversity of topographic elevation, climate, hydrogeomorphology and biotic communities (Ode et al., 2016b). The Santa Margarita River and its tributaries traverse strong natural gradients, from its headwater streams that drain off the slopes of the San Jacinto and Palomar mountains, the Santa Rosa Plateau, and the alluvial fan of Temecula Valley, through the Elsinore fault zone and into the Santa Margarita Estuary.

Human activities across the watershed have resulted in altered flows and increased nutrient loading to the tributaries of the Santa Margarita River. These alterations have resulted in elevated nitrogen and phosphorus concentrations, chronic algal blooms and, in certain parts of the main stems, low dissolved oxygen (Sutula et al., In Review). These eutrophication problems have triggered a 303(d)-impairment listing for nutrient pollution and stakeholders are pursuing a process to identify total maximum daily loads, for which in-stream eutrophication thresholds are needed. The watershed is also part of regional and state-wide bioassessment monitoring programs that use a combination of random ambient and targeted site sampling (e.g., Mazor, 2015, Rehn, 2021), with 46 discrete bioassessment monitoring sites that have been sampled in the watershed (Fig. 1).

As part of the monitoring programs, synoptically collected benthic macroinvertebrates, physical habitat characteristics, and eutrophication measures have been collected and analyzed from 46 locations within the Santa Margarita River watershed, as well as 6,200 others across the state. The implementation program for stream bioassessment in California is supported by well-established protocols, training and quality assurance, and a broad network of minimally disturbed reference sites (Ode et al., 2016a,b). As a result, a robust state-wide bioassessment data set exists; comprised of benthic macroinvertebrates and algal assemblage information, organic matter distribution, and a comprehensive set of eutrophication drivers (e.g., nutrients, flow, and water temperature) available for analysis.

Benthic macroinvertebrates, habitat, and eutrophication data collected between 2001 and 2019 were obtained from the from the Stormwater Monitoring Coalition (SMC)/Surface Water Ambient Monitoring Program (SWAMP) data portal (https://smc.sccwrp.org/). Benthic macroinvertebrates were collected across 150-m stream reaches using a 500-µm D-frame kicknet following Ode et al., (2016a). All fauna were preserved in ethanol and identified to Southwest Association of Freshwater Invertebrate Taxonomists (SAFIT) Standard Taxonomic Effort Level 2 (i.e., mostly to species, Chironomidae to genus) (Richards and Rogers, 2011). The condition of each sampling location was based upon macroinvertebrate assemblage composition scored with the California Stream Condition Index (CSCI) (Mazor et al., 2016). Following Mazor et al. (2022) benthic algal biomass as chl-a (mg m⁻²), benthic organic matter accumulation measured as ash-free dry mass (mg cm⁻²), macroalgal percent cover (%), water column total nitrogen (TN) and total phosphorus (TP) concentrations (mg L⁻¹) were selected as direct or indirect measures of eutrophication stress. These data were obtained from the SMC and SWAMP data portal (https://smc.sccwrp.org/). All eutrophication measures were collected following standardized field and lab protocols (Ode et al., 2016a).

2.3. Identification of ecologically similar sites

The ecological similarity approach is an extended application of Observed-to-Expected (O:E) bioassessment indices. It approximates the ability of any two stream reaches to support similar macroinvertebrate communities in the absence of anthropogenic disturbance using taxa profiles and capture probabilities predicted from underlying natural gradients modelled from reference sites (Gillett et al., 2019). Specifically, the O:E index predicts reference community composition using latitude, elevation, 10-year (2000–2009) mean annual precipitation, 10-year (2000–2009) mean annual air temperature, and watershed area for bioassessment sites following Mazor et al. (2016) or for NHD + stream segments following Beck et al. (2019). Expected ecological similarity allows us to infer the biological composition, as well as the expected response to stress, in each NHD + segment, irrespective if it had been sampled or not (Gillett et al., 2019). Expected biological similarity was calculated as pair-wise values based upon the Bray-Curtis dissimilarity formulation (Bray and Curtis, 1957) using the expected taxa capture probabilities from a Santa Margarita River site or segment and the expected taxa capture probabilities from other sites in the state-wide bioassessment data set (Gillett et al., 2019). Sites were considered similar to a given Santa Margarita River
2.4. Estimating eutrophication thresholds

The reach-specific data sets generated from the ecological similarity filtering process were subsequently used to generate biological response models for each distinct reach (i.e., bioassessment site or NHD + segment) in the watershed. For each reach, a logistic regression model was created from the eutrophication and biological data collected at their ecologically similar sites. If an ecologically similar site had been sampled multiple times, the sampling event with the highest CSCI score was selected for use. The logistic regression was structured with the probability of observing a CSCI score \( > 0.79 \) (i.e., 10th percentile of index scores at reference sites) as the response variable and one of the five eutrophication stress measures as the predictor variable (i.e., 5 regressions per site/reach).

The regression models were constructed to only consider the effects of one eutrophication indicator at a time. In reality however, the biota in many of the streams from which the data used to build those models were obtained were exposed to a variety of other stressors beyond the eutrophication stress. To allow us to compare results among logistic models whose underlying data represent varying degrees and diversities of non-eutrophication stressors in addition to the eutrophication, the probabilities at each stressor increment were re-scaled by dividing them by the maximum probability observed at the y-intercept for a given model. This approach, described in Mazor et al. (2018) and Mazor et al. (2022), effectively anchored the relative probability of observing reference condition biology under minimal eutrophication stress at 1 and creates relative probabilities versus absolute probabilities from the unadjusted regressions. The relative logistic regression approach, though a simplistic approach for partially separating eutrophication impacts from multiple stressors occurring in the field data, allowed us to standardize the effects of eutrophication stress across all the different models. This standardization allowed for clearer communication of potential threshold values with a management-focused audience.

The stressor value that produced a relative probability of 0.9 was selected as the threshold for maintaining reference condition benthic macroinvertebrates at the given Santa Margarita River reach. It is worth noting that though we used logistic regression models in this study, other types of response models (e.g., general additive models, linear regression, random forest regression) could be used with this reach-specific approach. Each of the five eutrophication measures were expected, \textit{a priori}, to have an inverse relationship with biotic condition (e.g., Dodds, 2006, Hilton et al., 2006, Johnson and Hering, 2009). As an illustration: increasing TN should reduce the probability of observing a CSCI score \( > 0.79 \). Consequently, if the coefficient for the eutrophication measure in the regression model was greater than 0, the regression was rejected as violating our conceptual model and the threshold was not retained. Similarly, if the \( p \)-value associated with the regression coefficient was \( > 0.1 \), the regression was rejected as a non-useful model and the threshold was not retained. All regressions were modeled using the glm function with a logit link in R v3.6.3. Likelihood ratio pseudo \( R^2 \) values (Nagelkerke, 1991) to approximate the goodness of fit for each model were calculated using the rsq function in the R package rsq v2.6 (Zhang, 2023). Receiver operating curve (ROC) characteristics for each model to characterize model performance were calculated using the roc function in the R package pROC v1.18 (Robin et al., 2011). The heterogeneity of the threshold values among the bioassessment sites within the watershed was approximated by calculating the coefficient of variation (CV) within each eutrophication measure. CV values were calculated using the cv function in the R package EnvStats (Millard, 2013). Heterogeneity of the segment-scale thresholds was evaluated by comparing values from the upper, middle, and lower sections of the watershed using a Kruskal-Wallis rank sum test with Dunn post-hoc comparisons (\( \alpha = 0.1 \)). Following local practice, the NHD + segments of watershed was divided into the three sections: 1) Upper - upstream of the Temecula Gorge, including the San Bernardino Mountain headwaters; 2) Middle – from Temecula Gorge and the coastal mountains down to the confluence with De Luz Creek; and 3) downstream of the confluence with De Luz Creek, which demarcates the beginning of the alluvial sediments (Fig. 1). Values for the reach-scale thresholds were compared between these upper, middle, and lower sections of the watershed using a Kruskal-Wallis rank sum test with Dunn post-hoc comparisons (\( \alpha = 0.1 \)). The Kruskal-Wallis tests were calculated with R v3.6.3 and the Dunn tests were conducted with the dunnTest function in the R package FSA v.0.8.32 (Ogle et al., 2021).

3. Results

Based upon expected biological similarity, between 482 and 1,138 ecologically similar sites were identified for each selected bioassessment site within the Santa Margarita River watershed. Of those groups of ecologically comparable sites, each group had at least 113 sites with an expected biological similarity value < 0.05 to their paired bioassessment site. For each of the Santa Margarita River bioassessment sites, nearly all of their ecologically similar sites were located within the coastal and inland chapparal regions of southern California, with a few sites located in central or northern coastal California (Fig. 3).

Across the NHD + stream segments, between 479 and 1,218 ecologically similar sites were identified for each segment, based upon their expected biological similarity values. The ecologically similar sites for the stream segment followed a similar geographic distribution to those of the bioassessment sites, with most of the similar sites found in the coastal chapparal of southern California and a handful of sites in central California, as well as northern/central parts of the coast.

Using the stressor and biological data from the ecologically similar sites in the logistic regression models, thresholds of critical biological response to each of the eutrophication measures were identified for nearly all of the bioassessment sites in the watershed (Fig. 4). The thresholds were not uniform across the different sites, though the majority of the site thresholds were clustered tightly together (if the extreme values are ignored). Quantitatively, there was a considerable degree of heterogeneity in thresholds for algal cover (CV = 0.55) and TP (CV = 0.94) across the different bioassessment sites (Fig. 4). The thresholds for AFDM (CV = 0.14) and Chl-a (CV = 0.08) showed less heterogeneity, while TN (CV = 0.3) showed moderate variability. Visually, one can see that the thresholds for AFDM and benthic Chl a fell out into three relatively distinct groups of values (Fig. 4). The thresholds for TP had two groups, with an additional high and a low value apart from the others. Similarly, the thresholds for TN were clustered together, with the addition of a high and low value apart from the central group. In contrast to the other eutrophication metrics, the % algal cover thresholds fell out in a continuous gradient, without any distinct grouping.

Thresholds for % algal cover, TN, and TP could not be identified for one bioassessment site as each regression coefficient had a \( p \)-value > 0.1 and therefore their models were rejected. Similarly, a threshold for % algal cover could also not be identified for another, nearby bioassessment site. Both of these sites were located in the upper portion of the Santa Margarita watershed. Both sites were within a relatively high elevation, spring-fed stream on the north side of Palomar Mountain (R. Mazor, pers. observation), which represents a relatively unique setting compared to other bioassessment sites in the watershed.

The segment-specific analyses provided a greater granularity in stressor response patterns than the site-specific approach. A visual inspection of the thresholds identified for each reach of the Santa Margarita River watershed indicates that the modeled sensitivity of the resident biota to eutrophication stressors was clearly not uniform across the watershed (Fig. 5). The lower reaches of the watershed (costal, low
gradient) had statistically lower thresholds for AFDM, algal cover, TN, TP than the upper (mountainous, high gradient) reaches of the watershed (Fig. 5, Table 1). Middle reaches had thresholds typically between the other two. A clear pattern in the magnitude of the thresholds for each eutrophication measure was also apparent geographically (Fig. 7). Thresholds tended to clump together, with the eastern reaches consistently being different than the other parts of the western and southern parts of the watershed. Of the 832 stream segments considered, thresholds for TN could not be identified for 26 segments (3% of total), as well as 160 (19%) for % algal cover and 15 (2%) for TP. As with the bioassessment sites, all of these instances were rejected. Thresholds for Chl a and AFDM could be calculated for all 832 stream segments.

Interestingly, many of the out-lying thresholds (i.e., beyond the 25th or 75th percentiles) for both the bioassessment sites and stream segments were from locations that had comparatively fewer numbers of ecologically similar sites with eutrophication data in their regression models (Figs. 4 & 5). However, those models had >200 data points, which would suggest that they had sufficient sample size. The patterns in standard error of the different threshold estimates for the bioassessment sites (Fig. 6) indicate that the models with lower sample sizes produced thresholds with greater amounts of associated variance than those with greater data density for AFDM, TN, and TP. For the lower data density AFDM models, the variance was ~2X the other models. The variance was much greater for the low data density TN (~3.5X) and TP (~15 X). This may suggest that the differences in threshold and variance around those estimates may be a function of some site-specific anomaly of these

Fig. 3. An example illustration of the ecologically similar sites used to model the stressor response relationships at one of the bioassessment sites (902SMRDRx) within the Santa Margarita River watershed. The yellow triangle illustrates the approximate location of the site and the dots represent other bioassessment sites within California. The color of the dots indicates their relative similarity to the Santa Margarita River site. Similarity was measured as expected biological similarity calculated from modeled taxa capture probabilities (see Gillett et al., 2019 for full details). Only sites classified with similarity ≤ 0.1 were used in the subsequent stressor-response modelling.
6

locations rather than the influence of 220 data points versus 300 + data points. Irrespective of the threshold values, the comparatively lower number of ecologically similar sites indicates that those locations were comparatively distinct in both their underlying ecology among the other parts of the Santa Margarita River watershed and the state as a whole. The high variance/lower data density models were mostly associated with higher elevation, mountainous sites. When these factors are combined with the relatively warmer temperatures and lower rainfall of southern California, these types of sites are rare within the California landscape (see Mazor et al., 2016 for detailed discussion) and would be expected to have relatively fewer environmentally similar sites.

Nearly every threshold identified for Santa Margarita River bioassessment sites or stream segments was higher than the thresholds derived from similar state-wide statistical modeling (Mazor et al., 2022). The exceptions were thresholds for algal cover, which were always below the state-wide number in the site-specific analyses. Within the segment-specific analyses, the thresholds of algal cover from the lower and middle portions of the Santa Margarita River watershed were all or nearly all below the state-wide threshold. In contrast, approximately 66 % of the reaches (n = 420) from the upper portion of the watershed had algal cover thresholds less than the state-wide model threshold. The reach-scale models yielded better fits of their reach-scale data sets (i.e., greater likelihood pseudo R² values) than their statewide counterparts did with the statewide data set (Fig. 8). Conversely, the reach-scale and statewide models had relatively similar classification accuracy (i.e., AUC values) to each other for all of the eutrophication models. The AFDM models were the lone exception, as the state-wide model performed better than all of the reach-scale models (Fig. 8).

4. Discussion

Our modelling of the biotic condition of streams in relation to eutrophication stressors indicated that even in moderately sized (1,927 km²) systems like the Santa Margarita River, it may be common to observe different responses to disturbance across a watershed. The nature of the response was similar in the majority (95.6 %) of reaches across the watershed – increases in the eutrophication measures decreased the likelihood of observing a reference-condition macrobenthic community. However, the magnitude of stressor exposure likely to degrade the biotic communities beyond a critical point varied in both the site- and segment-scale analyses. The degree of heterogeneity among the critical points did differ from indicator to indicator – total nitrogen thresholds were quantitatively similar to each other but benthic ash free dry mass and chlorophyll a were more varied. To our knowledge, this is the first study to characterize the heterogeneity of stressor-responses to eutrophication at these type of fine spatial scales within a single watershed.

Differences in the sensitivity to stress among distinct waterbodies has been observed in a variety of previous studies conducted at inter-watershed, ecoregional, or state-wide scales (e.g., Lamon and Qian, 2008, Heiskary and Bouchard, 2015). Many of these modeling studies were designed to account for natural gradients in their data sets. However, the results were still likely influenced by large scale variations in climate, topography, and geology that would obscure small scale heterogeneity. The differences between the eutrophication thresholds derived from reach-specific and state-wide response models (Mazor et al., 2022) that we observed were quantitatively different and of a scale that would be meaningful to regulated parties potentially responsible for achieving a given threshold in their streams. Moreover, there was concordance in the thresholds, where nearly every reach-specific threshold of four of the five eutrophication measures across the Santa Margarita were higher than those thresholds derived from the state-wide analysis. This type of pattern suggests that the Santa Margarita River as a whole may have a different, dampened response to eutrophication stress than other wadeable streams across the state. Conversely, it is logical to assume that there are other watersheds within

Fig. 4. Schematic box plots of site-specific thresholds derived from logistic regression models for five eutrophication measures across the 46 bioassessment sites of the Santa Margarita River watershed. The shade of the dots represents the number of data points used in the respective logistic regression. The dashed line indicates the state-wide model threshold for that measure from Mazor et al. (2022). Please note that for easier visual interpretation of the box plot, one total nitrogen threshold value (3.0 mg L⁻¹) and one total phosphorus threshold value (0.88 mg L⁻¹) were omitted from their respective panels.
California that are more sensitive to eutrophication stress than the threshold suggested by the single state-wide analysis. Beyond being a watershed with a different scale of response to eutrophication than others in its region, the Santa Margarita River also serves as an illustration of a watershed with intra-watershed heterogeneity in stressor response patterns. Three or four distinct portions of the watershed were apparent, each with different sensitivities to eutrophication. The spatial patterns in the eutrophic measures’ thresholds from the segment-scale analyses were indicative of some degree of a sub-catchment level of organization (Fig. 7). Contiguous reaches tended to have similar thresholds, which suggests that the abiotic, environmental setting across different parts of the watershed may have been mediating that stressor response relationship. The threshold patterns for TN and TP behaved the most similarly, with greater sensitivity in the middle and lower portions of the watershed and greater tolerance in the upper/eastern part of the watershed and some of the upper reaches of lower tributaries. The distribution of thresholds for AFDM and algal cover were similar to that of TN and TP, though not as bimodally distributed. The distribution of thresholds for benthic Chl a were the inverse of the other measures, most sensitive in the upper/eastern portions and higher thresholds in the central and lower parts of the watershed.

This kind of spatial organization of thresholds makes some sense, as the natural environmental setting of a waterbody can influence stressor response relationships at relatively small spatial scales irrespective of anthropogenic disturbance (e.g., Graeber et al., 2017, Kim et al., 2021, Turunen et al., 2021), especially with a complex stressor like eutrophication. Watershed characteristics (e.g., rainfall, geology, topography) influences the in-stream and riparian habitat, which can alter light availability and water temperature (Ferreol et al., 2005; Vannote et al., 1980). Furthermore, small-scale changes in slope will influence stream velocity and water residence time. These abiotic factors specifically influence primary production and accumulation of excessive organic detritus (Table 1).

### Table 1

Summary output of Kruskal-Wallis rank sum tests comparing reach-specific thresholds for each eutrophication measure among the three sections of the Santa Margarita River watershed. An underline in the post-hoc comparisons indicates that there were no differences in threshold values between sections of the watershed based upon a Dunn post-hoc test with $\alpha = 0.1$.  

<table>
<thead>
<tr>
<th>Eutrophication Measure</th>
<th>Number of reaches</th>
<th>df</th>
<th>Chi-sq statistic</th>
<th>p-value</th>
<th>Watershed Post-Hoc Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFDM (mg cm$^{-2}$)</td>
<td>829</td>
<td>2</td>
<td>232.3</td>
<td>&lt;0.001</td>
<td>Lower Middle, Upper</td>
</tr>
<tr>
<td>Chl a (mg m$^{-2}$)</td>
<td>829</td>
<td>2</td>
<td>6.8</td>
<td>0.034</td>
<td>Middle, Upper</td>
</tr>
<tr>
<td>Total Nitrogen (mg L$^{-1}$)</td>
<td>803</td>
<td>2</td>
<td>47.3</td>
<td>&lt;0.001</td>
<td>Lower, Middle, Upper</td>
</tr>
<tr>
<td>Total Phosphorus (mg L$^{-1}$)</td>
<td>814</td>
<td>2</td>
<td>18.1</td>
<td>&lt;0.001</td>
<td>Lower, Middle, Upper</td>
</tr>
<tr>
<td>Algal Cover (%)</td>
<td>669</td>
<td>2</td>
<td>78.9</td>
<td>&lt;0.001</td>
<td>Lower, Middle, Upper</td>
</tr>
</tbody>
</table>
matter (i.e., eutrophication) in a system (Nijboer and Verdonschot, 2004). Though we lack the data within this system to test these kinds of mechanistic relationships, the pattern we observed in the thresholds is suggestive of how reach-scale habitat may attenuate some types of anthropogenic pressures and stressors on the resident biota (Birk et al., 2020; Turunen et al., 2019, 2018). However, it is also

Fig. 6. Scatter plots of the site-specific thresholds derived from logistic regression models for five eutrophication measures across the 46 bioassessment sites of the Santa Margarita River watershed as a function of the number of data points used in their respective models. The error bars represent the +/- standard error of each threshold. The threshold values represent the magnitude of the eutrophication metric with 0.9 probability of supporting a reference condition (California Stream Condition Index score >0.79) as also presented in Fig. 4.

Fig. 7. Watershed maps depicting the segment-specific thresholds derived from logistic regression models of AFDM (a), Algal Cover (b), benthic Chlorophyll a (c), water column total nitrogen (d), and water column total phosphorus (e) across the 829 NHD + stream reaches of the Santa Margarita River watershed. Where thresholds could not be derived, the color was averaged between the next most proximate segments. Base maps obtained from www.maps.stamen.com with the ggmap package (v3.0.0) in R (Kahle and Wickham, 2013).
important to consider that reach-scale habitat may not influence the stressor-response relationships for other classes of stress like elevated conductivity or pesticides that have a more direct effect on in-stream benthic macroinvertebrates.

In addition to influencing how a stressor like eutrophication is manifest in a stream, the habitat of stream reaches will also influence the population dynamics and recruitment of the benthic fauna that live there in the absence of disturbance (a key tenant in our modelling approach) (Gasith and Resh, 1999; Herlihy et al., 2008; Johnson and Hering, 2009). Different taxa, “selected” by the filter of a stream’s environmental setting, will respond to a given stressor on a taxa-by-taxa basis, which combine together to create the community responses to stress that were observed between the Santa Margarita River and other systems used in the state-wide modelling efforts. Mechanistically disentangling the two ways in which natural environmental gradients concurrently affect both the stressor exposure and the biological response is not an easy proposition and beyond the scope of this work.

Overall, the Santa Margarita River watershed serves as an interesting case study for the development of reach-specific stressor response models and water quality criteria in general. There was intra-watershed variability in the responses of the stream biota to eutrophication, which may have been due to small-scale differences in the environmental settings (e.g., slope, riparian zone, channel morphology). Though they were not homogeneous across the watershed, nearly all of the different reaches had eutrophication thresholds higher than those derived using all of the wadeable streams in California. These Santa Margarita versus state-wide differences were likely the product of variability in environmental settings, as well as reach-to-reach differences in the expected fauna. What is unclear, is how this pattern of intra- and inter-watershed differences in response to stress would play out in other systems – both within an environmentally heterogeneous place like California (Ode et al., 2016b) or in a homogeneous location elsewhere in the United States or the world.

Our modelling approach to determine reach-specific stressor-response thresholds was unique because it pre-screens data to only include biotic and abiotic data from stream sites that were ecologically similar to the bioassessment site or stream segment of interest in lieu of simply using every available data point blindly. In post-hoc modeling or analysis of field collected data, there will always be the tradeoffs of maximizing sample size versus isolating the variance of interesting variables versus controlling for the heterogeneity of background variance that is not of interest to the question at hand. Our approach is designed to control the environmental and biogeographic differences/similarities among different sites a priori, which then allows the models to reflect the biological response more accurately to the stressor of interest. Accordingly, the reach-scale models did typically produce better fits between stream biota and eutrophication data than their state-wide analogs. However, though they dealt with variance in the data better, the reach-scale models were not more accurate in their predictions than the state-wide models. The prediction accuracies were reasonable, though not perfect (i.e., AUC values between 0.6 and 0.8). This pattern was likely reflective of multiple stressors affecting the condition of the stream fauna concurrently with the eutrophication measures that were modelled at reach and state-wide scales.

An alternative to building the reach-scale models could be to build more complex stressor response models with multiple predictors (e.g., the same predictors used to model expected taxa in the OE models used for our expected similarity). However, that inherent complexity could make it more difficult to simultaneously run the 800 + models that were needed to completely cover a medium sized watershed like the Santa Margarita. Furthermore, there is inherent value in the parsimony of models from a statistical perspective (e.g., Burnham and Anderson, 2002) and for the communication of complex ecological ideas to less technical audiences like water quality managers.

Reach-specific thresholds can improve watershed management because they can build stakeholder confidence that the unique environmental conditions in their areas are properly accounted for and they can highlight specific places where ecoregional or typological thresholds are inappropriate. The California State Water Resources Control Board is proposing to adopt water quality criteria protective for wadeable streams against nutrient pollution and eutrophication based on state-wide or regional statistical stress response modeling (Mazor et al.,

![Diagram](image-url)
The Santa Margarita River represented a situation where stakeholders were concerned that eutrophication targets derived from state-wide modeling might be inappropriate for their watershed. Both regulated and regulatory parties felt that biological response models that accounted for the watershed’s diversity and natural gradients were necessary. Our study showed that the regulators and regulated parties were not wrong and that state-wide model thresholds may indeed be overly stringent for most parts of the Santa Margarita River. However, the practical aspects of enforcing reach-specific thresholds may be too challenging for implementation. It may be that something more granular than a single, state-wide threshold, but more integrative than reach-by-reach thresholds (e.g., watershed or sub-watershed) could represent a reasonable compromise.

We believe reach-specific models can still be useful, even where their implementation may be impractical. For example, reach-specific models can be used to evaluate the appropriateness of a priori, homogeneous typologies, or they can provide a way to identify previously unrecognized sub-regions of a watershed requiring different management strategies. It is important to highlight that though we have only presented results from Southern California, the approach of identifying environmentally similar sites and building reach-specific models can be applied in other regions that have a sufficiently large bioassessment data set and an O/E predictive model. Within the United States, the National Rivers and Streams Assessment (NRSA) program provides both of these elements (US Environmental Protection Agency, 2020) and could support this type of reach-specific modelling approach.

5. Conclusions

We have presented an approach to create reach-specific eutrophication stressor response models and thresholds derived from ecologically similar streams. The application of the reach-specific relationships revealed a degree of heterogeneity in the potential response of benthic macroinvertebrates to eutrophication stressors across a single watershed. Those thresholds, in turn, were different than thresholds derived using the same statistical modeling approach for a large, state-wide dataset. Beyond the ecologically interesting patterns the delineation of reach-specific threshold reveals across heterogeneous watersheds, the reach-specific modeling approach could provide resource managers and regulators options to better manage their water resources by quantifying small scale differences in condition, biology, and biological response to stress. In most situations it is probably not practical to consider managing individual stream segments or sites with independent objectives and thresholds. However, some aggregation of these models and their derived thresholds may be useful in the development of watershed- or catchment-scale water quality improvement plans for 303d list waterbodies or strategic protective plans for high quality waterbodies in non-degraded conditions. The spatial heterogeneity of the thresholds that were identified and their relative difference to state-wide derived thresholds should provide managers the information to make informed decisions and balance the needs for uniform regulations and site specificity inherent to diverse, complex ecosystems like large stream networks.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All of the data are available from the SMC-SWAMP data portal https://smc.sccwrp.org/, which is linked in the body of the manuscript.

Acknowledgements

The authors thank the SCCWRP Commission’s Technical Advisory Group, Lester Yuan, and two anonymous reviewers for their help in improving the science and readability of the manuscript. Abel Santana (SCCWRP) assisted with creating the figures. Funding was provided by the California State Water Resources Control Board (Agreement No. 18-026-150), from Phase I and II, Proposition 84 Grant to the San Diego County Water Authority (Project No. 84-6-80006) and Riverside County Flood Control & Water Conservation District, the County of San Diego, and Rancho California Water District (Agreement No. 40009708).

References


CRediT authorship contribution statement

David J. Gillett: Writing – original draft, Writing – review & editing, Visualization, Formal analysis, Methodology. Raphael D. Mazor: Writing – original draft, Methodology, Data curation. Martha Sutula: Writing – original draft, Project administration, Funding acquisition, Conceptualization. Anne Holt: Methodology, Formal analysis, Data curation.