Quantifying biological constraints on stream integrity for classification and management priorization

Marcus W. Beck (marcusb@sccwpr.org), Raphael D. Mazor (raphaelm@sccwrp.org), Scott Johnson (scott@aquaticbioassay.com), Phil Markle (phil@lacsd.org), Peter D. Ode (peter.ode@wildlife.ca.gov), Ryan Hill (hill.ryan@epa.gov), Eric D. Stein (erics@sccwrp.org)

Introduction

- Degraded biological condition in streams can occur from individual or multiple stressors acting at different scales (Novotny et al. 2005; Townsend, Uhlmann, and Matthaei 2008; Leps et al. 2015). Identifying and mitigating causes of poor condition requires an understanding of how stressors propogate across space and time. Incomplete knowledge on drivers of change or high level of uncertainty in how biology is linked to drivers can lead to ineffective management actions. Placing bounds on effects of known drivers of change can reduce expenses and increase assurance of outcomes for targeted management (e.g., varying costs and challenges of urban stream restoration (Kenney et al. 2012; Shoredits and Clayton 2013))
- Effective stream management can depend on identification and prioritization of sites where activities are expected to have desired outcomes. This requires an understanding of how stressors affect biological integrity to place bounds on reasonable expectations for what is likely to be a possible outcome of a management action. This requires identifying biological constraints or limits on the potential range of biological conditions. Identifying an appropriate context for observed conditions can be used to prioritize. Context can be defined by models, expert knowledge, and/or defined value sets.
- We don't have good constraint tools to develop a context of expectation of what's possible at a site. This can help prioritize locations where management efforts will or will not have the intended outcomes. Biological filters act at different scales (Poff 1997) and we can use this information to describe an expectation for prioritization that is scale-specific. Landscape-level constraints are particularly relevant for macroinvertebrate communities in streams (Sponseller, Benfield, and Valett 2001)
- The goal of this study is to demonstrate application of a landscape model to classify and prioritize stream monitoring sites using estimated constraints on biological integrity. The model provides an estimate of context for biological condition that provides an expectation of what is likely to be achieved at a given site relative to large-scale drivers of stream health. The model was developed and applied to all stream reaches in California. A case study is used to demonstrate how the model can be used to classify and prioritize using guidance from a regional stakeholder group. Active stakeholder involvement was critical in applying the landscape models to define a framework for decision-making because priorities varied with management objectives.

Methods

Study area and data sources

- Brief description of CA, stream types and designated uses, PSA regions, management interests (e.g., southern vs. northern CA)
- Streamcat database used to quantify watershed land use at all sites (Hill et al. 2016)
- Streamcat data linked to National Hydrography Dataset Plus (NHD) (USGS (US Geological Survey) 2014), reach as individual unit for model output

The California Stream Condition Index (CSCI) (Ode et al. 2016; Mazor et al. 2016) was used as a measure of biological condition to assess site performance relative to reach classification. Benthic macroinvertebrate data used to calculate CSCI scores were collected at nearly 4800 sites between 2000 and 2014. Field data were collected during baseflow conditions typically between May and July following methods in Ode (2007). The CSCI is a predictive index of stream health that compares the observed taxa and metrics at a site to those expected under reference conditions. Expected conditions at a site are based on models that estimate the likely macroinvertebrate community in relation to factors that naturally influence biology, e.g., watershed size, elevation, climate, etc. The CSCI score at a site is based on an observed-to-expected ratio of taxa and a predictive multimetric index (pMMI) composed of six individual metrics that describe the structure and function of the macroinvertebrate community. The index score at a site can vary from 0 to 1.4, with higher values indicating an observed community with less deviation from reference conditions. Because the index was developed to minimize the influence of natural gradients, the index scores have consistent meaning across the state (Reynoldson et al. 1997). A threshold score based on a selected lower percentile of scores (e.g., 10%) at all reference sites is used to define nominally low and high scoring sites.

Building and validating landscape models

A prediction model of the CSCI was developed to estimate likely ranges of scores associated with land use gradients. Land use as urban and agricultural was quantified for the catchment of each stream reach in California using the StreamCat database (Hill et al. 2016). CSCI scores were modelled using only the estimates of urban and agricultural land use as the developed portion of the landscape within each stream reach. The model was incomplete by design to describe scores only in relation to large-scale constraints on biological condition that are not easily controlled by management actions or where costs to mitigate are likely to be excessive. The remainder of the variation in scores not related to landscape constraints could be attributed to additional, unmeasured environmental variables that influence stream biointegrity. Deviation of observed scores from the model predictions were considered diagnostic of variation not related to landscape effects.

Models were developed using quantile regression forests to estimate ranges of likely CSCI scores in different landscapes (N. Meinshausen 2006; Nicolai Meinshausen 2017). Quantile models evaluate the conditional response across the range of values that are expected, such as the lower and upper percentiles of the distribution, as compared to only the mean response with conventional models (Cade and Noon 2003). This allows use of model predictions to describe where bioassessment targets are unlikely to be met or where streams are unlikely to be impacted by placing bounds on the range of expectations relative to landscape constraints. Random forest models also provide robust predictions by evaluating different subsets of observations from random splits of the predictor variables. The final predictions are the averaged response across several models. These models have been used extensively in bioassessment applications (Carlisle, Falcone, and Meador 2009; Chen et al. 2014; Mazor et al. 2016) and can produce unbiased estimates that are relatively invariant to noisy relationships or non-normal distributions (Breiman 2001; Hastie, Tibshirani, and Friedman 2009). Quantile regression forests were used to predict CSCI scores in each stream reach from the 5^{th} to the 95^{th} percentile of expectations at five percent intervals (i.e., 5^{th}, 10^{th}, etc.).

Landscape estimates for the catchments of all NHD stream reaches in California were separated into calibration and validation data.

San Gabriel River watershed case study

Stream reach and bioassessment data from the San Gabriel River (SGR) watershed in southern California were used to develop reach classifications, site performance categories, and management priorities from the landscape models. A strong land use gradient occurs in the SGR watershed. Headwaters begin in the San Gabriel mountains where the land is primarily undeveloped or protected for reacreational use, whereas the lower watershed is in a heavily urbanized region of Los Angeles County. The San Gabriel river is dammed at four locations for flood control in the upper watershed and is hydrologically connected to the Los Angeles river to the west through the Whittier Reservoir in the lower watershed. Spreading grounds are present in the middle of the watershed for groundwater recharge during high flow. Nearly all of the stream reaches in the lower half of the watershed are channelized with concrete or other reinforcements.

Figure SGR watershed

The SGR watershed contains a diverse group of stakeholders from local municipalities, water districts, water quality regulatory agencies, consulting groups, and non-government organizations. Collectively, the San Gabriel River Regional Monitoring Program (SGRRMP) includes stakeholders from these groups that cooperatively work to increase awareness of issues in the SGR watershed and work to improve coordination of compliance and ambient monitoring efforts. The stakeholder workgroup included individuals from the SGRRMP with interests in water supply, improvements to water quality, habitat protection or creation, and storm water permitting. Individuals were selected for partipation to include a variety of mangement interests and based on willingness to adopt tools developed from the landscape models. The stakeholder workgroup met monthly over a six-month period to discuss model applications and to refine the interpretation of results. Stakeholder involement was critical for developing an assessment framework that met the needs of all engaged parties and ensured that final products were more likely to be incorporated into formal processes of decision-making.

Reach classification, site performance, and prioritization

A framework for identifying site priorities for management actions was developed using a three-step process. First, estimates of the range of expected CSCI scores at each stream reach in relation to land use were used to define reach classifications. Second, the relationship between observed CSCI scores and the reach classifications were then used to assign a relative performance value for each monitoring site. Third, site performance categories in relation to reach classification and bioassessment targets were used to define management priorities. This framework was developed through close interaction with the regional stakeholder group to demonstrate how the landscape model can be used as a management tool given that priorities will vary by interests and location. As such, the results are provided as a guide to facilitate decision-making rather than a prescription of targeted actions to manage stream health.

Identifying site priorities began with defining a classification framework for stream reaches to identify the possible or likely extent of biological constraints. Classifications were developed using the range of CSCI expectations at a reach relative to a chosen threshold for the CSCI to define nominally low or high scores. The reach classification was based solely on the intersection of the CSCI expectations at a reach with chosen CSCI threshold, where expectations could be below, above, or overlapping the threshold. Stream reaches with a range of CSCI score expectations entirely below the thresholds were considered likely constrained, whereas those with expectations entirely above were considered likely unconstrained. Reaches with score expectations that included the CSCI thresholds were considered possibly constrained or possibly unconstrained, where the distinction was based on location of the median expectation of a reach relative to the threshold.

CSCI scores from biomonitoring data were used to define performance of a sample site relative to the stream reach classification. For each of the four reach classifications (likely constrained, possibly constrained, possibly unconstrained, and likely unconstrained), the site performance was defined relative to the bounds of the expected CSCI scores. This provided a definition of site performance that can be used to understand the observed score relative to the biological context of a reach. Sites with observed scores above the upper limit of the reach expectation (e.g., above the 95^{th} percentile of expected scores) were considered "over-performing" and sites below the lower limit were "under-performing". Sites with CSCI scores within the range of expectations were as "expected".

Figure classification and performance

Site performance categories were further split relative to location to the selected CSCI threshold. This final split was created with the intent that description of site scores relative to a defined threshold (e.g., impairment threshold or restoration target) should also be considered. Specifically, a fourth category of site performance for each reach classifcation was added to define a site as above or below the threshold. For a

Table 1: Possible site types based on stream reach classification, site performance, and observed CSCI score. The observed score column describes where a CSCI score is observed relative to the lower and upper percentiles (e.g., 5^{th} and 95^{th}) of expected scores for a reach and the chosen CSCI threshold (e.g., 10^{th} percentile of scores at reference sites or 0.79) for nominally low or high values.

Reach expectation	Site performance	Observed score	Type
likely unconstrained	over scoring	$\geq 95^{th}$	1
	expected	5^{th} to 95^{th}	2
	under scoring	0.79 to 5^{th}	3
	under scoring	<0.79	4
possibly unconstrained	over scoring	$\geq 95^{th}$	5
	expected	0.79 to 95^{th}	6
	expected	5^{th} to 0.79	7
	under scoring	$<5^{th}$	8
possibly constrained	over scoring	$\geq 95^{th}$	9
	expected	0.79 to 95^{th}	10
	expected	5^{th} to 0.79	11
	under scoring	$<5^{th}$	12
likely constrained	over scoring	≥ 0.79	13
	over scoring	95^{th} to 0.79	14
	expected	5^{th} to 95^{th}	15
	under scoring	$<5^{th}$	16

likely unconstrained reach, underperforming sites below the minimum expected score were additionally defined as being above or below the CSCI threshold. Similarly, overperforming sites above the maximum expected score in a likely constrained reach were additionally defined as being below or above the CSCI threshold. For possibly constrained and possibly unconstrained reaches, sites that were performing as expected were addtionally defined as being below or above the CSCI threshold. In total, sixteen site types were defined for the three reach classification and three site performance classifications (table 1).

Table reach classification, site performance types and categories

Each site type was used to define a priority as a demonstration of how results from the landscape model can help achieve different stream management objectives. This final process relied exclusively on feedback from the stakeholder group that represented interests in monitoring, regulation, restoration, and protection. Priorities for each site type were defined accordingly with the expectation that site types will have different meanings for prioritization given the interest. Stakeholders from each sector were tasked with identifying their relevant priorities by ranking each site type from high to low priority using a blank template for reference (Figure). A brief description of the rationale for a site priority was also requested with the feedback.

Figure Site types template figure for prioritization

Sensitivity analyses and unclassified reaches

Stream reach classifications and site performance categories depend on the range of score expectations from the landscape model and the CSCI threshold for defining nominally low or high scores. This framework for identifying priorities was developed to allow flexibility in how the model could be applied. First, the framework can accommodate degrees of certainty in the model by allowing variation in the range of scores that are used to define a stream reach classification. The 5^{th} and 95^{th} percentile of expected scores at a reach are used as a default range in which a high degree of certainty in the model output is assumed. The ability to reduce this range (e.g., 25^{th} to 75^{th} percentile) to assume less certainty in the model is provided. The CSCI threshold can also be changed to assess effects of relaxing or increasing flexibility in a potential definition of a regulatory standard. A threshold of 0.79 is used by default as a measure of the 10^{th} percentile of scores at all reference (non-impacted) sites that were used to calibrate the CSCI index. This value can be increased to examine effects of a more stringent threshold or decreased for a more relaxed threshold. The combined effects of changing both the certainty in the model and the CSCI threshold were evaluated to estimate the changes in stream miles in each classification and the number of sites in each priority type.

Finally, some stream reaches were unclassifed following application of the landscape model to the statewide hydrography dataset. Unclassified reaches occurred when insufficient data in the StreamCat database were available to estimate CSCI predictions or if a stream catchment basin could not be defined for a particular reach. The latter was more common, particularly in developed areas where engineered channels or agricultural ditches were hydrologically removed from the natural stream network. Overall, unclassified reaches were not common in the statewide dataset but they may have regional importance depending on needs of local management groups. A preliminary approach for assigning biological expectations to unclassified reaches is demonstrated for 'typically' urban and agriculture reaches that relies on the range of expectations for reaches with similar land use by region.

Results

State-wide patterns

- Where does the model perform well, how does performance vary with validation and calibration datasets.
- What is the consistency of patterns? For example, percent stream miles as xyz by PSA.

Figure Statewide map.

Case study

- San Gabriel River Regional Monitoring Program
- Extent, classification, prioritization - probabilistic assessment to make broader conclusions.
- Relationships with environmental variables for constrained/unconstrained locations. Maybe apply to hardened/non-hardened reaches in constrained locations.
- What to do with unclassified streams - typical urban, typical ag.

Tables Priority by type, by perspective

Discussion

- What do priorities really mean? Depends on your interests, needs, values, etc.
- Changing certainty or CSCI treshold - mechanistic effets and implications. Mechanistically, reducing certainty in the model predictions by using a smaller range of score expectations will reduce the amount of stream reaches in the possibly constrained or possibly unconstrained category. Reducing the certainty will also change more sites from the "expected" perfomance category to the under- or over-performing category.
- Link with engineered channels study.
- Values of stakeholder interactions.

Supplement

Online application.

References

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Cade, B. S., and B. R. Noon. 2003. "A Gentle Introduction to Quantile Regression for Ecologists." *Frontiers in Ecology and the Environment* 1 (8): 412–20.

Carlisle, D. M., J. Falcone, and M. R. Meador. 2009. "Predicting the Biological Condition of Streams: Use of Geospatial Indicators of Natural and Anthropogenic Characteristics of Watersheds." *Environmental Monitoring and Assessment* 151 (1-4): 143–60. doi:10.1007/s10661-008-0256-z.

Chen, K., R. M. Hughes, S. Xu, J. Zhang, D. Cai, and B. Wang. 2014. "Evaluating Performance of Macroinvertebrate-Based Adjusted and Unadjusted Multi-Metric Indices (MMI) Using Multi-Season and Multi-Year Samples." *Ecological Indicators* 36: 142–51. doi:10.1016/j.ecolind.2013.07.006.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York: Springer.

Hill, R. A., M. H. Weber, S. G. Leibowitz, A. R. Olsen, and D. J. Thornbrugh. 2016. "The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States." *Journal of the American Water Resources Assocation* 52: 120–28. doi:10.1111/1752-1688.12372.

Kenney, M. A., P. R. Wilcock, B. F. Hobbs, N. E. Flores, and D. C. Martínez. 2012. "Is Urban Stream Restoration Worth It?" *Journal of the American Water Resources Association* 48 (3): 603–15. doi:10.1111/j.1752-1688.2011.00635.x.

Leps, M., J. D. Tonkin, V. Dahm, P. Haase, and A. Sundermann. 2015. "Disentangling Environmental Drivers of Benthic Invertebrate Assemblages: The Role of Spatial Scale and Riverscape Heterogeneity in a Multiple Stressor Environment." *Science of the Total Environment* 536: 546–56. doi:10.1016/j.scitotenv.2015.07.083.

Mazor, R. D., A. C. Rehn, P. R. Ode, M. Engeln, K. C. Schiff, E. D. Stein, D. J. Gillett, D. B. Herbst, and C. P. Hawkins. 2016. "Bioassessment in Complex Environments: Designing an Index for Consistent Meaning in Different Settings." *Freshwater Science* 35 (1): 249–71.

Meinshausen, N. 2006. "Quantile Regression Forests." *Journal of Machine Learning Research* 7: 983–99.

Meinshausen, Nicolai. 2017. *QuantregForest: Quantile Regression Forests*. https://CRAN.R-project.org/package=quantregForest.

Novotny, V., A. Bartosová, N. O'Reilly, and T. Ehlinger. 2005. "Unlocking the Relationship of Biotic Integrity of Impaired Waters to Anthropogenic Stresses." *Water Research* 39 (1): 184–98. doi:10.1016/j.watres.2004.09.002.

Ode, P. R. 2007. "Standard Operating Procedures for Collecting Benthic Macroinvertebrate Samples and Associated Physical and Chemical Data for Ambient Bioassessment in California." Surface Water Ambient Monitoring Program. Sacramento, CA.

Ode, P. R., A. C. Rehn, R. D. Mazor, K. C. Schiff, E. D. Stein, J. T. May, L. R. Brown, et al. 2016. "Evaluating the Adequacy of a Reference-Site Pool for Ecological Assessments in Environmentally Complex Regions." *Freshwater Science* 35 (1): 237–48.

Poff, N. L. 1997. "Landscape Filters and Species Traits: Towards Mechanistic Understanding and Prediction in Stream Ecology." *Journal of the North American Benthological Society* 16 (2): 391–409.

Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. "The Reference

Condition: A Comparison of Multimetric and Multivariate Approaches to Assess Water-Quality Impairment Using Benthic Macroinvertebrates." *Journal of the North American Benthological Society* 16 (4): 833–52.

Shoredits, A. S., and J. A. Clayton. 2013. "Assessing the Practice and Challenges of Stream Restoration in Urbanized Environments of the USA." *Geography Compass* 7 (5): 358–72. doi:10.1111/gec3.12039.

Sponseller, R. A., E. F. Benfield, and H. M. Valett. 2001. "Relationships Between Land Use, Spatial Scale and Stream Macroinvertebrate Communities." *Freshwater Biology* 46 (10): 1409–24. doi:10.1046/j.1365-2427.2001.00758.x.

Townsend, C. R., S. S. Uhlmann, and C. D. Matthaei. 2008. "Individual and Combined Responses of Stream Ecosystems to Multiple Stressors." *Journal of Applied Ecology* 45 (6): 1810–9. doi:10.1111/j.1365-2664.2008.01548.x.

USGS (US Geological Survey). 2014. "National Hydrography Dataset available on the World Wide Web."