Development of an Assessment Framework for Dry Ephemeral and Intermittent Streams in California and Arizona

Raphael D. Mazor
Jeff Brown
Eric Stein
John R. Olson
Matthew D. Robinson
Andrew Caudillo
Savannah Johnson
Gilbert Mak
Cara Clarke
Kevin O'Connor
Kamille Hammerstrom
Ross Clark

Southern California Coastal Water Research Project

SCCWRP Technical Report #1176

Development of an Assessment Framework for Dry Ephemeral and Intermittent Streams in California and Arizona

Raphael D. Mazor[1], Jeff Brown[1], Eric Stein[1], John R. Olson[2], Matthew D. Robinson[2], Andrew Caudillo[2], Savannah Johnson[2], Gilbert Mak[2], Cara Clarke[3], Kevin O'Connor[3], Kamille Hammerstrom[3], Ross Clark[3]

[1]*Southern California Coastal Water Research Project, Costa Mesa, CA*
[2]*California State University at Monterey Bay, Monterey, CA*
[3]*Central Coast Wetlands Group, Moss Landing, CA*

March 2021
Technical Report 1176

EXECUTIVE SUMMARY

Intermittent and ephemeral streams comprise a large portion of the arid southwest, yet tools to assess stream health have so far only been available for perennial and long-term intermittent streams, meaning that watershed assessments are incomplete — in some watersheds, substantially so. Managers therefore have only a limited ability to assess the effectiveness of their programs. Consequently, nonperennial streams, especially ephemeral streams, are often excluded from regulatory and management programs. To address this gap, researchers at the Southern California Coastal Water Research Project (SCCWRP), the Central Coast Wetlands Group (CCWG), and California State University at Monterey Bay (CSUMB) have developed new assessment tools to assess the ecological condition of intermittent and ephemeral streams when they are dry following the EPA's Level 1-2-3 framework. Specifically, we have developed Level 3 (L3) bioassessment indicators for use in dry ephemeral and intermittent streambeds, used them to validate a Level 2 (L2) tool, the California Rapid Assessment Method (CRAM) for episodic streams, and then conducted a Level 1 (L1) analysis where we estimate the extent and condition of non-perennial streams in California and Arizona.

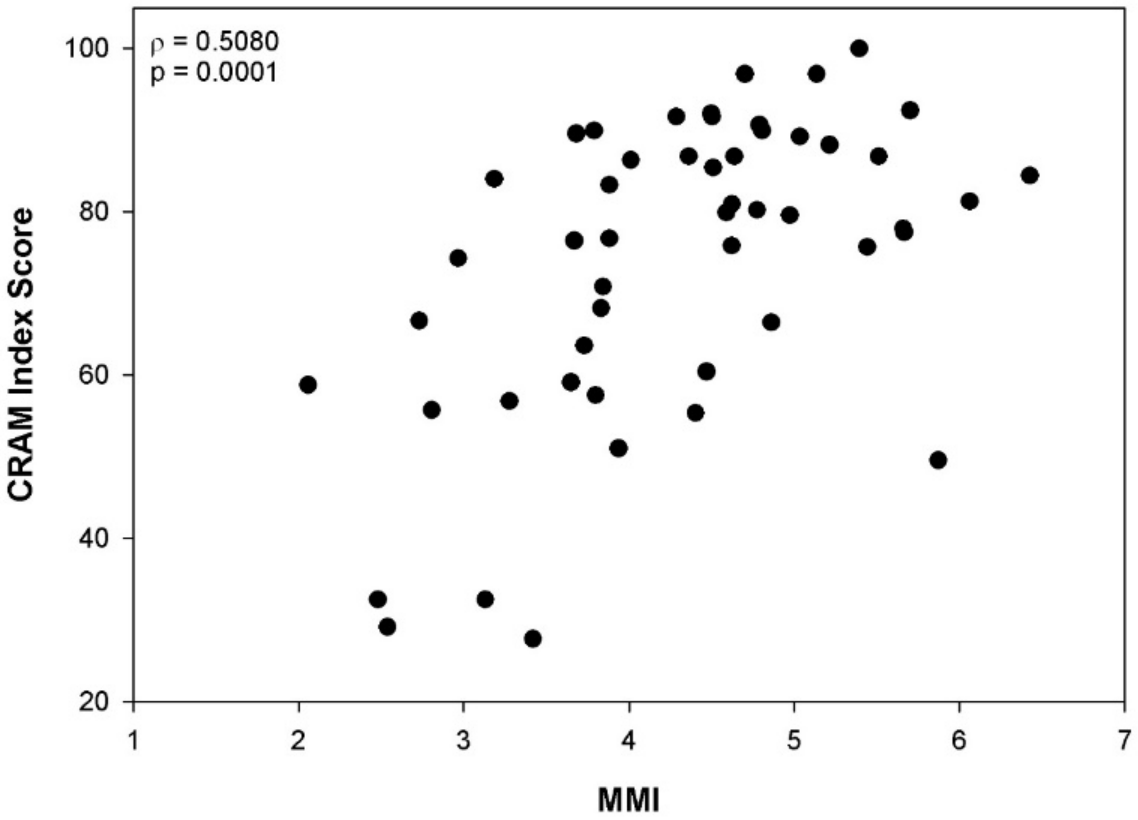
An L3 bioassessment index based on terrestrial indicators

Sampling terrestrial indicators in a dry streambed.

Using previously developed protocols for sampling terrestrial arthropods and bryophytes in dry streambeds, we developed predictive bioassessment index. Sampling 99 sites representing a range of natural and disturbed conditions in two states, we calibrated an index to predict metrics characterizing condition (e.g., combined richness of beetle and ant taxa) appropriate for different environmental settings. Differences from these natural expectations were scored and combined into an index that quantifies ecological condition.

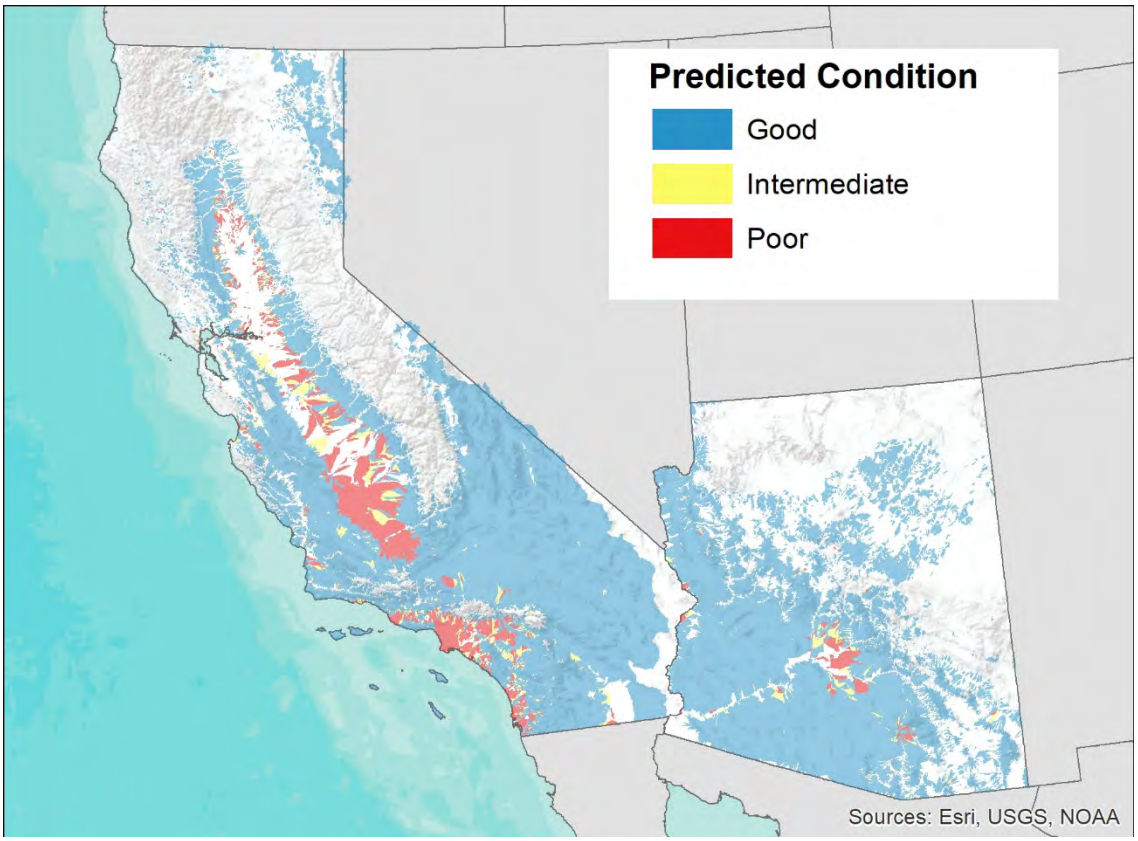
Validation of an L2 index for episodic streams

The L3 index and its component metrics provides an opportunity to validate a previously developed L2 assessment tool for episodic streams. The California Rapid Assessment Method (CRAM) module for riverine wetlands had been adapted for conditions in episodic streams, such as more appropriate expectations for biotic complexity in arid systems, more emphasis on naturally variable sediment regimes, and an overall larger assessment area. Several metrics showed strong correlations between L2 assessment scores and L3 measures of condition, indicating that the CRAM episodic module is a valid tool to assess conditions in these streams. Therefore, episodic systems can be included in regulatory programs where L2 assessments are needed (such as impact assessments and mitigation monitoring), as well as in ambient monitoring programs.



Scores for the episodic CRAM module were positively correlated with the multimetric index (MMI) derived from L3 indicators of dry stream condition.

L1 analysis shows that non-perennial streams are pervasive, and mostly in good shape in both states



Predicted conditions assessed by terrestrial L3 indicators in California and Arizona.

Models based on L2 and L3 assessment indicates that California and Arizona are dominated by ephemeral and intermittent streams, and the vast majority of these streams are likely in good biological condition. Poor conditions are most likely in certain areas, such as near the cities of Phoenix and Los Angeles, as well as in agricultural areas like the San Joaquin and Imperial Valleys of California. Notably, intermittent streams in California appear to experience higher levels of stress, as only a slim majority are likely to be in good condition, and more than 10% are likely to be in poor condition.

Recommendations

- **Integrate dry streams into monitoring programs**. This pilot study has generated much of the required infrastructure to begin large-scale monitoring, such as developing protocols and establishing a standard set of metrics and indices for generating and analyzing data. With sufficient training, monitoring practitioners in the California and Arizona can begin assessing the condition of dry streams.

- **Refine reference definitions for dry streams.** This study adapted an approach for identifying undisturbed streams that was originally developed for perennial systems, which heavily emphasizes minimizing activity in the upstream watershed. However, in

dry streams, upstream disturbance may be a poor indicator of local factors that have a larger influence on in-stream condition. Improved measurements of local disturbance (e.g., measures of habitat alteration, sediment contamination, or hydromodification) may be more useful for identifying reference sites in systems where upstream land-use is a poor proxy.

- **Collect data from additional reference sites, particularly in underrepresented regions.** The limited reference data generated by this study may not capture the full range of natural conditions, nor does it provide information about seasonal or interannual variability. Collecting these data may lead to the development of more precise indices, as well as better guidance on conditions where these indices are best suited.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This project was funded by the Environmental Protection Agency Grant CD-99T65301, and includes data supported under projects funded by the California State Water Resources Control Board (SWRCB), as well as California's Department of Fish and Wildlife (CADFW). Technical guidance was provided by Patricia Spindler (Arizona Department of Environmental Quality, ADEQ), Jason Jones (ADEQ), Peter Ode (CADFW), Andrew Rehn (CADFW), Maryann McGraw (New Mexico Environment Department, NMED), Kris Barrios (NMED), Alisha Steward (Queensland Government), Chad Loflen (Regional Water Quality Control Board of San Diego, RWQCB-SD), Betty Fetscher (RWQCB-SD), Michael Bogan (University of Arizona), Núria Cid Puey (University of Barcelona), Chris Solek (US Army Corps of Engineers, USACE), Melissa Scianni (U.S. Environmental Protection Agency, USEPA), Ken Fritz (USEPA), Tracie Nadeau (USEPA), Robert Leidy (USEPA), Rachel Stubbington (Nottingham Trent University, NTU), Chloe Hayes (NTU). and Chuck Hawkins (Utah State University).

INTRODUCTION

Nonperennial rivers and streams (NPRS) comprise a large portion of stream-miles in the arid southwest of the USA, but at this time, we have few tools that can be used to assess their condition. Consequently, monitoring programs may overlook these streams, despite their importance in providing beneficial uses or protecting adjacent perennial waters. With an increasing population and global change leading to extreme floods and droughts, land managers need to understand how freshwater systems respond to human impacts and relate to our clean water supply. Humans in industrialized countries have had a significant impact on freshwater ecosystems. To understand these systems, we can monitor and assess the relationship between the biota and the rivers, lakes, wetlands, and streams that create the above-ground freshwater network. Determining the best way to assess these systems is integral to evaluating their health.

Non-perennial stretches of rivers are common features in headwater systems, but they can also be found throughout river networks (Steward et al. 2012) and play key ecological roles in a watershed context during dry and wetted phases. Datry et al. (2014) described NPRS as continuously shifting habitat mosaics driven by alternating phase-changes (i.e., flowing, drying, and dry) which maintain habitat heterogeneity. These alternating phases can lead to temporal shifts in nutrient processing and availability which may affect nutrient balances and export downstream (von Schiller et al. 2011). Even when surface water is completely absent, dry river channels often have sub-surface flows that sustain river flows downstream (Levick et al. 2008), making them important for maintaining watershed connectivity. Additionally, dry river channels function as storage areas for nutrients and organic material (Steward et al. 2012). Alternating phase changes can act as disturbances for both aquatic and terrestrial biota, but NPRS provide habitat for organisms with various strategies and adaptations (physiological or behavioral) to cope with these changes (Datry et al. 2016. For example, some taxa (e.g., aquatic invertebrates), are present as juveniles during the flowing phase and are dormant as eggs during dry phases and require both phases to persist within a system (Armitage and Bass 2013; Stubbington et al. 2018). The wide range of NPRS biodiversity includes: prokaryotes, fungi and protozoans (Febria et al. 2015; Romani et al. 2017), diatoms (Tornés and Ruhí 2013), vascular plants (Sabater et al. 2017), aquatic invertebrates (during flowing phases and inhabiting the hyporheic zone during dry phases) (Wood et al. 2010; Stubbington and Datry 2013; Stubbington et al. 2017), fish (Kerezsy et al. 2017), terrestrial and semiaquatic invertebrates (Corti and Datry 2015; Sánchez-Montoya et al. 2016b; Steward et al. 2017), as well as amphibians, reptiles, birds, and mammals (Sánchez-Montoya et al. 2016a; Sánchez-Montoya et al. 2017). Given their widespread distribution, abundance, and important ecosystem functions including hydrologic connectivity with adjacent perennial waters, the condition of non-perennial systems and their ability to function properly can greatly influence the health of entire watersheds.

Episodic streams are particularly abundant and widespread in drier regions of California and Arizona. Most streams in California and Arizona exhibit some degree of ephemeral flow (NHD 2020; Levick 2010). Despite their intrinsic values and importance to hydrologically connected waterbodies, they are typically excluded from ambient surveys and overlooked in management programs, because most wetland and stream assessment tools have been focused on perennial streams (Rehn et al. 2015). Ephemeral streams (especially in desert locales) are under increasing pressure from development, including new urban/suburban and infrastructure projects, and, most recently, alternative energy production facilities (e.g., wind and solar) (Hamada et al. 2016).

Assessment tools for ephemeral and intermittent streams are necessary to allow managers to prioritize streams for protection or restoration, assess impacts, and develop and evaluate performance standards for mitigation or remediation.

To address this need, we developed a suite of tools following the U.S. Environmental Protection Agency's three-tiered framework for assessing the condition of wetlands and other aquatic resources. Level 1 (L1) assessments are map-based and require no field visits. Level 2 (L2) are rapid, field-based assessment tools of waterbody structure and function but require no sample collection. Level 3 (L3) are intensive methods that require lab analysis of field-collected samples, such as tools based on biological indicators or water chemistry. In general, L3 methods provide the most accurate measures of wetland condition, whereas L1 assessments are most suitable for large-scale application. L3 assessments may be used to L2 assessments, and both L2 and L3 may be used to validate L1 assessments.

Prior to this effort, California had an extensive assessment toolkit focused on perennial streams. This toolkit included L3 indices for benthic macroinvertebrates (Mazor et al. 2016), benthic algae (Theroux et al. in review), as well as L2 tools for riparian wetlands (CWMW 2013). Arizona, too, had an L3 index for perennial streams based on benthic macroinvertebrates (ADEQ 2015) and is developing one for diatoms (P. Spindler, personal communication), but no L2 assessment tools. Adaptation of L3 indices in flowing intermittent streams has been completed in California (Mazor et al. 2014) and is underway in Arizona (P. Spindler, personal communication). However, apart from a pilot study in the San Diego hydrologic region (Mazor et al. 2019a), there have been few efforts to apply or create new L3 assessment tools for dry intermittent and ephemeral streams in the USA.

California's L2 assessment tool, the California Rapid Assessment Method (CRAM), has multiple modules for different wetland types, and the traditional riverine module applies in many intermittent streams even when they are not flowing. However, this module was less effective in drier streams characterized by more episodic flow events, where the module exhibited relatively poor scores where systems were largely undisturbed. In particular, two of the CRAM attributes that comprise the final score — the hydrology and biotic structure attributes — did not adequately capture the dynamics in natural episodic streams. Large, episodic floods in dryland climates means that morphological channel features are frequently reworked and may lack the distinctiveness of these features in perennial streams in more temperate climates. For temperate-climate streams, where precipitation is more evenly distributed both temporally and spatially, lower and more frequent intermediate flow events tend to govern the equilibrium channel shape and size. However, because dryland ephemeral channels exhibit a more rapid response to rainfall and shorter-duration flows than temperate-region streams, they tend to be shaped by high magnitude flow events. These streams experience such extreme and rapid variations in flood regime, that they rarely reach process-form equilibrium where flow conditions change too rapidly for channel bed forms to develop a form matching that flow. Thus, existing sedimentary structures can give a misleading picture of the flow that occurred (Levick et al. 2008). The transitory nature of morphological features in these climates is further enhanced by generally less cohesive soils and poorly vegetated banks. These processes create a fabric of highly-varied, transient channel forms that confound determinations of active versus relict stream processes and conventional notions of stable and unstable channel forms (Vyverberg 2010; Stein et al. 2011). Therefore, a new CRAM module was developed for Episodic Riverine systems.

This report describes efforts to develop and validate assessment tools for dry ephemeral streams and intermittent rivers. First, we describe the development of L3 indices based on a variety of biological indicators (i.e., terrestrial arthropods in the streambed, arthropods on riparian vegetation, and bryophytes on the banks and channel) using data collected from arid portions of California and Arizona. We then use these L3 indices to validate the CRAM module for episodic streams described above. Finally, we conduct two L1 assessments: one based on estimating the extent and condition of streams where the new L3 indices should be used, and the other based on the L2 episodic CRAM module.

L3 ASSESSMENT TOOLS: DEVELOPMENT OF BIOLOGICAL INDICATORS FOR DRY INTERMITTENT AND EPHEMERAL STREAMS BASED ON TERRESTRIAL ARTHROPODS AND BRYOPHYTES

Non-perennial rivers and streams are estimated to make up more than 50% of all river systems worldwide (Datry et al. 2016), consisting of intermittent rivers and ephemeral streams (IRES) which both cease to flow for extended periods of time. Intermittent rivers can maintain flows seasonally, while ephemeral streams maintain flows only after large rain events (Nadeau et al. 2015). NPRS are expected to become more prevalent with longer dry periods as growing human population, urban development and climate change place increased stress on water resources (Sabater and Tockner 2009). This change will be especially apparent in arid to semi-arid regions where droughts and water shortages are already common, creating challenges for river monitoring and management that rely on tools that require the presence of surface water for assessment (e.g., indices based on benthic arthropod composition).

Traditional bioassessment indicators that rely on aquatic assemblages (e.g., benthic arthropods or algae) can be difficult to use in certain non-perennial systems due to the unpredictability of flows (Steward et al. 2018), and may provide an incomplete picture of stream health by overlooking dry-phase biota. Monitoring programs often have optimal index periods which are generally set over a span of months when baseflow conditions are expected, flow variability is low and aquatic communities are relatively stable (Barbour et al. 1999). In some arid climates, such as Southern California where the optimal index period for sampling may be difficult to predict, streams that have a high probability of drying are often excluded entirely from monitoring programs (Hall et al. 1998). In some instances, streams may remain dry for months to years and assessing their ecological health using traditional bioassessment methods on a timeline needed to inform management decisions is impossible. This can bias ambient surveys such as the United States Environmental Protection Agencies National Rivers and Streams Assessment that excludes streams that are dry during the index period. The inclusion of dry streams in ambient surveys would expand the target population of streams and result in a more comprehensive assessment of ecological condition of streams across the U.S.

We developed sampling methods to characterize terrestrial arthropod and bryophyte assemblages of NPRS and assessed the ability of biological metrics that characterize these assemblages to distinguish reference condition streams from those impacted by human activities. This information will support the development of bioassessment tools for NPRS that can be used during the dry phase. For the purposes of this study, a reference condition site is characterized as having minimal anthropogenic disturbance in the watershed (Stoddard et al. 2006; Ode et al. 2016). We developed metrics characterizing aspects of terrestrial arthropod (ground-dwelling and vegetation-dwelling) and bryophyte assemblages known to respond to anthropogenic disturbances. These metrics described the richness, taxonomic composition, diversity, and feeding groups or growth forms for each assemblage. We accounted for bias in metric values caused by natural variation by adjusting metrics that were influenced by naturally occurring environmental gradients. We evaluated metric responses to anthropogenic disturbance and assessed the ability of metrics to discriminate between reference and non-reference sites (e.g., background variability, signal-to-noise ratio). We also evaluated the role of human activities (i.e., % urbanization, % agriculture and % urban and agriculture land cover in the watershed) and percent fines as limiting factors to the biological responses. Combined with traditional protocols

and metrics for flowing systems, the development of biological assessment tools for non-perennial systems during the dry phase would allow management agencies to assess stream condition regardless of presence or duration of flow, reduce the number of streams excluded from bioassessments, and result in more comprehensive assessments of watershed condition.

Approach

There are two common approaches for assessing the biological integrity of waterbodies (Hawkins et al. 2000, Mazor al. 2019b): multimetric indices (MMIs) and measures of taxonomic completeness (e.g., ratios of observed-to-expected taxa, or O/E indices). Both methods share the goal of transforming complex taxonomic data into simple measures of biological condition. MMIs are comprised of metrics based on taxonomic composition, pollution tolerance values, or life history traits; condition is interpreted to be healthy when these metric values are close to values observed at reference sites. In contrast, O/E indices use species-distribution models to predict which taxa are likely to occur. When all likely taxa occur (i.e., the ratio of observed to expected taxa is close to 1), conditions are interpreted to be healthy.

Methods

Study Area and Site Selection

The study focused on arid (xeric) portions of California and Arizona, regions where ephemeral and intermittent streams comprise the majority of stream-miles (Figure 1). We divided the study area into 10 regions. Two regions in California (i.e., the South Coast and the Central Valley) were sampled under a variety of other programs, but using the same methods described below. We attempted to identify at least 2 reference sites and 3 non-reference sites in each region. Reference sites were defined as those with minimal human activity in the watershed, combined with little evidence of local habitat disturbance (detected through consultation with local experts, as well as physical habitat data, screened as described below).

Landscape-scale measures of human activity were derived from the Streamcat dataset (Hill et al. 2016) using thresholds modified from Ode et al. (2016). Specifically, we looked for catchments that had low signs of human activity at both the watershed and local catchment scale (Table 1). A total of 62 reference sites were identified based on professional judgment of project leads, and confirmed by evaluating evidence of local and landscape-scale human activity (Table 2).

Table 1. Reference thresholds applied to Streamcat variables.

Human activity measure	Watershed threshold	Catchment threshold
% agricultural land use	3	3
% urban land use	3	10
% agricultural or urban land use (combined)	5	10
Road density (km/km^2)	2	5
Road crossings	50	3
Canal density (km/km^2)	10	2
Mine density	0	0

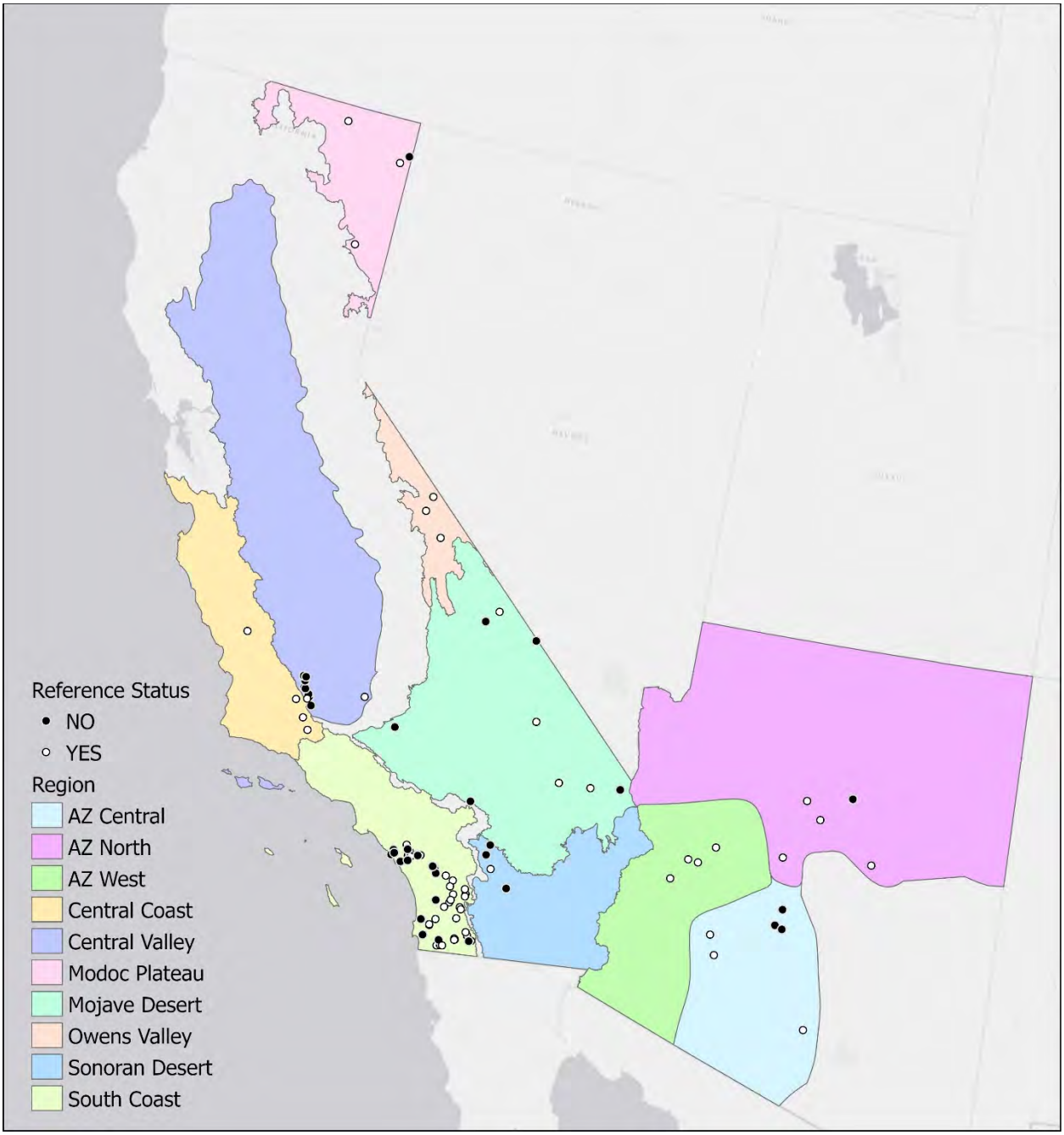


Figure 1. Map of the study area.

Table 2. Number of reference and non-reference sites sampled in each region.

Region	Reference sites	Non-reference sites	Total
California	*51*	*33*	*84*
South Coast*	33	16	49
Central Coast	4	0	4
Central Valley*	3	8	11
Modoc Plateau	3	1	4
Mojave Desert	3	3	6
Sonoran Desert	1	3	4
Eastern Sierra	4	2	6
Arizona	*11*	*4*	*15*
Central	3	3	6
Northern	4	1	5
Western	4	0	4
Total	62	37	99

Data collection

Following the protocol described in Robinson et al. (2016), we collected biological and habitat data at 99 sites in California and Arizona. At each site, we designated a representative 160-m reach at each site, which we separated into eight sections. In each section, we collected channel and vegetation-dwelling arthropods using ramped pitfall traps and a canvas bag, respectively (Robinson et al. 2016). Ramped pitfall traps offer advantages over traditional pitfall traps because they reduce disturbance to the habitat, and they are also more suitable for sampling in stream beds with hard substrates (i.e., cobbles, bedrock, or concrete) that make digging pitfall traps impractical (Pearce et al. 2005; Patrick and Hansen 2013). The traps were left out for 24 hours to collect both diurnal and nocturnal arthropods, which were stored in jars along with the contents of the traps for later identification.

Vegetation dwelling arthropods were collected on plants in or near the channel, following Robinson et al.'s methodology of visually picking the healthiest plant in each section. We wrapped the plant in a 1-m^2 canvas bag and hit it a total of 30 times (Robinson et al. 2016), using a plastic pipe to dislocate any vegetation-dwelling arthropods. The contents of the bag were placed in a jar and preserved with 70% ethanol for later identification.

Along with arthropods, we also collected bryophytes (moss) at each site, which were collected using a floristic approach (Newmaster et al. 2005; Robinson et al. 2016). We designated three mesohabitats (Robinson et al. 2016): right and left banks and the channel. We designated 20 minutes to search for moss in each habitat and allotted 12 minutes to collect moss (Robinson et al. 2016). We collected up to a total of five samples of moss from each mesohabitat, collecting

them by hand in a pattern from most diverse to least diverse patches in each microhabitat (e.g., soil, rock, or wood) present (Robinson et al. 2016).

We measured aspects of physical habitat such as channel depth, sediment size, and slope for each transect in the sample reach following Robinson et al. (2016). At each of the 8 transects, we measured sediment particle size, riparian vegetation, channel morphology, and microhabitat types. Along with physical habitat information, we also recorded any stressors observed. Using stressor categories such as fire breaks, walking paths, and other anthropogenic disturbances, we assigned a categorical value to each stressor based on how prevalent it is. These values ranged from “Not present” to “25% of the reach” or “over 75% of the reach”.

Metric calculation and initial screening

We calculated 233 metrics that characterize composition of the three assemblages (Table 3). Metrics were largely based on taxonomic composition, but a few reflected life history traits (specifically, feeding strategies of certain Coleoptera and spiders). For the arthropod metrics, we also calculated metrics based on invasive species (e.g., Argentine ants, *Linepithema humile*). For each type of metric, we calculated richness, relative richness, abundance, and log abundance; for some higher-level groups (e.g., Araneae), we also calculated diversity and evenness metrics.

Metrics were screened to eliminate those with insufficient information to continue analysis. This step included removing richness ranges less than 5 and removing metrics with a high number of zero values (> 2/3). Metrics passing these screens were then modeled to account for natural variability, as described below. Adjusted metrics were then screened to identify those with the highest level of responsiveness to watershed alteration.

Table 3. Summary of metrics used to characterize stream conditions.

Bryophytes	Riparian arthropods	Streambed arthropods
Bryophyte morphospecies	Arthropods	Arthropods
Bryophyte families	Coleoptera	Coleoptera
Bryophyte genera	Hymenoptera	Ground beetles
Acrocarps	Ants	Rove beetles
Pleurocarps	Hemiptera	Ground + Rove beetles
Bryaceae	Thysanoptera (thrips or silverfish)	Predator beetles
Pottiaceae	Spiders	Herbivore beetles
	Other arthropods	Fungivore beetles
	Coleoptera + Ants	Fungivore, dead wood, and detritivore beetles
	Coleoptera + Ants + Spiders	Hymenoptera
	Coleoptera + Spiders	Ants
	Ants + Spiders	Thysanoptera
		Diptera
		Hemiptera
		Archaeognatha (bristletails)
		Earwigs
		Spiders
		Wolf spiders
		Ground spiders
		Web spiders
		Ground-hunting spiders
		"Other" hunting spiders
		Mites
		Isopods
		Collembola
		Other arthropods
		Coleoptera + Ants
		Coleoptera + Spiders
		Coleoptera + Ants + Spiders
		Ants + Spiders
		Invasive arthropods

Assessing and Reducing Influence of Natural Gradients in Characterizations of Biological Responses to Human Activity

We modeled each of our biologic metrics as a function of 89 natural environmental gradients by constructing a 500-tree RF model. The environmental gradients used as predictors included local catchment and watershed scale measures of climate, topography, geology, and hydrology and were derived using geographic information systems analysis or obtained from the StreamCat data set (Hill et al. 2016) and field observations. These predictors were chosen based on their known influence on stream hydrology and other habitat features that may affect the terrestrial assemblages we sampled. After fitting the RF models using all 89 environmental predictors, we assessed the importance for each predictor in each RF model by calculating the percent increase mean squared error (%IncMSE). Percent increase MSE is calculated as the difference between the MSE of the model when all values of a predictor are permuted and the original MSE rate divided by the standard error (Cutler 2007). If the %IncMSE of any given predictor was < 0, we removed the predictor from the model and fit the model again using only those predictors with %IncMSE > 0. By using this method, each model potentially has its own unique suite of predictors. Following Vander Laan and Hawkins (2014), we adjusted the metric values by substituting the residual value (observed value – expected value) as the new metric value if the models explained > 10% of the variation in an individual metric's values observed at reference sites. All statistical analyses related to the RF modeling were completed using the Random Forest package (Liaw and Wiener 2002) using R software (R core team 2016).

Assessing Metric Ability to Distinguish Ecological Condition Between Reference and Non-Reference Sites

We assessed the ability of metrics to distinguish between reference and non-reference sites using multiple criteria previously used in MMI development studies. We used criteria modified from Herbst and Silldorff (2009) designed to quantitatively assess a metrics ability to provide clear discrimination of human activity on benthic macroinvertebrate communities. These criteria were developed to assess metrics that have a negative relationship with human activity (i.e., "decreaser" metrics) as well as metrics that have a positive relationship with human activity (i.e., "increaser" metrics). Following Herbst and Silldorff (2009), we eliminated metrics that did not pass any of the following criteria: 1) background variability measured as the coefficient of variation less than 0.2 (i.e., standard deviation of reference site metric values divided by the mean metric value at reference sites), 2) signal from human activity greater than 1.5 or less than 0.67 for increaser metrics (measured as the ratio between the mean of reference site metric values and the mean of non-reference site metric values), 3) signal-to-noise ratio greater than 1.5 (measured as the absolute difference between the mean of reference metric values and the mean of non-reference metric values divided by the standard deviation in the reference site metric values), and 4) for all decreaser metrics discrimination efficiency defined as having less than 50%, 35%, and 25% of non-reference metric values greater than the 10^{th}, 25^{th}, and 50^{th} quantiles of reference site metric values, respectively. For all increaser metrics, discrimination efficiency criteria is met if less than 50%, 35%, and 25% of non-reference metric values measured below the 90^{th}, 75^{th} and 50^{th} quantiles of reference site metric values, respectively.

For all metrics passing at least one of the Herbst and Silldorff criteria, we calculated t-statistics between mean metric values at reference and non-reference sites to assess their ability to respond to human activity. We considered metrics with the greatest absolute t-statistics to be the most

responsive to disturbance and be the most likely to distinguish reference sites from non-reference sites. We expect metrics that are minimally influenced by natural variation, pass at least one of the metric assessment criteria and are the most responsive to human activity will have the greatest potential to be used in bioassessment tools in NPRS during the dry phase. Only metrics passing at least one of the Herbst and Silldorff criteria and those with t-statistics > 1.80 were retained to be used in MMI development.

Removing Duplicate Metrics

Following Schoolmaster et al. (2013), we removed duplicate metrics, and scaled and reflected metrics to prepare for the combination of individual metrics into candidate MMIs. To identify duplicate metrics, we calculated Pearson's R squared for each combination of metrics and removed duplicate metrics that had a Pearson's R > 0.95 or < -0.95. This method differs from removing redundant metrics that simply have high correlations by instead only removing the metrics with high correlations that contain the same information, such as metrics describing percent native species and percent non-native species (Schoolmaster et al. 2013).

Scoring Metrics

We scored metrics so that they ranged from a scale of 0 (indicating poor conditions) to 1 (indicating reference conditions), such that all metrics (whether they were increasers or decreasers) could be interpreted the same way. This scoring facilitates their incorporation into an MMI. To score metrics, we used the following equations from Schoolmaster et al. (2013):

$$m_{scaled} = \frac{m - L}{U - L}$$

$$m_{scored} = \max(m_{scaled}) - m_{scaled}$$

where m is the raw metric value, m_{scaled} is the scaled metric, m_{scored} is the scored metric, and L and U are the 2.5th and 97.5th percentiles of m, respectively. For all metrics above and below the 2.5th and 97.5th percentiles of m, the m was set to L and U, respectively.

Prior to calculating metric scores for MMI development, we removed all negative metric values that resulted from adjusting metric scores that were correlated with natural variation. To remove negative metric values, we used the following equation:

$$m_{pos} = (m + abs(\min(M))$$

where m_{pos} is the positive metric value, m is an individual site metric and $abs(\min(M)$ is the absolute minimum value of the given metric across all sites. To ensure all metrics that were candidates for MMI development had negative relationships with stress and resulted in an MMI with a negative correlation to stress, we reflected all metrics that were increasers following Schoolmaster et al. (2013). We reflected all metrics with a positive relationship with stress (i.e., metrics with negative t-statistics) using the following equation:

$$m_{ref}(\max(M) - m)$$

where m_{ref} is the reflected metric score, $\max(M)$ is the maximum value of a metric score across all sites, and m is an individual site metric.

Multimetric Index Development

We used a modified algorithm from Schoolmaster et al. (2013) to select combinations of candidate metrics and create an empirical based MMI that results in an MMI with the greatest negative correlation with stress. The algorithm creates candidate MMIs by using each metric as a starting metric and adding metrics that result in the greatest negative correlation with stress until each candidate metric is used. This results in multiple candidate MMIs for each beginning metric. We calculated candidate MMIs using the following steps:

1) For each combination of an individual metric m_1, find the best combination of each other metric, m_j, and find the best combination of $m_1 + m_j$ that has the strongest negative correlation (Pearson's R) with watershed condition (specifically, the Index of Watershed Integrity [IWI], a variable in StreamCat that combines multiple measures of land cover and land use). Record the correlation.
2) Add m_j to m_1 and create a candidate MMI that resulted in the most negative correlation with IWI.
3) Continue this process until all metrics have been added to the candidate MMI based on criteria from step 1.
4) Repeat steps 1-3 using a new starting metric, m_1, until additive MMIs have been created for all m_j metrics.

This algorithm creates the best possible combinations of metrics based on negative correlations with IWI and gives each candidate metric the chance at being the first metric in the MMI. After all candidate MMIs have been created, the best MMI for each beginning metric is chosen based on the most negative correlation with IWI. All duplicate MMIs are removed from the list of candidate MMIs for further analysis.

After determining the best candidate MMI for each starting metric, we fit a linear model using the MMI as the response variable and IWI as the predictor variable. We calculated the Akaike information criterion (AIC) for that model and calculated the difference in all model AIC scores from the lowest AIC score (Δ AIC). We calculated Δ AIC as:

$$\Delta\, AIC = AIC_j - \min\,(AIC)_j$$

where $\Delta\, AIC$ is the difference between an individual MMI AIC value and the minimum AIC value of candidate MMIs, AIC_j is the AIC value of each individual MMI and $\min\,(AIC)_j$ is the minimum AIC value for all candidate MMIs. We considered models that had a Δ AIC < 2 to be the best performing models. If no MMIs had a $\Delta\, AIC$ value < 2, we selected the MMI with the minimum AIC value as the best performing MMI.

Results

Assessing and Reducing Influence of Natural Gradients in Characterizations of Biological Responses to Human Activity

Natural environmental gradients explained greater than 10% of the variation in 117 of the 233 metrics we evaluated (Table 4). The models were most successful in explaining the variation in ground-dwelling arthropod metrics (45 models > 10% variation explained), followed by vegetation-dwelling arthropod metrics (43 models > 10% variation explained), and bryophyte metrics (29 models >10% variation explained). Models that explained more than 10% of variation in scores at reference sites were used to adjust metric values by subtracting the predicted value from the observed value (henceforth called "adjusted metrics"). A total of 67 metrics (or adjusted metrics, where appropriate) passed screens recommended by Herbst and Silldorff (2009). Sixty-two percent of these metrics reflected riparian arthropod communities, while the rest reflected streambed arthropod communities; no bryophyte metrics met these criteria.

Table 4. Top 8 metrics within each assemblage group, ranked by the variance in reference site scores explained by natural factors in a random forest model.

Metric	R^2	Assemblage	Form
Streambed arthropods			
T_Is_La	0.62	Isopoda	Log abundance
T_Site_Ri	0.41	Arthropoda	Richness
T_Ar_Ri	0.4	Araneae	Richness
T_Ar_Di	0.39	Araneae	Diversity
T_Ot_Ri	0.38	Other Arthropoda	Richness
T_ArGh_La	0.37	Ground-hunting Araneae	Log abundance
T_Ly_Ab_RO	0.35	Lycosidae	Relative abundance
T_Cl_Ab_RS	0.35	Collembola	Relative abundance
Riparian arthropods			
V_FoAr_La	0.59	Mixed Arthropoda	Log abundance
V_Ar_La	0.54	Araneae	Log abundance
V_CoFoAr_La	0.48	Coleoptera Formicidae and Araneae	Log abundance
V_CoAr_La	0.45	Coleoptera and Araneae	Log abundance
V_CoFo_Ev	0.41	Coloeoptera and Formicidae	Evenness
V_FoAR_Ev	0.39	Formicidae and Araneae	Evenness
V_CoFoAr_Di	0.39	Coleoptera Formicidae and Araneae	Diversity
V_Fo_La	0.39	Formicidae	Log abundance
V_CoFo_Di	0.38	Coleoptera and Formicidae	Diversity
Bryophytes			
CB_BrFrm_Ri	0.58	Bryophyte families on channel and banks	Richness
B_BrFm_Ri	0.51	Bryophyte families on banks	Richness
CB_BrGe_Ri	0.44	Bryophyte genera on channel and banks	Richness
B_BrGe_Ri	0.39	Bryophyte genera on bank	Richness
CB_Br_Di	0.35	Bryophytes on channel and bank	Diversity
CB_Br_Ri	0.33	Bryophytes on channel and bank	Richness

CB_Ba_Ri	0.33	Bryaceae on bank	Richness
C_BrFm_Ri	0.33	Bryophyte families on channel	Richness
C_BrGe_Ri	0.31	Bryophyte genera on channel	Richness

Evaluation of Metric Response to Human Activity

From these 67 metrics, we identified 45 metrics with the greatest absolute t-statistics with t-statistics > 1.80 (referred to hereafter as "responsive metrics"). Of the three assemblages analyzed, only ground-dwelling and vegetation-dwelling arthropod assemblages had metrics with t-statistics > 1.80 which accounted for 17 and 28 of the 45 responsive metrics, respectively (Table 5). These 45 metrics included measures of richness, abundance, taxonomic composition and functional feeding groups. Two examples of highly responsive metrics are shown in Figure 2.

Table 5. Top 8 responsive metrics with the greatest absolute t-statistics.

Metric	t-statistic
Streambed arthropods	
Coleoptera richness	3.84
Formicidae richness	3.82
Percent inasive abundance	3.41
Richness of fungivore, deadwood, and detrivore Coleoptera	3.11
Relative richness of Lycosidae	3.05
Relative richness of Archaeognatha	3.03
Coleoptera diversity	3.02
Formicidae diversity	2.93
Riparian arthropods	
Log abundance of Coleoptera and Formicidae	6.47
Log abundance of Formicidae	4.98
Log abundance of Coleoptera, Formicidae, and Araneae	4.94
Log abundance of Formicidae and Araneae	4.89
Log abundance of Coleoptera and Araneae	4.32
Log abundance of Araneae	4.18
Relative richness of Coleoptera, Formicidae, and Araneae	3.22
Relative richness of Coleoptera and Araneae	2.94

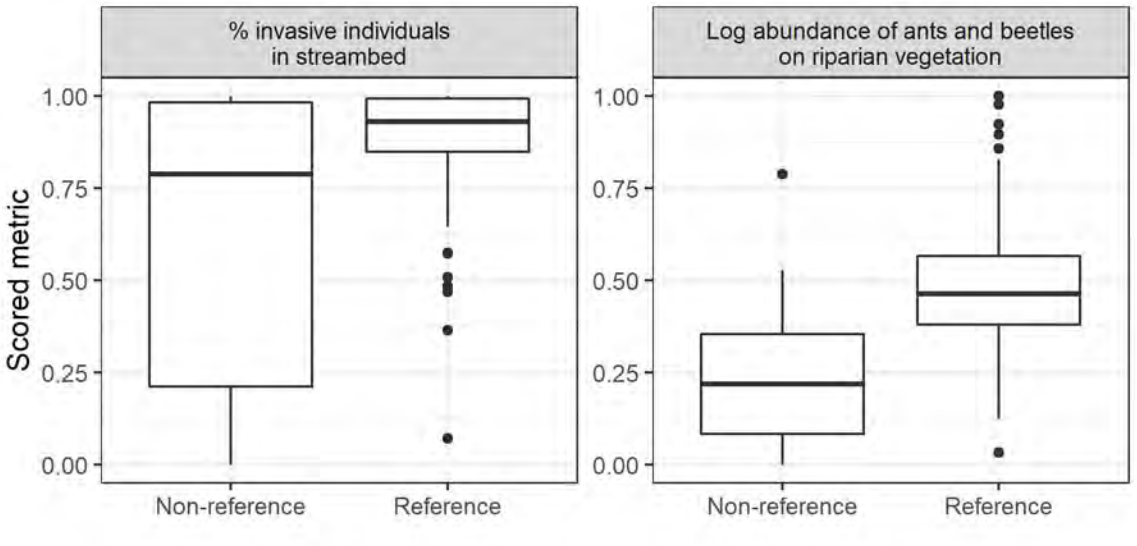


Figure 2. Scored metric values at two example responsive metrics.

Remove Duplicate Metrics and Scale and Reflect Remaining Metrics

Following Schoolmaster et al. (2013), we removed 4 duplicate metrics that had a Pearson's R > 0.95 or < -0.95 (Table 6). The 4 metrics we removed from further analysis were metrics that included combinations of taxonomic groups (e.g., combined Formicidae and Araneae richness). We selected the simplest version of the correlated to retain the metrics that were the easiest to calculate and contained the same information without adding more taxonomic groupings.

Table 6. Metric pairs with high correlations and their respective Pearson's R values.

Metric	Correlated metric (excluded)	Pearson's R
Richness of Araneae	Richness of Formicidae and Araneae	0.96
Richness of Araneae	Relative richness of Formicidae and Araneae	0.98
Log abundance of Coleoptera and Araneae	Log abundance of Coleoptera, Formicidae, and Araneae	0.98
Log abundance of Araneae	Log abundance of Formicidae and Araneae	0.97

Multimetric Index Development Results

Metric screening

We calculated a total of 1640 candidate MMIs. For each starting metric, m_1, we found the best combination of m_j metrics that created an MMI with the most negative correlation with the IWI. We excluded all other candidate MMIs whose starting combination of metrics began with the same m_1 metric. From the remaining 41 candidate MMIs we excluded 7 MMIs that contained the same metrics as another candidate MMI, leaving 34 candidate MMIs to consider for further analysis. We did not find a candidate MMI that had a $\Delta\ AIC < 2$. We therefore chose the candidate MMI with the lowest AIC value to be the best MMI to respond negatively to stress referred to hereafter as the "final MMI").

The final MMI was MMI15 (Table 7), which was the candidate MMI with the least amount of combined metrics. Only two of the three assemblages were included in the final MMI and included 3 ground-dwelling arthropod and 5 vegetation-dwelling arthropod metrics for a total of 8 metrics; no bryophyte metrics were selected for inclusion. The 8 metrics included in the final MMI contained metrics that characterized richness, taxonomic composition (including percent abundance of invasive taxa), diversity and feeding groups. This MMI had a Pearson correlation coefficient of -0.60 with the IWI.

Table 7. Metrics selected for inclusion in the final MMI.

	Final MMI
Metric	**Metric description**
m10	Ground-dwelling herbivorous Coleoptera richness
m15	Ground-dwelling Formicidae diversity
m16	Ground-dwelling Thysanoptera relative abundance
m18	Vegetation-dwelling invasive percent total abundance
m23	Vegetation-dwelling Araneae relative richness
m24	Vegetation-dwelling Araneae log abundance
m35	Vegetation-dwelling Hemiptera log abundance
m39	Vegetation-dwelling combined Coleoptera and Formicidae evenness

Performance of Final MMI

The MMI had a mean score of 4.71 at reference sites, with a standard deviation of 0.74. Scores were somewhat lower and more variable at nonreference sites (mean: 3.97; standard deviation: 0.95). Reference scores were somewhat lower in California (mean: 4.67) than Arizona (mean: 4.91), but this difference was not significant. MMI scores differed significantly between reference and non-reference sites ($t = -4.24$, $p < 0.001$; Figure 3). In addition, they had a strong negative correlation with the IWI (Pearson's $R = -0.60$; Figure 4).

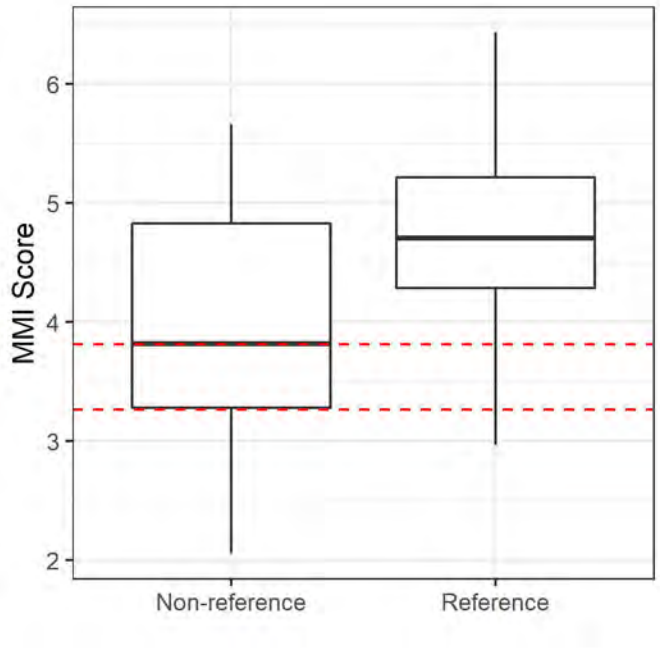


Figure 3. MMI scores at reference and non-reference sites. Red dashed lines indicate the 10th (i.e., 3.81) and 1st (i.e., 3.26) percentiles of reference sites.

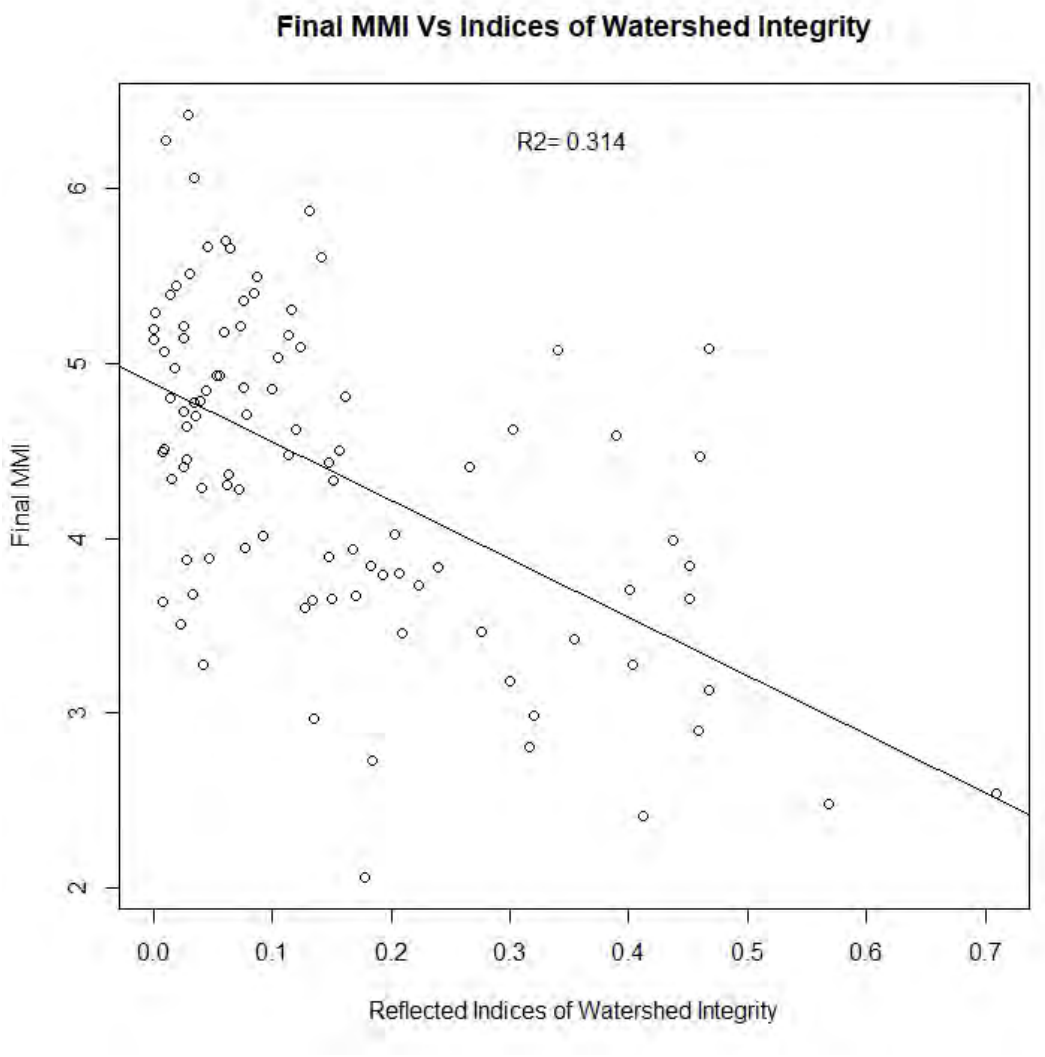


Figure 4. Relationship between MMI scores and the Index of Watershed Integrity from StreamCat (Hill et al. 2016).

Discussion

We were able to create a multimetric index (MMI) to quantitatively measure condition in dry intermittent and ephemeral streams, proving the feasibility of including these streams in watershed management programs. This assessment index for dry streams will fulfill a major gap in monitoring and regulatory programs in the arid southwest. However, few steps are needed to reach the stage where intermittent and ephemeral streams can be fully integrated in these programs.

The indices require validation with independent data, particularly at reference sites that represent the full range of conditions where the index may be applied. Validation activities should also assess inter- and intra-annual variability to fully assess the precision and repeatability of these

assessments. Greater taxonomic resolution may also improve observed strength of metric-stressor relationships, as multiple species within the same families may exhibit different responses to the same stressor.

These analyses were based on the best taxonomic data available at the time, which in general was family or genus level, with morphospecies identified to provide additional resolution. It is likely that additional information may be gained with greater taxonomic effort, which would allow incorporation of available life-history information (Steward et al. 2018; Stubbington et al. 2019). For example, non-native Argentine ants (Formicidae: *Linepithema humile*) may be more common in ephemeral streams that receive urban runoff than in natural ephemeral streams due to the soil moisture preferences of this species (Holway 1998; Menke et al. 2007). Because our study was limited to morphospecies, we are unable to tell if we are observing this pattern in the San Diego region. Trait-based efforts have been productive for bioassessment applications in perennial streams, and would likely apply here as well. Molecular methods (e.g., DNA barcoding) may also enhance our ability to generate highly resolved taxonomic data for these biological indicators.

To identify reference sites, we followed the approach of Ode et al. (2016), which set criteria for identifying reference-quality perennial and intermittent streams in California (which are generally higher-order than some of the ephemeral headwaters included in this study). It remains unknown if the criteria identified there are meaningful for the indicators we studied. Ode et al. (2016) emphasized screens based on measures of human activity in the upstream watershed (e.g., urban development) under the assumption that these activities can impact in-stream communities. The low frequency of flow in some of our intermittent and ephemeral sites may decouple or weaken the links between upstream disturbance and condition at a site, which may account for the relatively weak relationships we observed between metric scores or the MMI and land use. A reference definition that incorporates locally measured stressors and human activity, covering both habitat and water or sediment quality, should be considered. Our use of a "proximity of local activity" metric represents a first-cut approach to incorporating local information in defining reference sites.

The value of multiple assemblages in bioassessment

The final MMI incorporates metrics representing two of three assemblages we sampled: streambed arthropods and riparian arthropods. No bryophyte metrics were selected for the final MMI, largely due to the inability to account for natural variation in many of these metrics at reference sites. Although individual bryophyte metrics may have value as bioindicators, they do not at this time offer a useful tool to incorporate into monitoring programs.

In contrast, numerous arthropod metrics showed large responses to measures of stress, and metrics reflecting both riparian and streambed assemblages were included in the MMI. As a multi-assemblage index, this MMI is comparable to hybrid algal indices developed for southern California (Fetscher et al. 2014; Theroux et al. 2020), based on both diatoms and soft-bodied algae. Multi-assemblage indices incorporate a greater diversity of lines of evidence when assessing condition, leading to a more complete picture, as well as superior index performance. However, cost concerns may make single-indicator indices (which require less sampling and analysis effort) desirable, and these should be explored in the future, despite their potentially weaker performance.

Our ability to quantify human stressors across the sites we sampled in the study is limited by the lack of data on several important stressors like hydrological alterations caused by groundwater extraction or the effects of cattle grazing. This limited our ability to successfully develop a single combined gradient that would allow us to parse reference from non-reference sites, or to evaluate fine-scale gradients of condition within non-reference sites. Because of this, the majority of metrics that correlated well with human activity did not show a consistent response to the stressor gradients we examined.

Our understanding of the mechanisms driving biological responses to disturbance in dry streams is also limited. We can speculate why some metrics increase with stress whereas others decrease, but a better understanding of how upstream stressors affect local channel environments and how this translates into changes in local biota of dry streams is needed. For this reason, our study and its implications for management may be limited by our binary classification of sites (e.g., reference or non-reference) which limits the resolution needed to make direct links between individual stressors and metric responses. Although we were able to show evidence that four of our metrics were likely affected by human land use, we do not know the drivers (e.g., increased runoff, increased sedimentation, water extraction, pollution) associated with developed land use that underlie this relationship. Other factors including the intensity of disturbance, duration of the disturbance, interactions between disturbances and differences among sites (e.g., hydrologic regime, topography, geomorphology) may also play roles in determining how terrestrial biota respond to human disturbance. More quantitative methods of evaluating certain impacts (e.g., using wildlife cameras to measure grazing intensity) may elucidate these relationships.

Although we have demonstrated the feasibility of assessing dry intermittent and ephemeral streams, further work still needs to be done to better understand the community dynamics and complex biotic and abiotic interactions that exist in the dry channels of nonperennial streams, which would give managers better confidence for incorporating these tools in their monitoring programs. Future studies should focus on developing and testing the causal mechanisms driving biological responses to better understand the direct effects of human disturbance on terrestrial dry stream communities.

L2 Assessments: Validation of the California Rapid Assessment Method Module for Episodic Streams

Introduction

Rapid assessment of streams allows for cost-effective and repeatable characterization of stream health at scales from local sites to entire watersheds to regions or statewide, yet they are developed with a set of assumptions that require validation with appropriate data, such as bioassessment indices or other L3 assessment tools. This validation gives managers assurance that L2 assessments provide useful information about waterbody condition, and they can confidently incorporate L2 assessments in a range of management decisions.

Background on CRAM

The California Rapid Assessment Method (CRAM) is the primary L2 assessment tool in California, and it is currently used in a range of monitoring and management programs. CRAM provides an overall Index score (ranging from 25 to 100) that indicates the general health of a stream or wetland and its capacity to perform important functions and services. The CRAM Index score is an average of four main "Attributes" of condition (i.e., buffer and landscape context, hydrology, physical structure, and biotic structure). Each Attribute is calculated from two to five metrics and submetrics (Table 6). The assessment of each metric or submetric is based on visual indicators surveyed during a field visit of less than half a day.

Table 8. CRAM attributes and metrics with summaries of each metric.

Attributes	Metrics	Metric Summary
Buffer and Landscape Context	Stream Corridor Continuity	Measures presence of intact habitat upstream and downstream 500 m
	Percent with Buffer	Percent of area surrounded by at least 5 m of buffer land cover
	Buffer Width	Average of 8 buffer width measurements up to 250 m
	Buffer Condition	Degree of soil disturbance, impact of human visitation, and vegetation quality (native vs. non-native)
Hydrology	Water Source	Anthropogenic influence on water sources (extractions or inputs) within local watershed up to 2 km
	Sediment Transport	Alterations to natural sediment transport processes
	Hydrologic Connectivity	Access to adjacent slopes without levees, road grades, or other obstructions

Physical Structure	Structural Patch Richness	Number of habitat structures present from a list of potential patch types for episodic streams
	Topographic Complexity	Complexity of micro- and macro-topographic features
Biotic Structure	Number of Plant Layers	Number of plant height classes that cover at least 5% of the area
	Number of Co-dominant Species	Total number of living plant species that comprise at least 10% of any plant layer
	Percent Invasive Species	The percent of the total number of co-dominant species that are listed by Cal-IPC as invasive
	Horizontal Interspersion	The complexity of plant zones (species assemblages or mono-specific stands)
	Vertical Biotic Structure	Overlap of plant layers

Validation and the CRAM Development Process

There are six steps to CRAM development, as described in Sutula et al. (2006) and outlined on the CRAM website (http://www.cramwetlands.org/about). These steps include:

1. Definition phase
2. Basic design phase
3. Verification phase
4. Validation phase
5. Module production phase
6. Ambient survey phase

Previous work, funded by the USEPA and others, accomplished phases one through three. Initial field testing occurred in 2010 in association with the Solar II energy development project and the related Sunrise Powerlink Transportation Project. Verification was completed in 2013 and 2014 with a survey of episodic streams in Southern and Central California. The method design phase involved an extensive literature review and experts were convened to provide input on indicators of condition. The 2010 and 2013 development efforts provided a solid foundation to launch the current Validation phase project and demonstrated that the episodic stream module can effectively differentiate between “good”, “fair”, and “poor” sites. The initial field book for Episodic Riverine CRAM was drafted as part of the 2013 project for field testing. While the

initial method proved to differentiate good and poor condition systems, several thorny issues still needed resolution in order to finalize the field book.

This validation effort documents relationships between CRAM results and independent measures of condition (specifically, L3 assessment metrics described in the preceding section) in order to establish CRAM's defensibility as a representative and repeatable measure of wetland condition (Stein et al. 2009). This validation is an essential step in establishing CRAM's scientific defensibility, which is needed to support its utility for local, state, and federal programs.

Methods

Validation of the Episodic CRAM module followed the systematic process described by Stein et al. (2009), which prescribed several steps to validation:

1. Identify the gradient of stress
2. Identify Level 3 data to validate the CRAM module
3. Identify metrics that will be calculated from the detailed Level 3 data
4. Create conceptual models of the expected relationship between the detailed data and CRAM scores
5. Identify field sites where Level 3 data are available or possible to collect
6. Conduct CRAM assessments at the sites identified
7. Analyze relationships between CRAM scores and Level 3 metrics

Identifying the Gradient of Stress

Episodic streams, like all wetlands, can be impacted by surrounding land use (Chiu et al. 2017). Landscape conditions can therefore be an effective predictor of wetland health (Roth et al. 1996; Micacchion and Gara 2008). Adjacent and upstream land cover affects wetlands and streams through many processes, including polluted runoff, habitat loss, and alteration of hydrologic dynamics. In episodic streams the impacts are more closely associated with local land use practices, while the entire upstream watershed has more indirect effects (Shaw and Cooper 2008). Most studies of land use impacts on streams focus on the full watershed (e.g., Taka et al. 2016), while for episodic streams the immediately adjacent activities seem to have more influence (Levick et al. 2008). This difference may be due to the flashy episodic flows in these systems are less connected to the upstream watershed, compared to perennial streams where water is always flowing from upper watershed to lower watershed. When episodic streams are surrounded by natural open space they are much more likely to support flora and fauna, and provide other important functions, such as groundwater recharge and nutrient cycling (Levick et al. 2008). Conversely, when they are close to developed areas such as urban or agricultural land covers, they are more likely to have reduced functions and species diversity (Krueper 1996; Pima County 2000). This study selected a range of sites along a gradient of development pressure, including some sites in open space preserves or parks, and others in cities and agricultural areas (Figure 5).

Identify Level 3 Data

We used the three biological indicators described in the preceding section: terrestrial arthropods in the dry streambed, arthropods on riparian vegetation, and bryophytes in the channel. As part of an earlier literature review (Mazor et al. 2019a), these assemblages were identified as having high potential for use as biological indicators because of their ease of sampling and plausible relationships with environmental quality in dry stream channels.

Identify Metrics from Level 3 Data

As described in the preceding section, the species richness, abundance of individuals, diversity, and evenness of terrestrial arthropods and bryophytes populations were analyzed to establish condition metrics. Some metrics combined taxonomic groups for analysis; for example, Coleoptera (beetles) and Formicidae (ants) were combined into one metric that measures the

diversity or species richness of both groups. As described a multimetric index (MMI) was created from a suite of high-performing metrics that characterized the two arthropod assemblages (no bryophyte metrics were included in this index).

Create Conceptual Models

The expected relationship between CRAM Index and Attribute scores and Level 3 data were predicted a priori for each Level 3 indicator. The establishment of a priori relationships between CRAM attributes and Level 3 metrics was complicated because the level 3 methods were being developed concurrently with CRAM validation. The development team relied on other terrestrial invertebrate studies to inform the predictions. Generally, we assumed that there would be a positive correlation between CRAM Attributes and Index scores and invertebrate population metrics, except for taxa that are more responsive to human disturbance. The MMI was predicted to have a positive correlation with CRAM scores since MMIs are designed to be responsive to a similar condition gradient from highly disturbed to pristine reference condition.

Identify Field sites

To complete the validation process, the development team selected sites across California and Arizona with a range of anticipated condition. This project partnered with the Arizona Department of Environmental Quality to include the state of Arizona in the development process, allowing for the use of this CRAM module in both states. We selected sites in several regions, including the Mojave Desert, Sonora Desert, Owens Valley, Modoc, the Central Valley, the Central Coast of California, Northern Arizona, Central Arizona, and Western Arizona. CRAM data in the South Coast of California was collected under related projects. Altogether, these sites represented gradient of human disturbance and stress. Most sites were on public land, with some located within private preserves or other privately-owned lands. A total of 56 sites were analyzed (Figures 2-4).

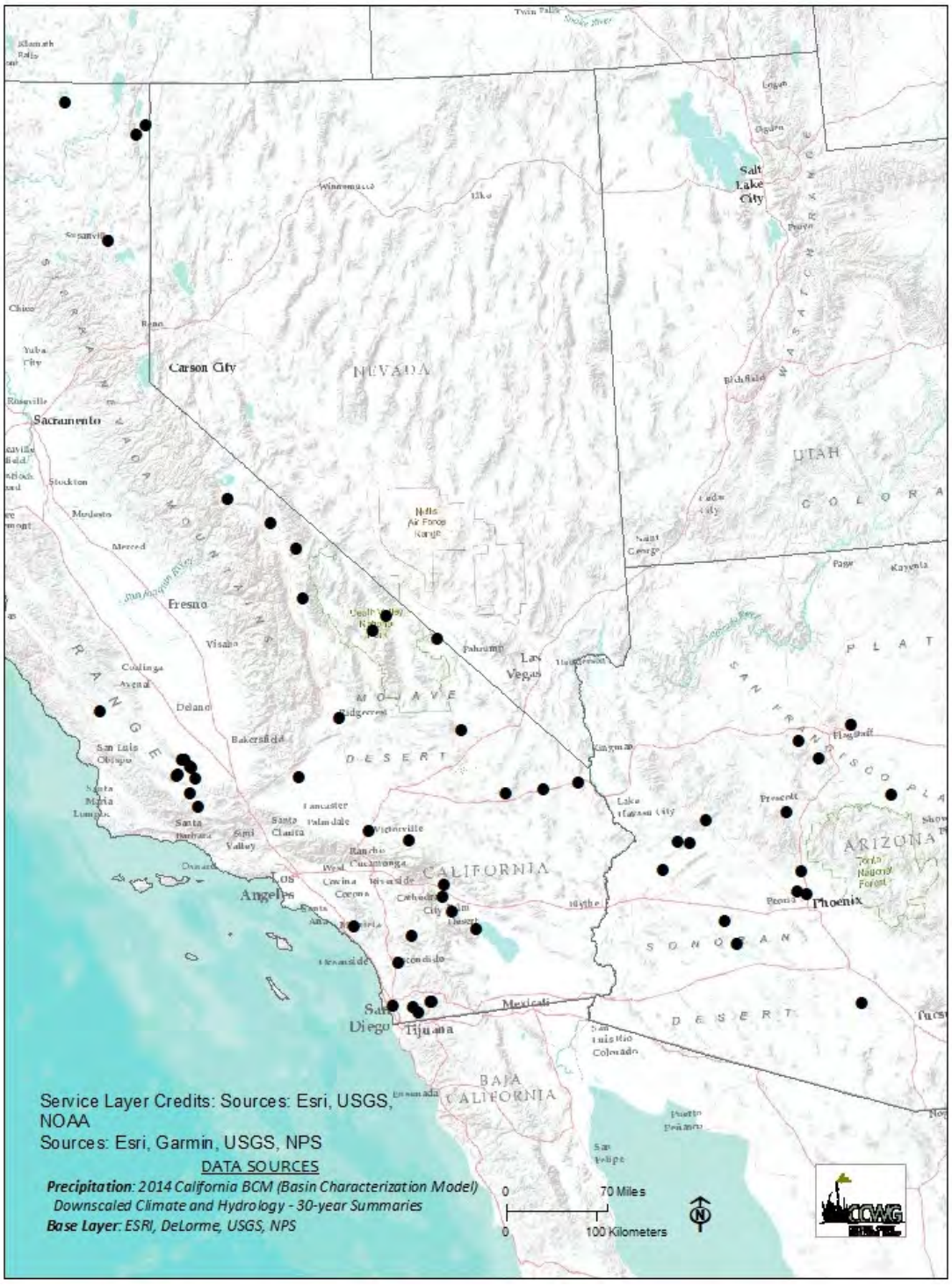


Figure 5. Map of all sites selected for L2 sampling and analysis.

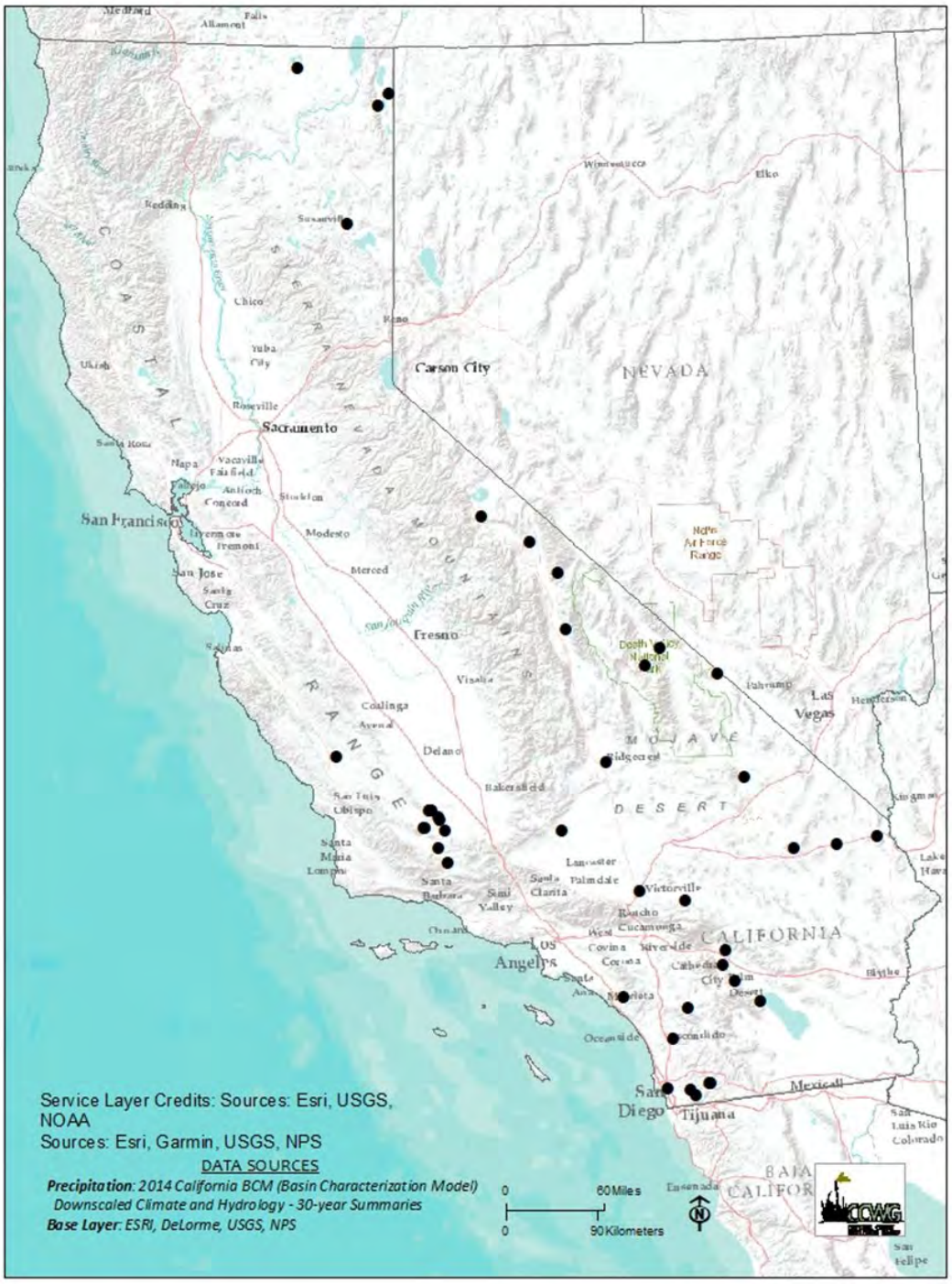


Figure 6. Map of sites selected in California for L2 sampling and analysis.

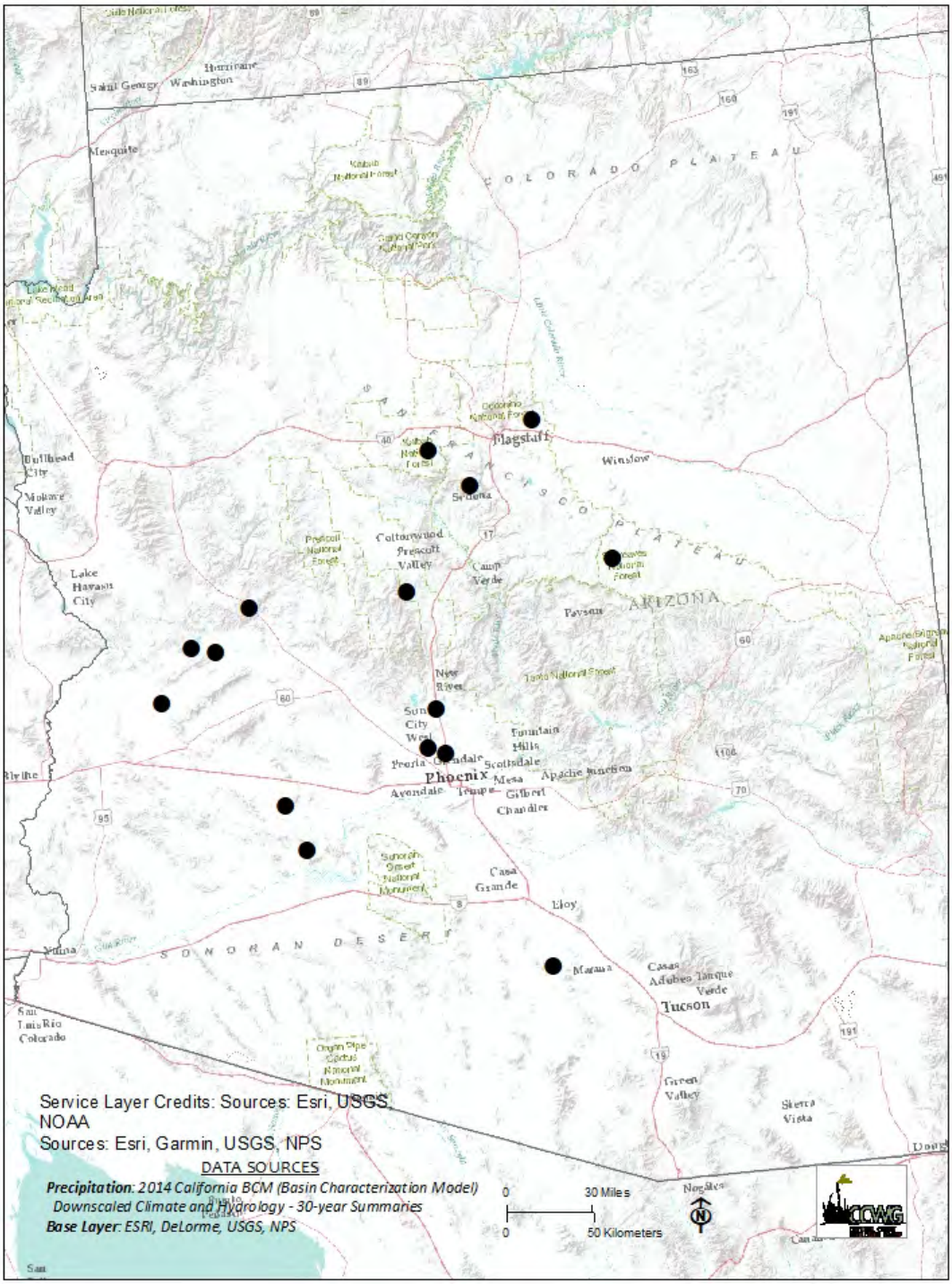


Figure 7. Map of sites selected in Arizona for L2sampling and analysis.

Conduct Field Assessments

Field assessments were conducted using the Episodic CRAM module (version 2.0, February 2018) at 35 sites for this project during 2018 and 2019 (Figure 8). Project partners conducted 15 assessments in Arizona during 2018. CRAM was conducted at the sites in Southern California in 2018 and 2019 as well. All assessments followed the quality assurance procedures outlined in the CRAM QA Plan (CWMW 2016) and the QAPP for this project (SCCWRP 2018). These same sites were assessed during the same time period by CSUMB researchers using the Level 3 arthropod and bryophyte assessment tool described in the previous section.

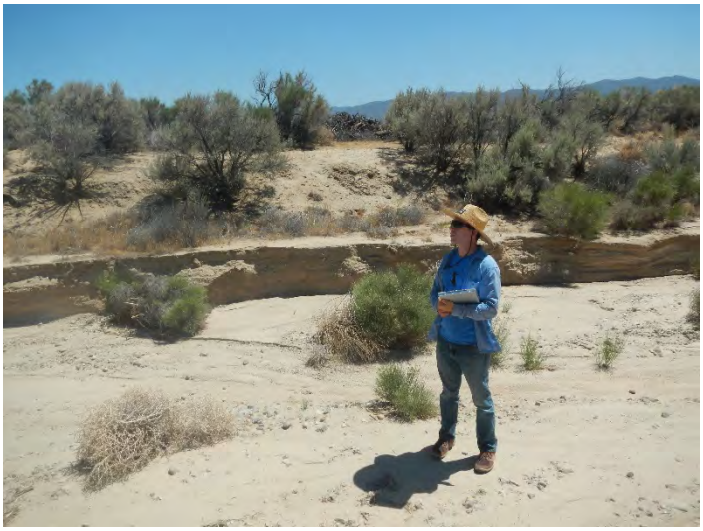


Figure 8. Conducting CRAM in a dry wash.

Analyze relationships between CRAM scores and Level 3 metrics

Spearman rank correlations were conducted for the CRAM Index score and each of the CRAM Attribute scores relative to the terrestrial arthropod and bryophyte metrics and MMI. The non-parametric Spearman rank correlation was used because data did not meet assumptions of bivariate normality (Dodge 2010). Each metric within CRAM was treated as independent, so p-values were calculated separately for each CRAM metric and independent measures (arthropod and bryophyte metrics). We also evaluated correlations between CRAM and climate and geographic variables and land use (EPA's StreamCat database; Hill et al. 2016). Climate and geographic variables included mean temperature, maximum temperature, elevation, and watershed area. Land use related variables included an Index of Watershed Integrity (IWI), Index of Catchment Integrity, measures of inorganic fertilizer application and manure application. These last two indicators serve as a proxy for agricultural land use, and while there isn't typically regular rainfall to transport fertilizers from fields into streams, episodic flow events could still bring contaminants from farmed areas into stream systems. The data were examined among reference sites (least disturbed) and non-reference sites. A simple T-test was used where the data conformed to a normal distribution, and in cases of non-normality, a non-parametric Wilcoxon Two Sample Test was used. This helped to confirm that CRAM assessments statistically

differentiated reference and non-reference sites, which were designated a priori based on landscape factors. Landscape variables were tested between reference and non-reference to determine if geographic factors (elevation, temperature, or watershed area) affected the outcome of the assessments. Correlations among StreamCat variables and CRAM Index and Attribute scores were investigated. All calculations were conducted using SAS 9.3 software (SAS Institute Inc. 2011).

Results

An effective rapid assessment method must be responsive to a range of conditions and be sensitive to human disturbance (Sutula et al. 2006; Stein et al. 2009). The CRAM Index score is a composite of the four Attribute scores and represents the overall ecological condition of the wetland. The CRAM tool generates a minimum value of 25 and a maximum value of 100. The CRAM Index scores collected for this project ranged from 28 to 100, with a median score of 78 (Figure 9). We determined that the scores did not biased high or low values, although data were moderately skewed (Bulmer 1979) towards higher scores (skewness = -0.96). The range of scores collected from our survey (20-100) confirms the responsiveness of the Episodic CRAM module to score a full range of expected conditions within southwest United States ephemeral streams.

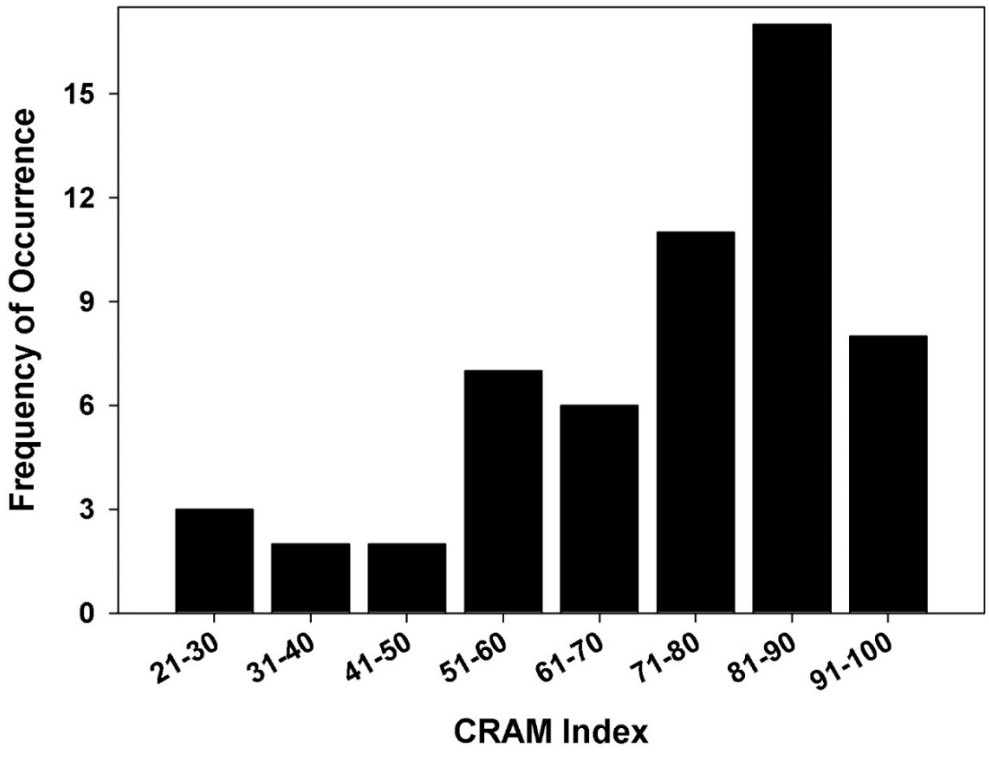


Figure 9. Histogram of CRAM Index scores (n=56).

A full range of scores (minimum of 25 and the maximum of 100) were measured for each CRAM Attribute except for hydrology (Buffer and Landscape Context 25-100, Hydrology 33-100, Physical Structure 25-100, and Biotic Structure 25-100) (Figure 10). These data support our assumption that each Attribute is responsive to the varying conditions of Ephemeral Streams.

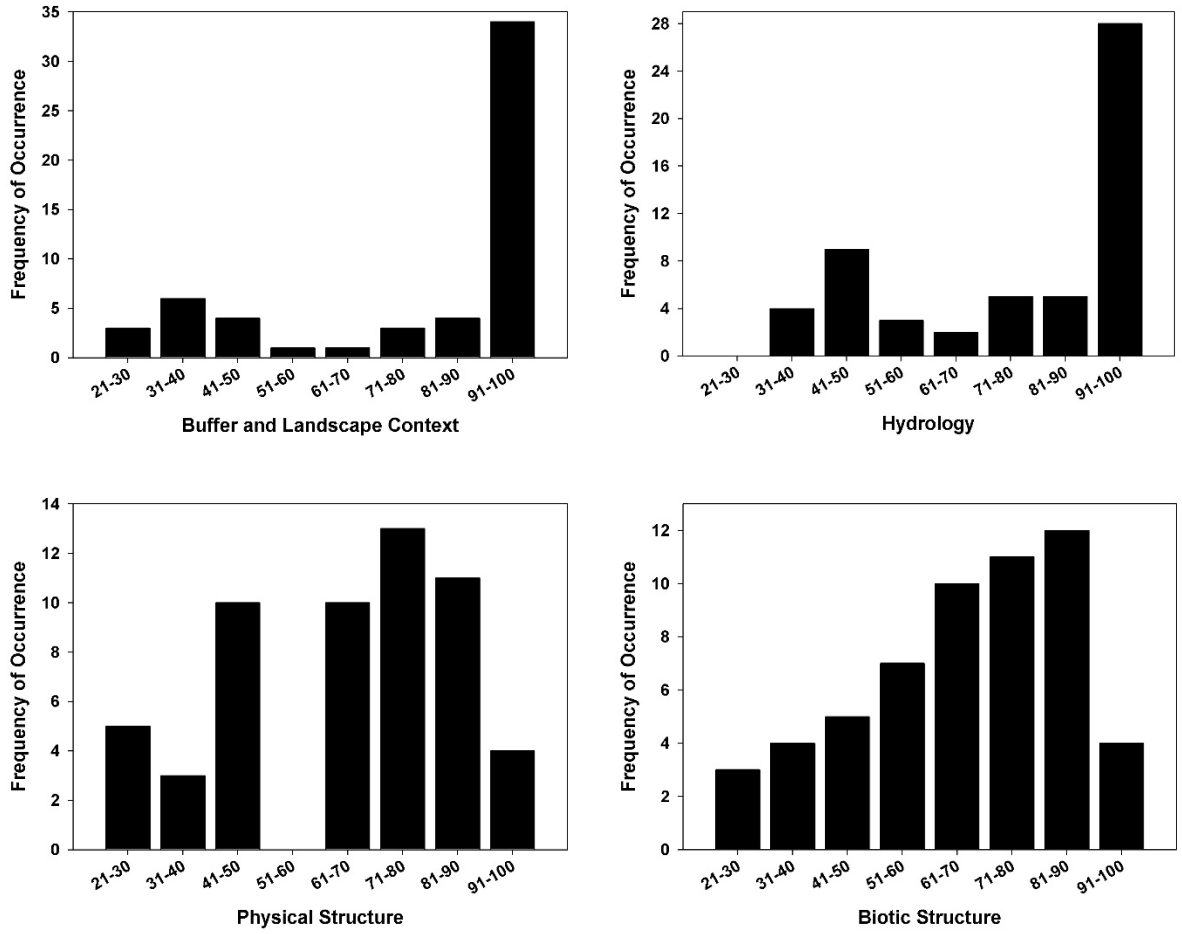


Figure 10. Histograms showing the distribution of data for each CRAM Attribute at all sites (n=56).

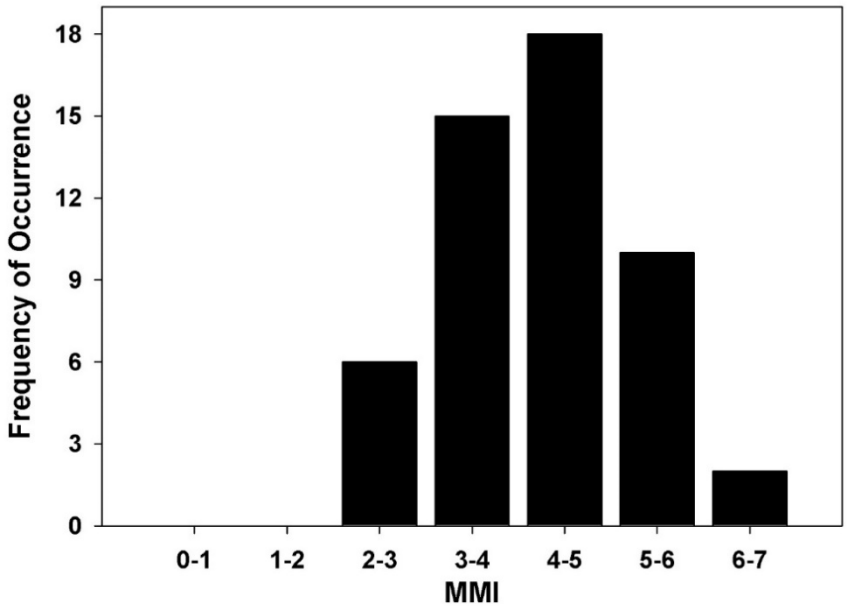


Figure 11. Histogram showing the distribution of data for the MMI at all sites (n=56).

The multimetric Index (MMI) is a composite of several arthropod metrics, and it is designed to discriminate between sites with minimal human activity (i.e., reference sites) and more disturbed sites (see previous section). MMI scores ranged from 2 to 6.5, with a median of 4.5 (Figure 11). The distribution of scores was not skewed (skewness = -0.13).

The overall CRAM Index score and each Attribute score were tested for significant correlations with Level 3 data, including several individual arthropod and bryophyte metrics as well as the MMI. Correlations among CRAM Index and Attribute scores and arthropod metrics are presented in Table 7. For each CRAM parameter the strongest correlation was selected (largest absolute value of the correlation coefficient). Collection protocol for arthropod metrics are noted with a T for "trap" or a V for "vegetation" (see other sections of this report for terrestrial arthropod collection methods). Taxa included within each metric are noted: Co for Coleoptera (beetles); Fo for Formicidae (ants); and Ar for Araneae (spiders). Diversity, abundance, and richness indices were evaluated and identifies as: Di for Shannon Diversity; La for the log of the abundance of individuals of those taxa; and Ri for species richness. The metric most often

correlated with CRAM scores was the T_CoFo_Di: diversity of Coleoptera and Formicidae taken from ground traps. The CRAM Index, Buffer and Landscape Context Attribute, and Physical Structure Attribute all had the strongest correlation with T_CoFo_Di among all of the arthropod metrics (Table 9, Figure 12 to Figure 13). There was only one top correlation with vegetation arthropod sampling. Specifically, the V_FoAr_La (log of the abundance of Formicidae and Araneae) was negatively correlated with the Hydrology Attribute. The Biotic Structure Attribute was correlated with richness of ground trapped Coleoptera, Formicidae, and Araneae species (T_CoFoAr_Ri).

Table 9. Spearman's rank correlations (ρ) among CRAM parameters and various Level 3 independent variables.

CRAM index or attribute score	L3 Metric	Sample Number (N)	Correlation Coefficient	P-Value
Index Score	T_CoFo_Di	56	0.61	<.0001
Buffer and Landscape Context	T_CoFo_Di	56	0.51	<.0001
Hydrology	V_FoAr_La	56	-0.54	<.0001
Physical Structure	T_CoFo_Di	56	0.59	<.0001
Biotic Structure	T_CoFoAr_Ri	56	0.51	<.0001

Figure 12 demonstrates the relationship between the CRAM Index Score and the Shannon Diversity of Coleoptera and Formicidae. Some sites that had relatively high rainfall were included in the analysis but marked separately (open circles). When the analysis was run without these wetter sites there was no statistically significant difference in the correlation.

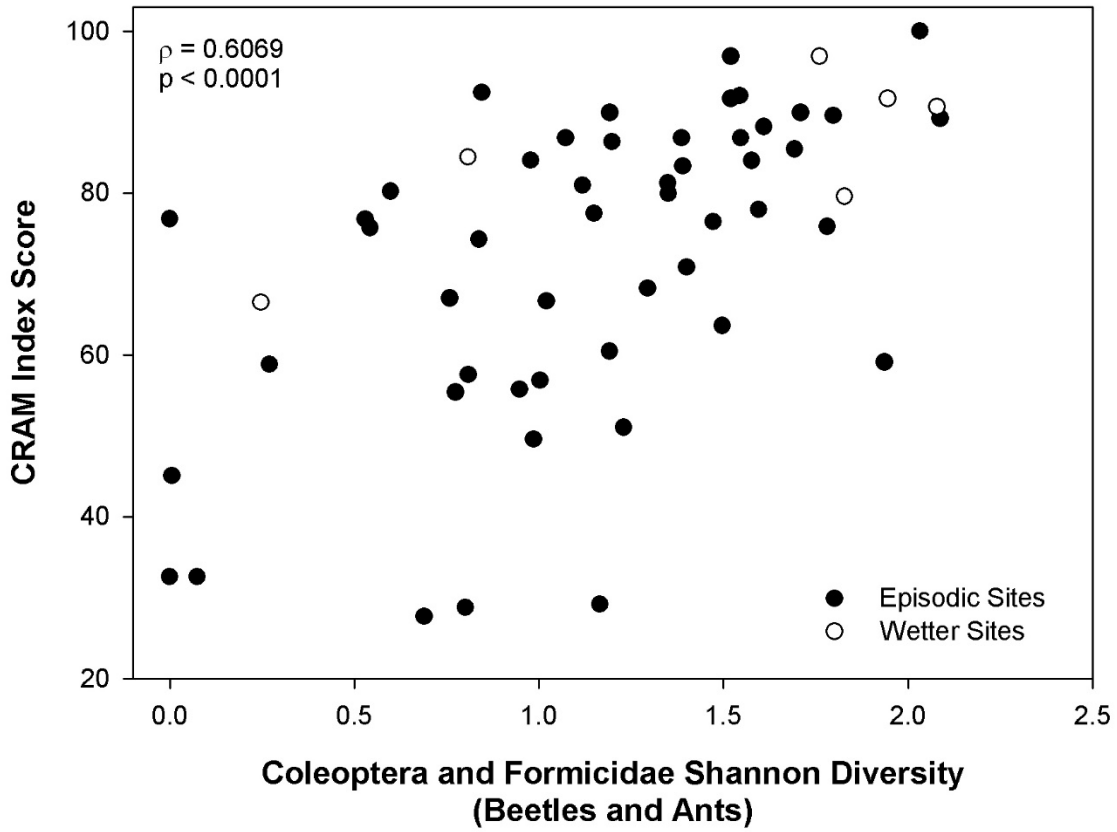


Figure 12. Correlation plot of CRAM Index score vs. T_CoFo_Di.

Figure 13 presents the most signficant relationship among each CRAM Attribute and the arthropod metrics. Again, sites with higher precipitation are noted as open circles. Only the Biotic Structure metric analysis found wetter sites to clustered at the higher end of both CRAM scores and the arthropod metric (species richness of Coleoptera, Formicidae, and Araneae). Most of the Attributes were positively correlated with their top arthropod metric. The Hydrology Attribute, however, was most strongly negatively correlated with the log of abundance of Formicidae and Araneae.

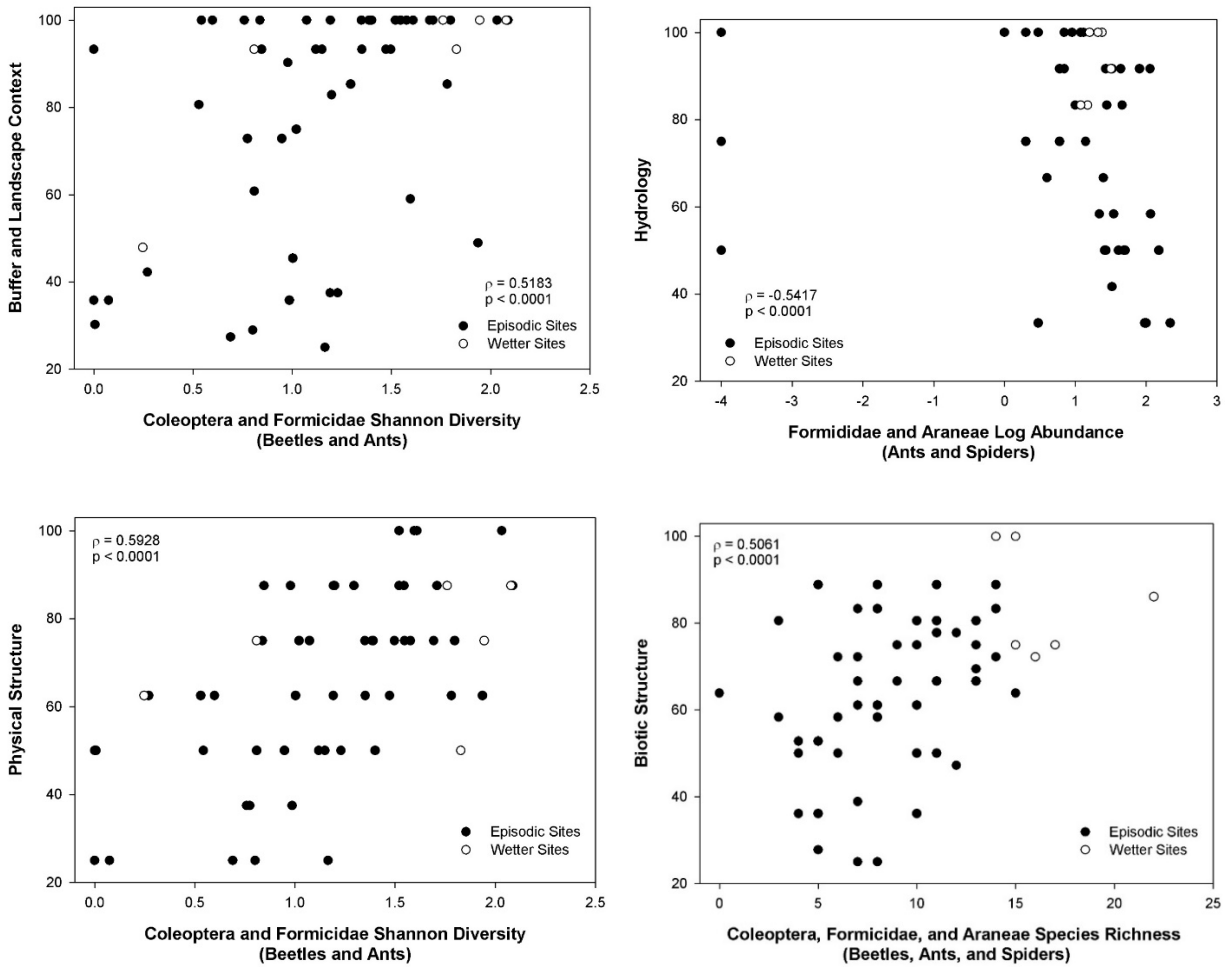


Figure 13. Correlation plot of CRAM Attributes vs. terrestrial arthropod metrics.

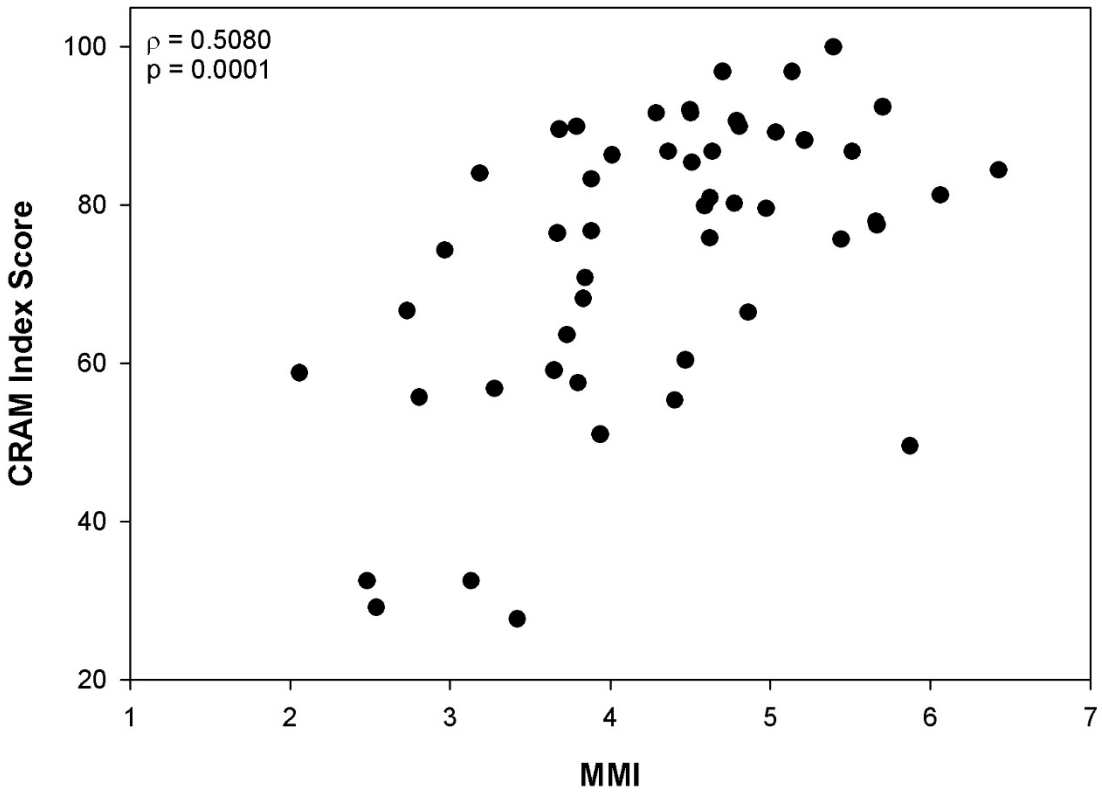
The CRAM scores were tested for correlations with the MMI (Table 10). The CRAM Index score and all of the CRAM Attribute scores were found to be significantly correlated with the MMI. Hydrology had the strongest correlation, while Physical Structure had the weakest correlation with the MMI.

Table 10. Spearman's rank correlations (ρ) for CRAM comparisons to the MMI.

CRAM Parameter	L3 Metric	Sample Number (N)	Correlation Coefficient	P-Value
Index Score	MMI	51	0.51	0.0001
Buffer and Landscape Context	MMI	51	0.41	0.002
Hydrology	MMI	51	0.74	<.0001
Physical Structure	MMI	51	0.31	0.03
Biotic Structure	MMI	51	0.40	0.003

Figure 14 below shows the CRAM Index scores plotted with the MMI scores for each site. The two variables were moderately and significantly correlated. While a regression analysis is not appropriate, as these measures are independent of each other, there is a definite trend of higher CRAM scores along with higher MMI scores. The significant correlation among indices confirms that both methods describe the range of condition among the sample sites while the variability in the correlation suggests (similar to other validation efforts in california) that the two metrics responds differently to site specific variables (Sutula et al. 2006; Stein et al. 2009). Such variablity among Level 2 and Level 3 metrics is anticipated and leads to a robust toolbox of assessment tools.

Figure 14. Correlation plot of CRAM Index vs. MMI.



Each CRAM Attribute was also tested for a significant correlation with the MMI, and all four Attributes were found to be positively correlated (Figure 15). It is evident in these plots that Hydrology has the tightest relationship with the MMI, while the other Attributes have a weaker, but significant relationship.

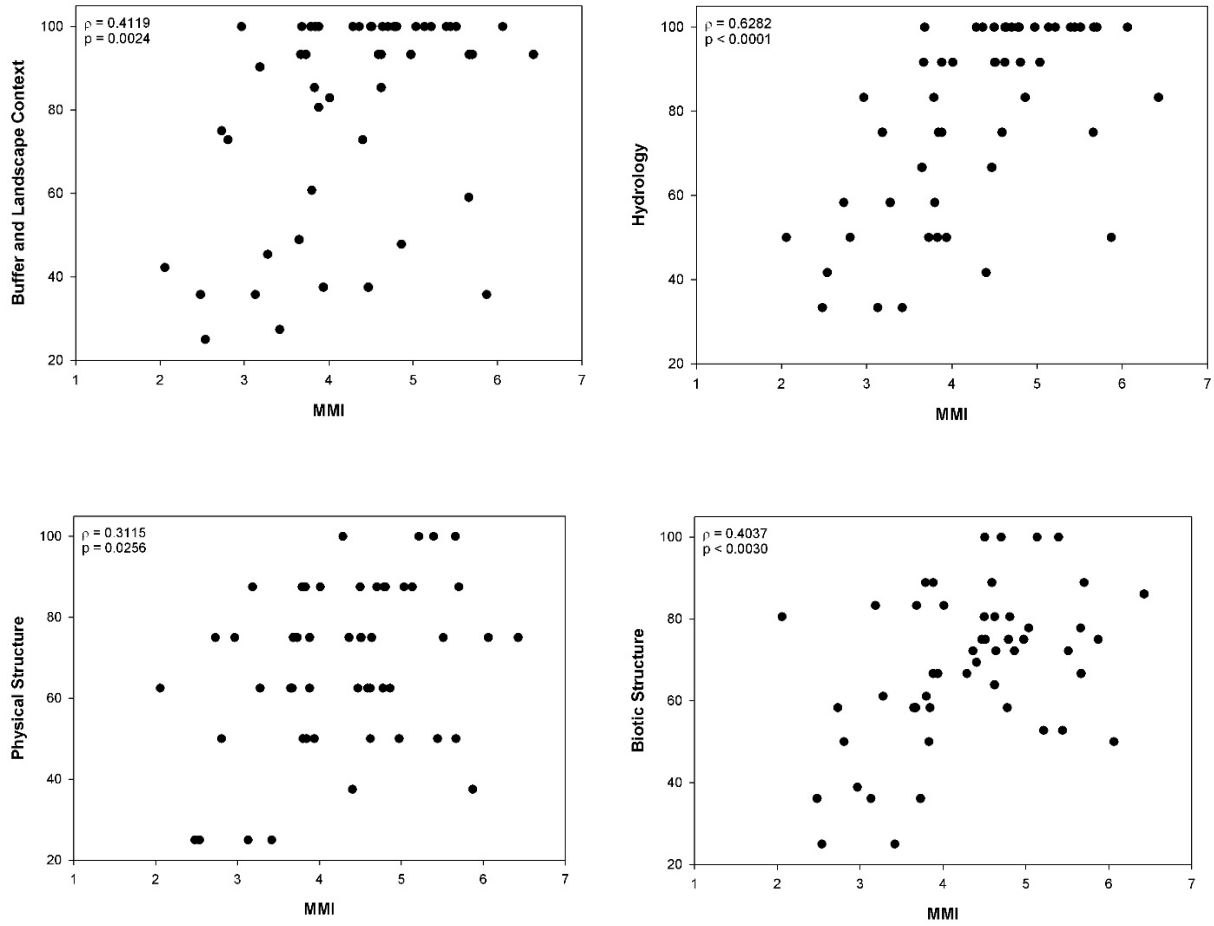


Figure 15. Correlation plots of CRAM Attributes vs. MMI

Correlations between CRAM Index and Attribute scores and watershed land use data (percent impervious, percent urban, and percent agriculture) were tested. The entire upstream watershed was included in these calculations. The percent of watershed that was impervious (i.e., paved or covered with buildings) was the land use metric most closely correlated with CRAM scores (Figure 16, Figure 17). Biotic Structure had the strongest correlation with the percent urban area in the watershed (Figure 17).

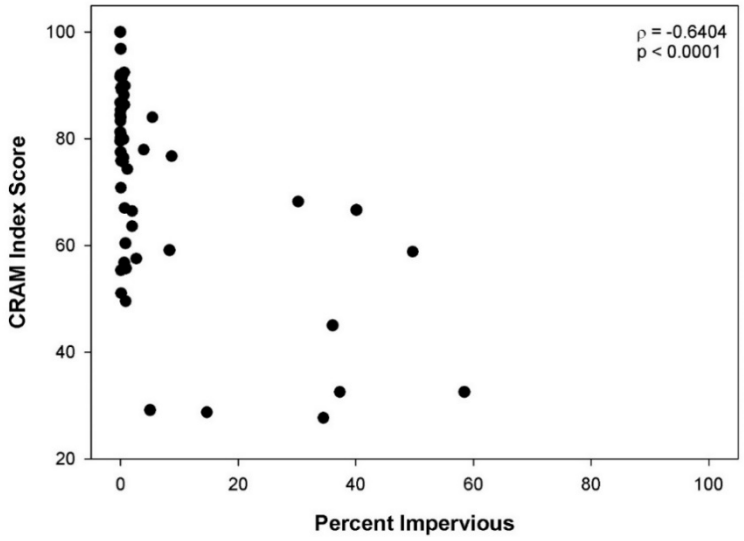


Figure 16. Correlation plot of CRAM Index vs. Percent Watershed Imperviousness.

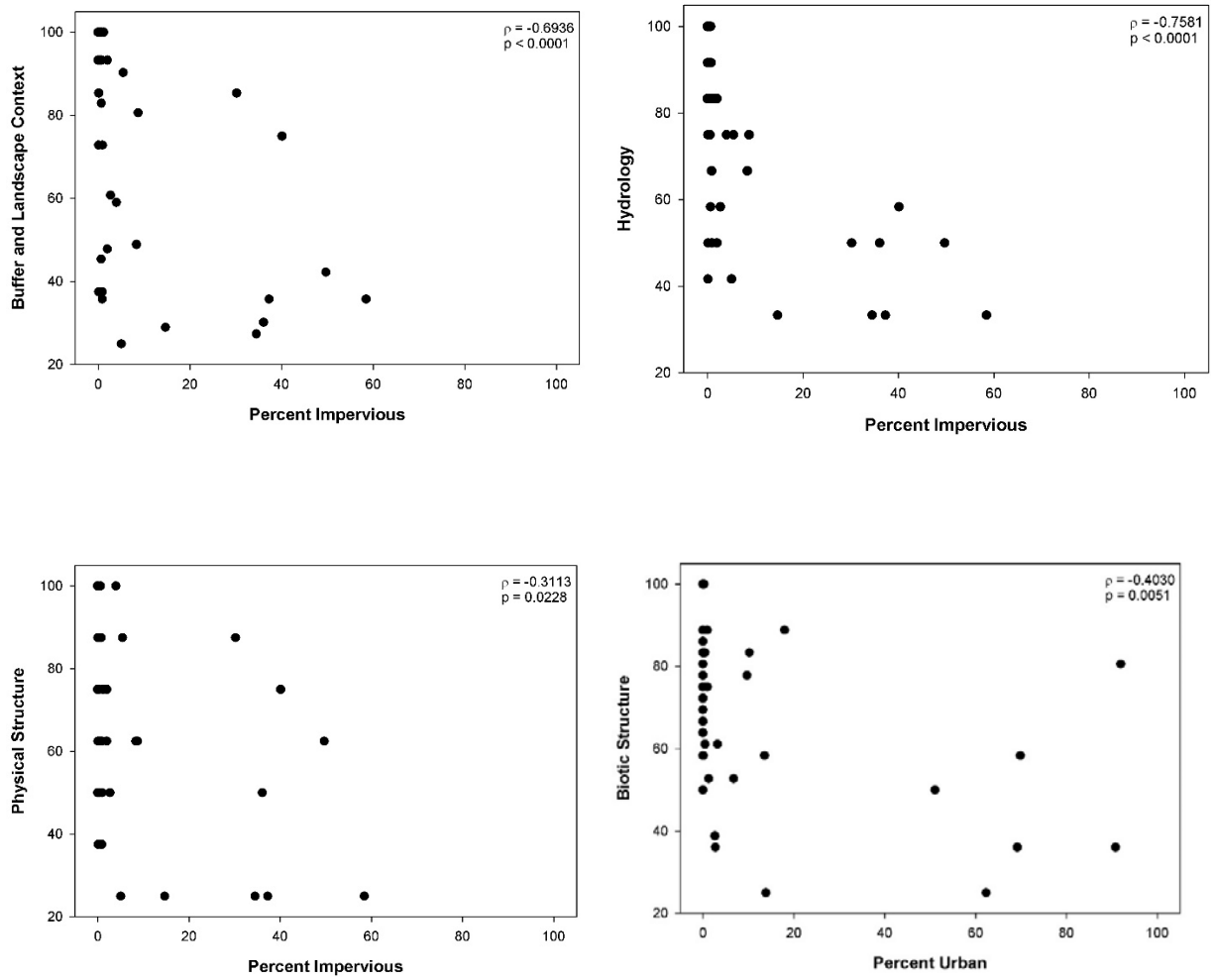


Figure 17. Correlation plots of CRAM Attributes vs. Land use Metrics.

The sites sampled in this study were characterized as reference sites (relatively undisturbed watersheds and local land use), and non-reference sites (more impacted by human activities). We tested whether CRAM scores and land use characteristic were statistically different between reference and non-reference sites. The CRAM Index and Attribute scores between reference and non-reference sites were statistically different (Table 11). The natural variables such as precipitation, elevation, temperature, and watershed area were not significantly different between reference and non-reference sites.

Table 11. Reference vs. Non-reference Site Statistics (significant P-values in bold).

Variable	Test Type	T- or Z-statistic	P-value
CRAM Index	Wilcoxon Two-Sample	Z = 4.2353	**<0.0001**
Buffer and Landscape	Wilcoxon Two-Sample	Z = 4.1603	**<0.0001**
Hydrology	Wilcoxon Two-Sample	Z = 4.3730	**<0.0001**
Physical Structure	Wilcoxon Two-Sample	Z = 2.3810	**0.0086**
Biotic Structure	Student's T	T = -2.41	**0.0198**
Precipitation in Catchment	Student's T	T = -0.45	0.6563
Elevation (natural log transformed)	Student's T	T = -0.64	0.5284
Mean Temperature in Catchment (natural log transformed)	Student's T	T = 0.17	0.8641
Maximum Temperature in Catchment	Wilcoxon Two-Sample	Z = 0.1738	0.8621
Watershed Area (km^2)	Wilcoxon Two-Sample	Z = -0.4981	0.6184

The plots below give a visual representation of the differences in CRAM scores between reference and non-reference sites (Figure 18, Figure 19).

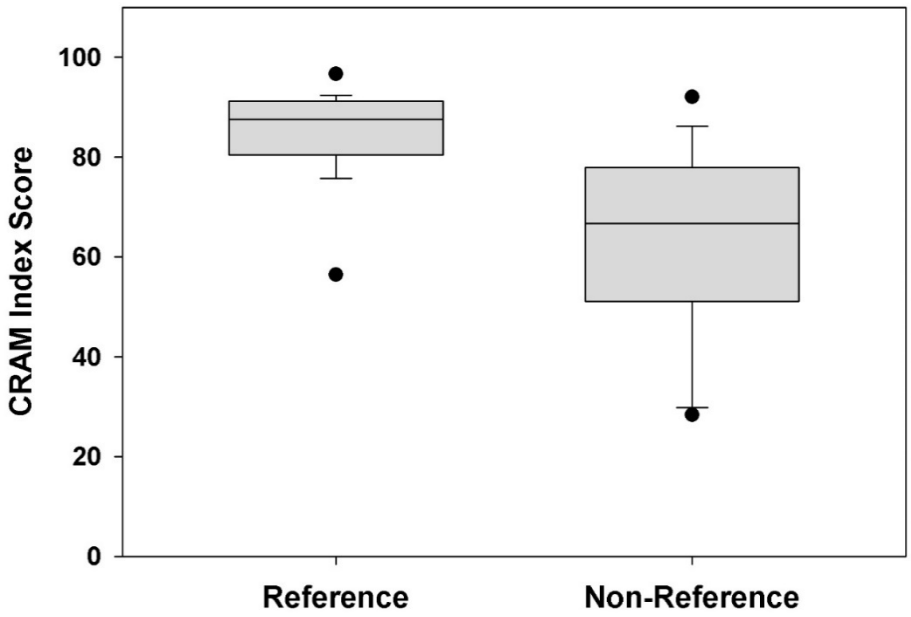


Figure 18. Boxplot of CRAM Index scores for reference and non-reference sites.

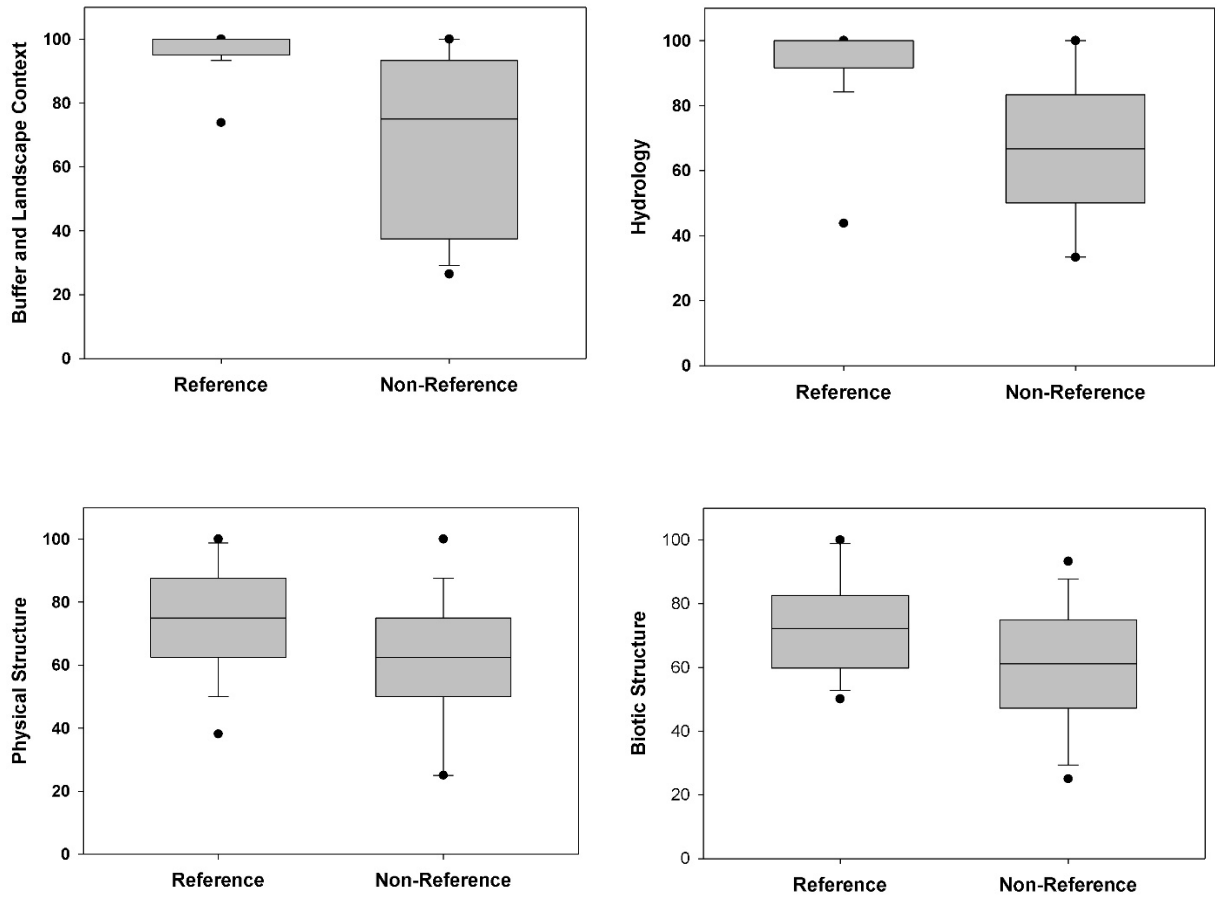


Figure 19. Boxplots of CRAM Attribute scores for reference and non-reference sites.

The environmental variables were not significantly different between reference and non-reference sites, as shown in the plots below (Figure 20).

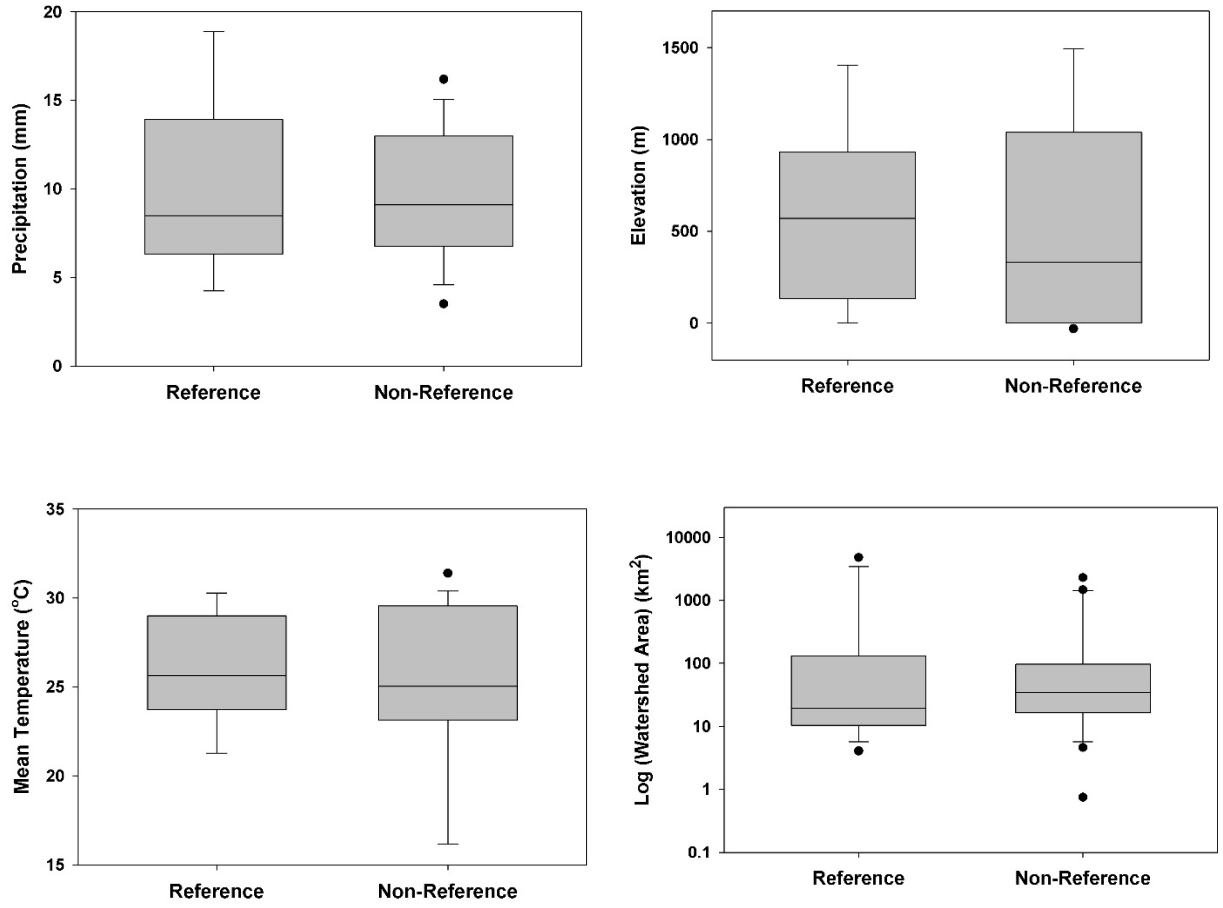


Figure 20. Boxplots of StreamCat variables at the catchment level for reference vs. non-reference sites. Note that these plots show raw data, and some variables were transformed for analysis.

Discussion

The goal of this project was to validate the Episodic Riverine CRAM module. To ensure that the CRAM method meets established CRAM development guidelines (Stein et al. 2009), the CRAM Validation team set out to confirm that a CRAM module for episodic systems met a set of key criteria (i.e., responsiveness to a range of condition, correlate with Level 3 indicators of condition). This validation exercise found the CRAM module to generate scores which appropriately represent the full range of wetland conditions found within the ecoregion. The tool was also found to correlate with other trophic or function specific indicators of condition.

The site selection process ensured that the sampled ephemeral and intermittent streams represented the full range of ecological condition within California and Arizona. The sites in the dataset also represent a range of climatic conditions within the drier areas of the state that contain episodic streams. By partnering with scientists throughout the two states with extensive experience in arid episodic streams, we have developed a tool that can be used successfully by California and Arizona wetland practitioners.

The results showed that the CRAM Index and Attribute scores encompassed the full range of potential conditions in episodic streams. Most of the Attributes had scores ranging from the minimum of 25 to the maximum of 100, except for Hydrology, which had a minimum score of 33. These results confirm that the method is measuring the variation in condition from extremely degraded to a high-functioning least disturbed ecosystem. For the Hydrology Attribute, no site received D scores for all metrics, but each metric displayed the full range of condition across all sites (i.e., for each Hydrology metric there were sites that scored A, B, C, and D).

The Buffer and Landscape Context and Hydrology Attributes were skewed towards higher scores, while Physical and Biotic Structure had more evenly distributed values. The Buffer and Hydrology Attributes are largely influenced by local land use, while scores for the other two Attributes are influenced by climate, geology, and landscape position.

It is useful to develop a conceptual model of what drives condition within the specific wetland class from which to predict and test relationships between CRAM scores and various Level 3 indicators of condition.

Our analysis found that CRAM Index scores were significantly correlated (as expected) with the MMI and the diversity of Coleoptera and Formicidae, as well as several other arthropod metrics. The CRAM Attributes Buffer and Landscape Context and Physical Structure were also significantly correlated with the diversity of Coleoptera and Formicidae. The Biotic Structure Attribute had the strongest correlation with the species richness of Coleoptera, Formicidae, and Araneae. Hydrology had the strongest correlation with the log of the abundance of Formicidae and Araneae, an unexpected negative correlation. However, Hydrology was strongly and positively correlated with the MMI (Figure 15) and the diversity of Coleoptera and Formicidae (Figure 21), so that confirms that the Hydrology attribute is measuring a true difference in overall condition between sites. These correlations make sense because terrestrial arthropods are responsive to disturbance in streams (Mazor et al. 2019a).

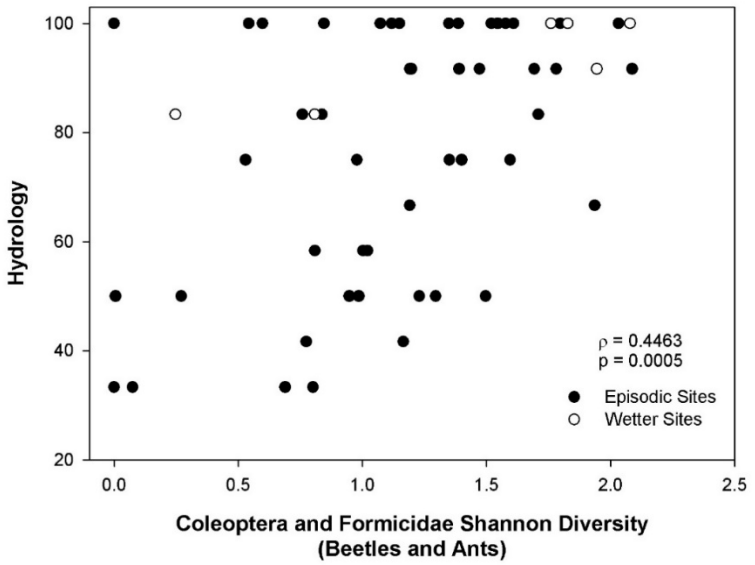


Figure 21. Correlation plot of CRAM Hydrology Attribute vs. T_CoFo_Di.

Since the terrestrial arthropod MMI was developed to respond to disturbance and stress in the stream, CRAM attributes that evaluate similar functions and areas of condition should reflect a similar gradient of impacts. Both the overall Index score and all of the CRAM Attribute scores were correlated with the MMI scores. The Buffer and Landscape Context Attribute measures anthropogenic impacts from surrounding land use. Similarly, the MMI is sensitive to those same impacts. The Hydrology Attribute evaluates the sources of water and potential contamination, the artificial manipulation of water flow, and the connection to adjacent transitional habitat. These factors similarly affect the composition of the terrestrial arthropod communities. The Physical Structure Attribute measures the diversity of topography and structures that create habitat, which facilitate diverse habitat for terrestrial arthropods. The Biotic Structure Attribute looks at the health of the plant community based on structural and species diversity, and the MMI reflects the response of the arthropod communities to these habitat niches.

The goal of this CRAM module validation exercise was to document predicted correlations between multiple L3 metrics and CRAM attributes along a complete gradient of condition which effect ecological functions and services. It is not the intent of this exercise to develop a condition method that responds exactly (i.e., high Spearman's Rho values), to the other independent variables of condition (landscape and arthropods) as this would negate the need for developing a new method of assessment. Specifically, CRAM is meant to integrate multiple wetland functions into a single condition assessment rather than focus on specific functions, as represented by the L3 data. The Spearman's Rho values between CRAM and independent variables (ranging from 0.4 – 0.6) along with significant P-values of less than 0.05 that were found in this study document the predicted correlation between habitat condition and species population dynamics without being redundant.

Conclusions

It is the conclusion of the development team that the goals of validation have been met (Stein et al. 2009) and the Episodic Riverine CRAM module meets the goals defined by the Level 2 Committee for CRAM (CWMW 2011). Our analysis shows a significant correlation between CRAM Index and Attribute scores and Level 3 intensive measures of condition and function. The Episodic Riverine CRAM module provides a meaningful and accurate assessment of wetland condition across the states of California and Arizona.

L1 Assessments: Landscape Models for Estimating the Extent and Condition of Intermittent and Ephemeral Streams

Introduction

Level 1 (L1) assessment tools provide the most rapid ways to assess the condition and extent of aquatic resources across large geographic areas. Currently available resources, such as the National Hydrography Dataset Plus (NHD Plus; McKay et al. 2012) provides a good foundation for conducting L1 assessments, but is challenging to use in assessing ephemeral and intermittent streams due to limitations in the underlying data. First, the NHD Plus does not distinguish between intermittent and ephemeral streams. Second, these designations are notably inaccurate, with error rates exceeding 50% in certain regions (Mazor et al. 2012).

This study presents an opportunity to adapt these tools for L1 assessments in two ways: First, we can develop models to predict appropriate L2 California Rapid Assessment Method (CRAM) module (episodic vs. traditional riverine) and L3 bioassessment tools (terrestrial vs. benthic invertebrates) for NHD Plus stream segments in California and Arizona. This model will not only allow estimation of the extent of episodic streams in both states, it will also provide guidance for practitioners to anticipate the appropriate tools that might be needed prior to conducting field visits. Second, we can evaluate levels of landscape alteration and other disturbances in these episodic streams using thresholds identified in the previous section to determine the extent of likely healthy or impacted episodic streams.

Different habitats require different types of assessment tools, and L1 assessments can create maps that assist practitioners in determining which tools may be appropriate prior to visiting a site in the field. We created models to conduct two different types of L1 assessments: one for the highly ephemeral streams covered by the L2 episodic CRAM module, and one for the dry intermittent and ephemeral streams covered by L3 bioassessment tools based on terrestrial arthropods and bryophytes. Thus, the L1 assessment based on the L3 tools targets a broader set of streams than the L1 assessment based on the L2 tool. Both assessments were conducted in California and Arizona, with unique models built for each state.

Our goals were to use these models to 1) create maps showing the location of likely ephemeral or intermittent streams in California and Arizona, and 2) estimate the likely extent of ephemeral or intermittent streams in good ecological condition. To do this, we first assembled a data set of locations where the appropriate L2 and L3 tools were determined based on field visits. Then, we explored two modeling approaches (generalized linear models and random forest models) to predict the appropriate tool based on landscape or modeled natural hydrology. We selected the best model based on its accuracy. We then applied this model to streams in California and Arizona, and calculated the extent of ephemeral streams (i.e., streams requiring the episodic module) in each state. Finally, we calculated the extent of ephemeral streams where disturbance levels exceeded thresholds associated with poor assessment scores.

Methods

Data aggregation

This analysis requires two types of data: sites where the type of resource (e.g., episodic vs. traditional riverine wetlands) is known, and environmental data that could be used to predict resource type in a statistical model.

Aggregating sites with known appropriate assessment tools

L2 assessment locations

Sites were classified into two categories for L2 assessments: Sites where the episodic CRAM module was required vs. sites where the traditional riverine CRAM module was required. These two categories approximately correspond to ephemeral streams and intermittent/perennial streams, respectively.

In both California and Arizona, sites where the episodic CRAM module had been implemented as part of this study and related studies were used for L1 model development. In California, the eCRAM database provided a large data set of over 2000 sites where the traditional riverine CRAM module had been used. In Arizona, such data were not available; instead, we used site locations where traditional bioassessment sampling for benthic macroinvertebrates had occurred under State and National assessment programs, assuming that the traditional riverine CRAM module would be appropriate at these locations. Sites that were not located on NHD Plus flowlines were excluded from analysis, and sites of the same sample type located on the same flowline were treated as a single observation.

L3 assessment locations

Sites were classified into three categories for L3 assessments: Sites where aquatic indicators (such as BMI) were appropriate (which approximately corresponds to perennial streams), sites where terrestrial indicators were appropriate (which approximately corresponds to ephemeral streams and short-duration intermittent streams), and sites where both indicators are appropriate (which approximately corresponds to long-duration intermittent streams).

In both California and Arizona, sites where sampling for terrestrial arthropods had occurred as part of this study and related studies were used in model development. In addition, we queried state databases for bioassessment data to determine where BMI sampling had occurred in the past. In California, site evaluation data from probabilistic surveys was also reviewed; these data indicate when sites were deemed appropriate for BMI sampling, and when sites were too dry for these aquatic indicators. Catchments containing exclusively BMI samples were classified as appropriate for aquatic indicator-based L3 assessment tools; catchments containing exclusively terrestrial arthropod samples were classified as appropriate for “dry” L3 assessment tools; and catchments containing both sample types were classified as appropriate for both types of L3 assessment tools. Sites that were not located on NHD Plus flowlines were excluded from analysis, and sites of the same sample type located on the same flowline were treated as a single observation.

Predictor data

To acquire environmental predictor data, we first snapped each location to the closest stream segment in the NHD plus and determined its COMID (i.e., the unique identifier of each segment). Based on the COMID, we could then associate data points with other data sets.

We used two sources of environmental data to predict the appropriate CRAM module at each site: 1) landscape characteristics of site watersheds and local catchments, and 2) natural modeled hydrology. Landscape metrics were derived from the StreamCat dataset (Hill et al. 2016), which covers nearly every stream segment in the NHD Plus. We selected metrics likely to relate to stream hydrology (Table 10), including landcover, geology, climate, and estimated runoff. In California, we supplemented StreamCat with additional data sources, including habitat information from the USGS GAP/LANDFIRE Land Cover data set, and plant community classifications from the Revised Hierarchy for Natural Vegetation Classification and Standard (USNVC). In addition, we supplemented California data with predicted natural flow classes from the California eFlows database (Lane et al. 2020), and the USGS database of modeled natural monthly flows (Miller et al. 2018). For the latter, we calculated the 10^{th} percentile of annual natural flows for 1985 to 2015 to estimate the magnitude of low flows. A total of 103 predictors were considered for analysis (Table 12).

Table 12. Environmental predictors evaluated in the models. Asterisks indicate that the variable was evaluated at both the watershed and catchment scales.

Predictor code	**Full Name of Predictor**	**Predictor Source** (table)
Al2O3*	Mean aluminum oxide	StreamCat (GeoChemPhys)
BFI*	Base Flow Index	StreamCat (BFI)
CaO*	Mean calcium oxide	StreamCat (GeoChemPhys)
CCHEM_v2	Catchment chemistry	StreamCat (ICI_IWI_v2)
CCONN_v2	Catchment connectivity	StreamCat (ICI_IWI_v2)
CHABT_v2	Catchment habitat provision	StreamCat (ICI_IWI_v2)
CHYD_v2	Catchment hydrologic regulation component score	StreamCat (ICI_IWI_v2)
Clay*	Clay soil raster units (%)	StreamCat (STATSGO)
CompStrgth*	Compressive strength (%)	StreamCat (GeoChemPhys)
CSED_v2	Catchment sediment regulation	StreamCat (ICI_IWI_v2)
CTEMP_v2	Catchment temperature regulation	StreamCat (ICI_IWI_v2)
Elev*	Mean elevation	StreamCat (Elevation)
Fe2O3*	Mean iron oxide	StreamCat (GeoChemPhys)
HydrlCond*	Hydrologic conductivity (%)	StreamCat (GeoChemPhys)
ICI_v2	Index of catchment integrity	StreamCat (ICI_IWI_v2)
IWI_v2	Index of watershed integrity	StreamCat (ICI_IWI_v2)
K2O*	Mean potassium oxide	StreamCat (GeoChemPhys)
MgO*	Mean magnesium oxide	StreamCat (GeoChemPhys)
N*	Mean nitrogen	StreamCat (GeoChemPhys)
Na2O*	Mean sodium dioxide	StreamCat (GeoChemPhys)
Om*	Organic Matter raster units (%)	StreamCat (STATSGO)

P2O5*	Mean phosphorous pentoxide	StreamCat (GeoChemPhys)
PctBl2011*	Bedrock (%)	Stream Cat (NLCD)
PctConif2011*	Coniferous/evergreen forest (%)	Stream Cat (NLCD)
PctCrop2011*	Row crop (%)	Stream Cat (NLCD)
PctDecid2011*	Deciduous forest (%)	Stream Cat (NLCD)
PctGrs2011*	Grass/herbaceous (%)	Stream Cat (NLCD)
PctHay2011*	Pasture/hay (%)	Stream Cat (NLCD)
PctHbWet2011*	Herbaceous wetland (%)	Stream Cat (NLCD)
PctIce2011*	Ice/snow (%)	Stream Cat (NLCD)
PctMxFst2011*	Mixed forest (%)	Stream Cat (NLCD)
PctOw2011*	Open water (%)	Stream Cat (NLCD)
PctShrb2011*	Shrub/scrub (%)	Stream Cat (NLCD)
PctUrbHi2011*	Developed, high intensity (%)	Stream Cat (NLCD)
PctUrbLo2011*	Developed, low intensity (%)	Stream Cat (NLCD)
PctUrbMd2011*	Developed, medium intensity (%)	Stream Cat (NLCD)
PctUrbOp2011*	Developed, open (%)	Stream Cat (NLCD)
PctWdWet2011*	Woody wetland (%)	Stream Cat (NLCD)
Perm*	Permeability units (cm/hr)	StreamCat (STATSGO)
Precip8110*	Average annual normal precipitation from 1981-2010 (mm)	StreamCat (PRISM)
RckDep*	Depth of bedrock soils (cm)	StreamCat (STATSGO)
Runoff*	Mean runoff (mm)	StreamCat (Runoff)
S*	Mean sulfur	StreamCat (GeoChemPhys)
Sand*	Sand raster units (%)	StreamCat (STATSGO)
SiO2*	Mean silicon dioxide	StreamCat (GeoChemPhys)
Tmax8110*	Average maximum air temperature from 1981-2010 (°C)	StreamCat (PRISM)
Tmean8110*	Average mean air temperature from 1981-2010 (°C)	StreamCat (PRISM)
Tmin8110*	Average minimum air temperature from 1981-2010 (°C)	StreamCat (PRISM)
WCHEM_v2	Watershed chemistry	StreamCat (ICI_IWI_v2)
WCONN_v2	Watershed connectivity	StreamCat (ICI_IWI_v2)
WHABT_v2	Watershed habitat provision	StreamCat (ICI_IWI_v2)
WHYD_v2	Watershed hydrologic regulation component score	StreamCat (ICI_IWI_v2)
WSED_v2	Watershed sediment regulation	StreamCat (ICI_IWI_v2)
WtDep*	Water table depth of soil (cm)	StreamCat (STATSGO)
WTEMP_v2	Watershed temperature regulation	StreamCat (ICI_IWI_v2)
California models only		
Habitat	Habitat type (level 3 formation)	Gap-USGS
meanp10Q	Mean natural monthly flow at the 10th percentile from 1985 to 2015	NaturalMonthlyFlow-USGS

Natural flow regime class	Natural flow regime class	Lane et al. (2020)

Development of random forest models to predict stream type

We used random forest models to predict appropriate L2 and L3 assessment tools. Random forest is a machine learning approach based on creation of a large number of decision-trees that split data based on dichotomous decisions. Random forest is popular in environmental science due to its handling of non-linear relationships and complicated interactions among variables.

We determined predictors through a simple stepwise process. First, we ran all the predictors (except for natural monthly flows, eFlow hydrologic regime classes, and NLCD ice land cover, due to incomplete data coverage) through an "unbalanced" random forest model with 1,500 trees (unbalanced because one class, such as riverine CRAM assessments, were more heavily represented in the calibration data than other classes). Next, we repeated this process with stratified samples to balance the data and select the best stratification (one that would produce the least amount of out of bag (OOB) error and class error). For the L2 models in both states, we selected an episodic stratum size of 20 samples and a riverine stratum size of 40 samples due to the relatively low number of episodic sites. For the L3 models, we selected a stratum size of 22 for each of the three classes in Arizona, and 150 for California. We then reviewed the partial dependence plots for all the predictors and removed any land use predictors with less than 5% cover. Finally, we removed negative importance predictors and ran a balanced random forest model with 1500 trees to obtain our final model.

We decided on this process to improve the OOB and class errors even though random forest models with large numbers of environmental predictor variables tend to perform just as well as pre-selected variables or stepwise process for selecting predictors (Fox et al. 2017), and are robust to overfitting. For each model, we calculated accuracy as the out-of-bag error rate, as well as Cohen's Kappa statistic (which accounts for the likelihood of getting correct classifications by random chance). We also calculated within-class error rates for each model, and summarized the top 5 most important variables.

Application of models

One of the primary goals of L1 assessments is to estimate the extent of aquatic resources in a region. To achieve this goal, we applied the models described above to every stream in California and Arizona included in the StreamCat dataset (Hill et al. 2016). This application results in the classification of all streams as episodic or riverine (for L2 assessments); or as requiring terrestrial, aquatic, or both types of indicators (for L3 assessments).

Predicting ecological condition of ephemeral and episodic catchments

In addition to estimating the extent of aquatic resources, a second major goal of L1 assessments is to estimate the ecological condition of these resources. In order to do so, we identified levels of watershed alteration associated with high and low likelihood of good conditions, and used these as thresholds to classify streams that were previously predicted to be ephemeral or episodic, or to require terrestrial indicators. Thresholds were determined by evaluating the distribution of condition scores (CRAM scores for L2 condition, and bioassessment index scores for L3 condition) at reference sites (Table 13).

Table 13. Thresholds for identifying dry streams in good, intermediate, or poor conditions.

Conditions	L2 scores	L3 scores
Good	Above 71.9	Above 3.81
Intermediate	Between 52.6 and 71.9	Between 3.26 and 3.81
Poor	Below 52.6	Below 3.26

We then used quantile regression to predict the median L2 and L3 score based on the percent of developed land in the contributing watershed (derived from StreamCat; Hill et al. 2016), providing thresholds for percent development where good, intermediate, and poor conditions are likely. Quantile regression is appropriate for exploring ecological relationships where stressors are predicted to limit biological condition. We applied these landscape thresholds to all episodic stream segments to predict L2 condition, and to all intermittent or ephemeral stream segments to predict L3 condition. We then tabulated the extent of stream miles in each condition class for each state.

Results

Data aggregation

For the L2 models, a total of 34 episodic and 2035 riverine sites were identified in California. Much smaller numbers were identified in Arizona, where we found 22 episodic sites and 213 presumed riverine sites (Figure 22).

For the L3 models, we had 2939 locations were aquatic indicators were appropriate, 7095 sites where terrestrial indicators were appropriate (largely, sites that were too dry for traditional bioassessment sampling in probabilistic surveys), and 400 sites where both indicators were appropriate. In Arizona, we had 213 sites where aquatic indicators were appropriate, 24 where terrestrial indicators were appropriate, and 28 where both types of indicators were appropriate (Figure 23).

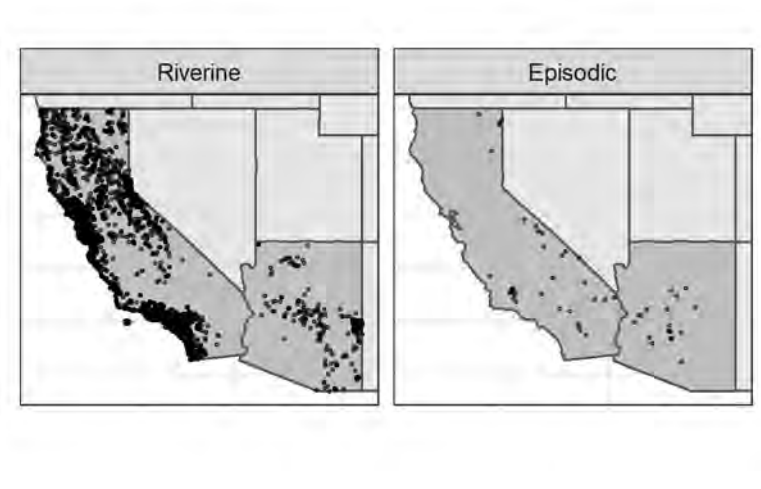


Figure 22. L2 assessment site locations used to calibrate L1 models in California and Arizona.

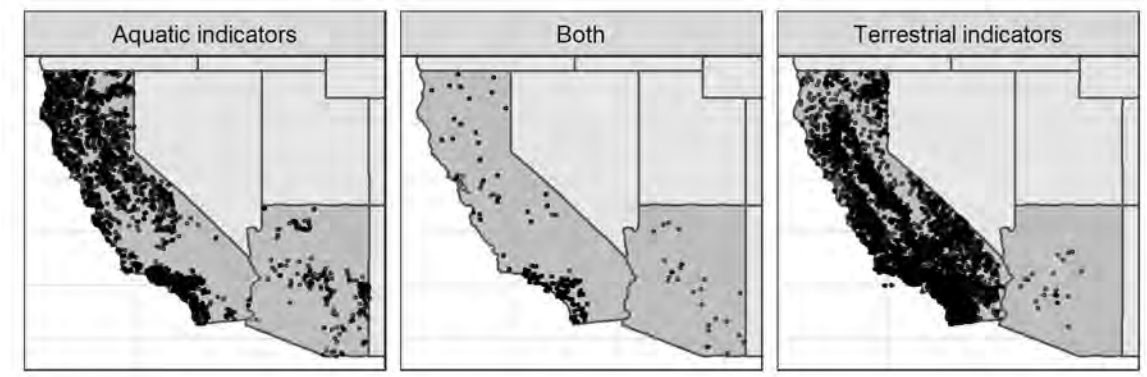


Figure 23. L3 assessment site locations used to calibrate L1 models in California and Arizona.

Model Characteristics

Overall the models for California were more accurate than the models for Arizona, likely due to the much larger number of data points used to calibrate the models (Table 14). In fact, Cohen's Kappa statistics for both Arizona models showed that model predictions agreed with the true classes only slightly better than chance. In contrast, the California models predicted the true class somewhat (for the L3 models) or substantially (for the L2 models) better than chance. For both states, models were more successful at predicting the wetter habitats (i.e., riverine for the L2 models, and aquatic indicators for the L3 models). Both California models relied more extensively on predictors based on precipitation, whereas the Arizona models made more use of temperature-related predictors.

Table 14. Summary of models to predict appropriate L2 and L3 tools. Overall accuracy is measured as the out-of-bag error rate (lower is better), as well as Cohen's Kappa statistic (higher is better). Top five predictors in each model are indicated in the final column. Predictors followed by (cat) are calculated at the catchment scale in StreamCat (Hill et al. 2016); all other predictors are calculated at the watershed scale.

	Accuracy					
	Overall	Kappa	Class-wise error			Top predictors
L2 models			*Episodic*	*Riverine*		
California	1.79	0.74	20.6	1.5		Precipitation, precipitation (cat), Habitat, % bedrock, organic matter content in soil (cat)
Arizona	5.96	0.14	31.8	3.3		Elevation, mean temperature, max temperature, min temperature, mean temperature (cat)
L3 models			*Aquatic*	*Both*	*Terrestrial*	
California	27.7	0.48	26.7	48.3	27.5	Precipitation, % coniferous forest, watershed area, max temperature, precipitation (cat)
Arizona	21.3	0.13	19.4	25.0	33.3	Base flow index (cat), elevation, base flow index, min temperature, mean temperature

Application of models

Despite the relatively poor accuracy of some of the models, they nonetheless produced maps that were consistent with expectations. For example, in California, episodic streams dominated the Sonoran and Mojave Deserts, as well as portions of the San Joaquin Valley and the Modoc Plateau in the northeastern part of the state (Figure 24). In Arizona, much of the southwestern part of the state was predicted to be episodic, with the exception of the mainstem of the Gila River (Figure 25). The predominance of riverine streams in eastern Arizona likely reflects the lack of such streams in that region in the calibration data set (Figure 22).

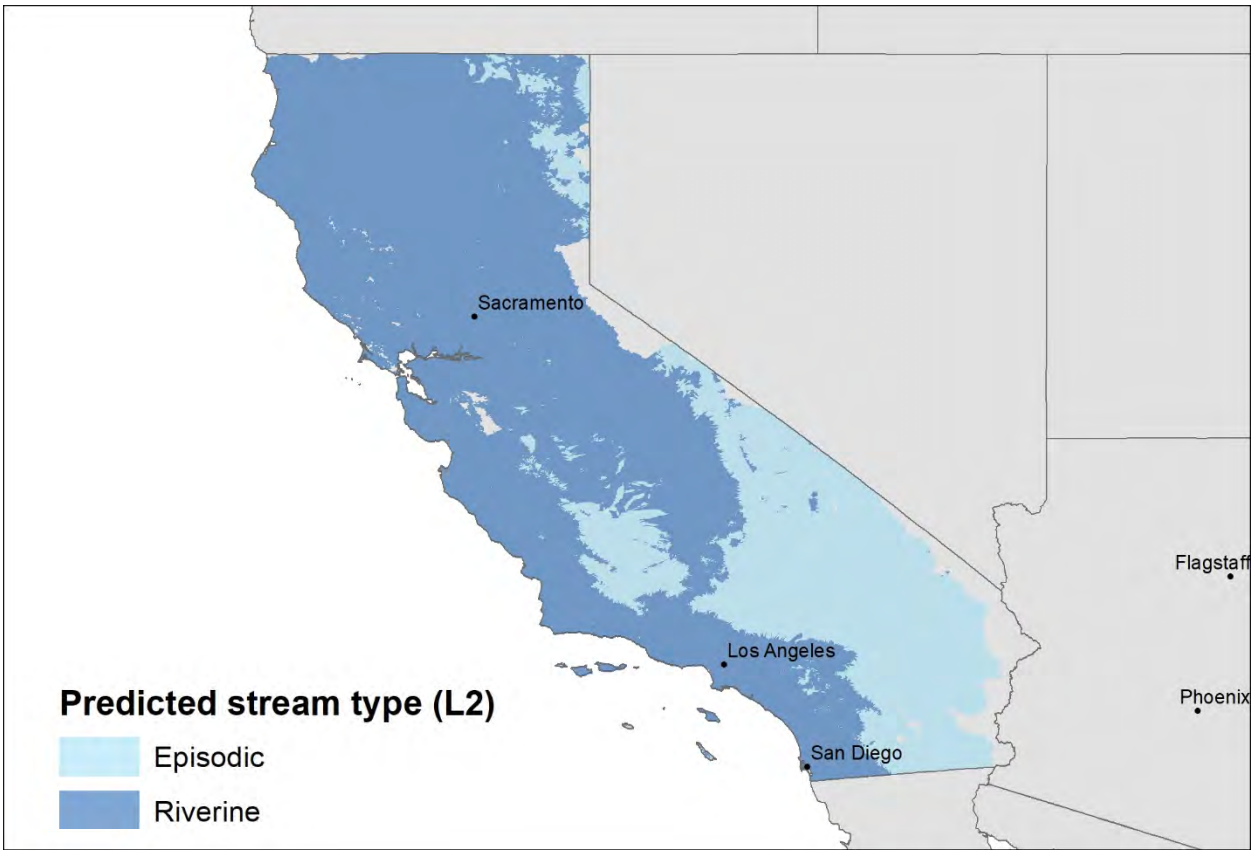


Figure 24. Predicted L2 assessment tool (episodic vs. riverine) for catchments in California.

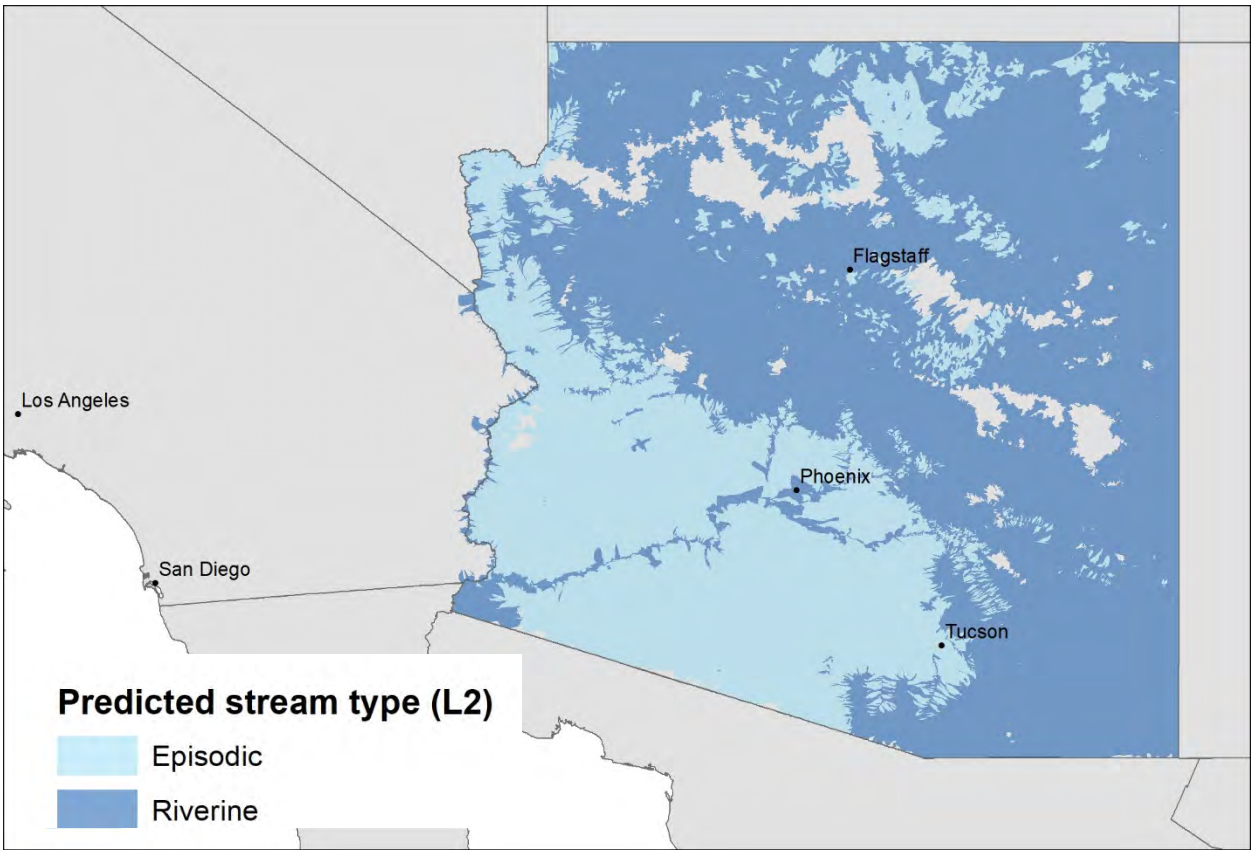


Figure 25. Predicted L2 assessment tool (episodic vs. riverine) for catchments in Arizona.

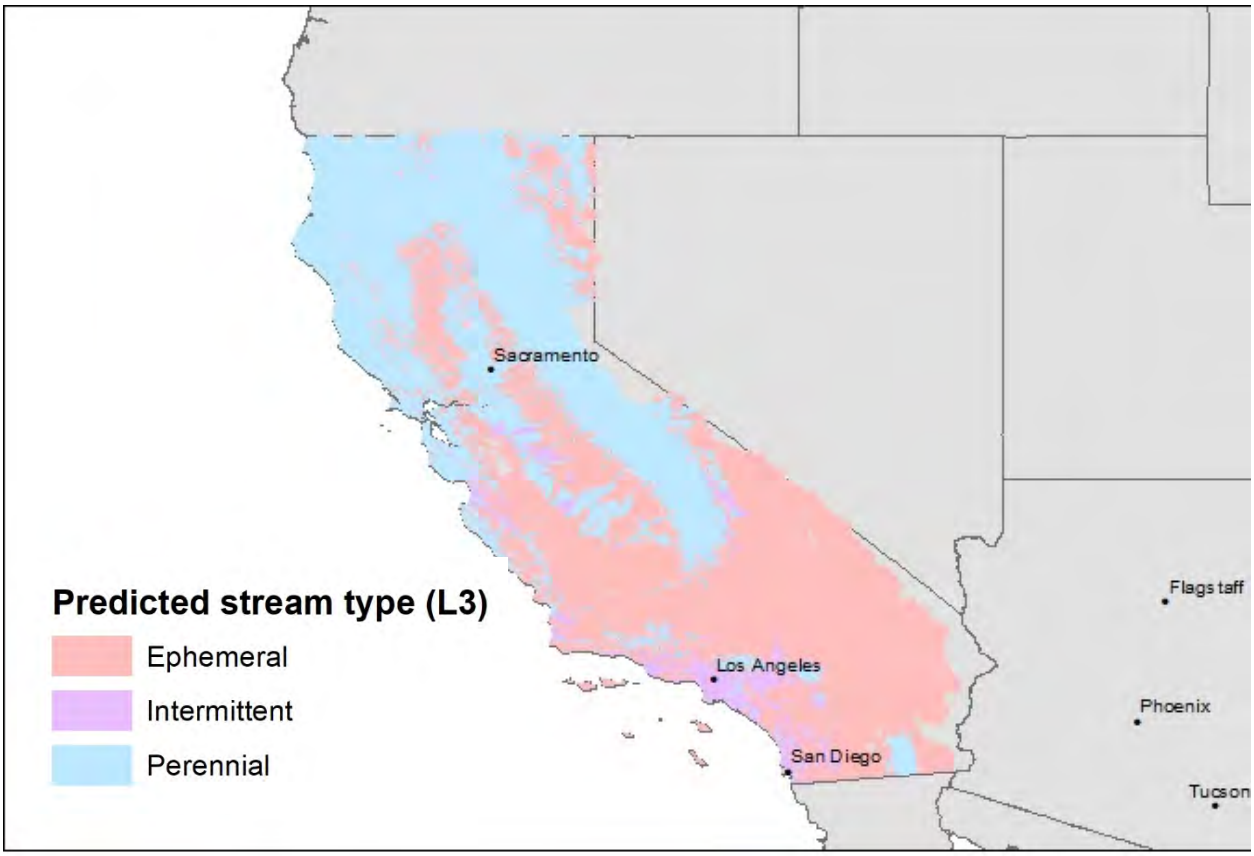


Figure 26. Predicted L3 assessment tool (aquatic vs. terrestrial vs. both) for catchments in California. Streams where terrestrial indicators are recommended are assumed to be ephemeral or short-term intermittent; streams where aquatic indicators are recommended are assumed to be perennial streams; and streams where both are recommended are assumed to be long-term intermittent.

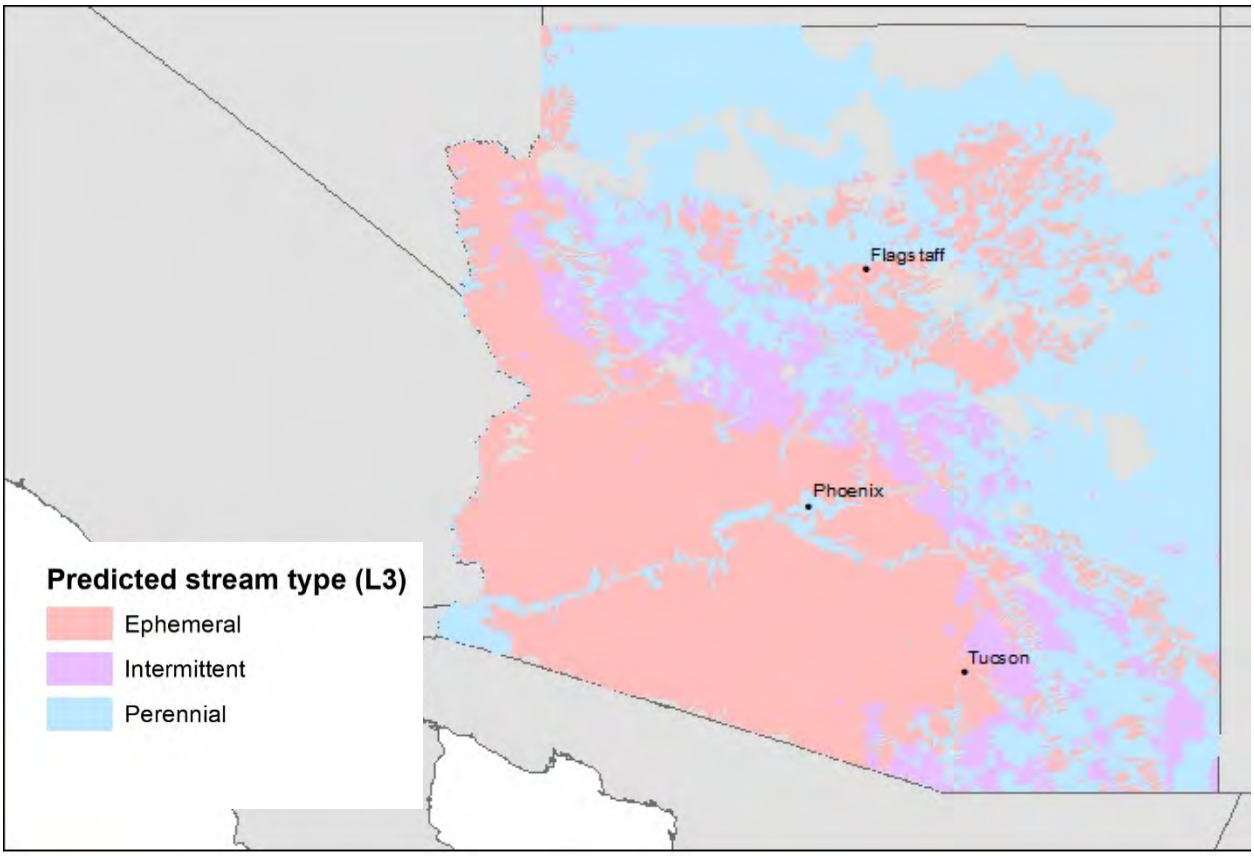


Figure 27. Predicted L3 assessment tool (aquatic vs. terrestrial vs. both) for catchments in Arizona. Streams where terrestrial indicators are recommended are assumed to be ephemeral or short-term intermittent; streams where aquatic indicators are recommended are assumed to be perennial streams; and streams where both are recommended are assumed to be long-term intermittent.

Predicted conditions of episodic and ephemeral catchments

Quantile regressions showed a stronger relationship between watershed development and L3 scores than L2 scores. Although both showed negative relationships, there was considerably more variability in L2 scores, and the regression model had much weaker significance than the L2 model (i.e., $p < 0.1$ vs. $p < 0.01$; Table 15). Notably, a few sites with highly developed watersheds had very high L2 scores. However, thresholds for landscape derived from these models were not terribly different (Table 16); that is, for both indices, good conditions were most likely when less than ~30% of the watershed was developed, and poor conditions were most likely when more than ~50% of the watershed was developed.

Application of these thresholds to non-perennial catchments predicted by the models described above showed that many streams in both states are likely to be in good condition (Figure 28). However, predictions of poor conditions were clustered in urban areas near Los Angeles, Phoenix, and San Diego, as well as in the southern San Joaquin valley of California. L3 assessments showed that good conditions were likely in more than 95% of intermittent and

ephemeral streams in Arizona, and over 93% of ephemeral streams in California (Table 17). However, the extent of good conditions for intermittent streams in California was more limited, likely in just more than half of such streams; moreover, poor conditions were likely in over 14% of intermittent streams in California.

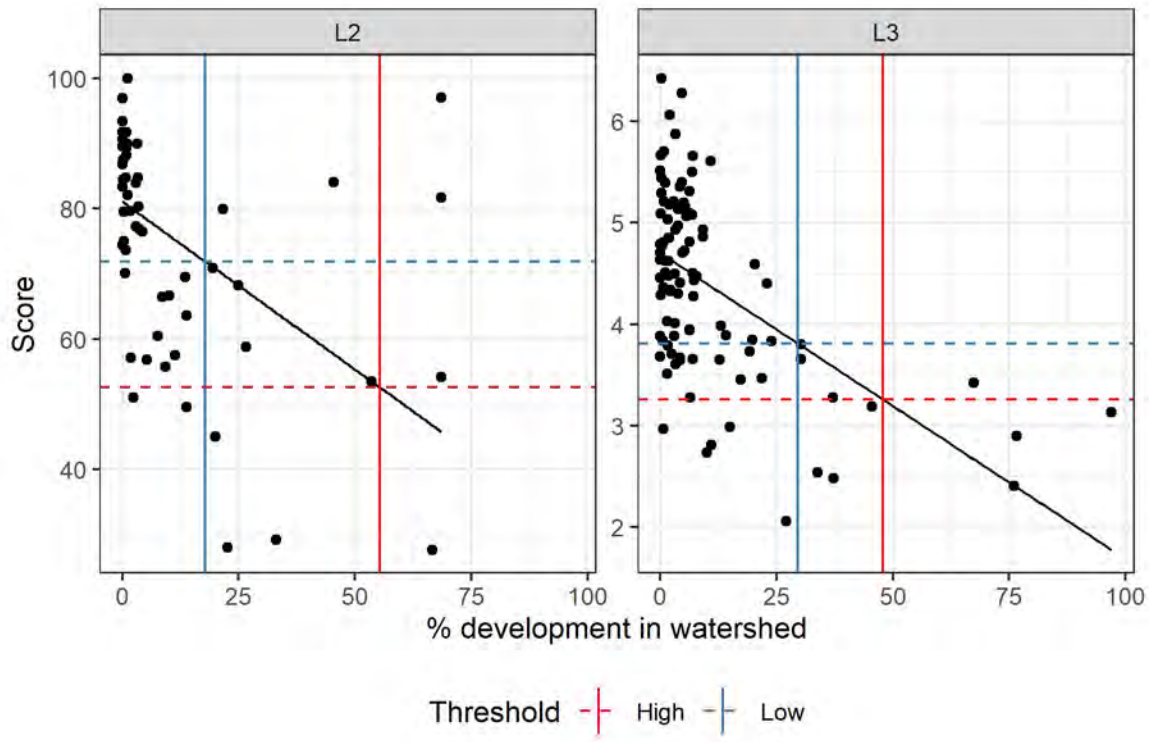


Figure 28. Quantile regressions of L2 and L3 scores (specifically, episodic CRAM scores, and MMI scores, respectively) against percent development in the watershed. The solid black line represents the quantile regression of the median score. The horizontal dashed lines represent the index score thresholds based on the 10th and 1st percentile of reference site scores (blue and red, respectively). The solid vertical lines represent the levels of development where good (blue) or intermediate (red) conditions are as likely as worse conditions.

Table 15. Summary of quantile regressions of index scores against percent development in the watershed.

Index	Value	Value	Standard error	t-value	p-value
L2	Intercept	81.0	2.792	29.0	0
	% development	-0.515	0.286	-1.8	0.078
L3	Intercept	4.70	0.172	27.4	0
	% development	-0.030	0.009	-3.2	0.002

Table 16. Maximum levels of % development in the watershed used to predict likely condition class for L2 and L3 assessments.

Index	Good conditions	Intermediate conditions
L2	< 17.8	< 55.3
L3	< 29.6	< 47.8

Table 17. Extent of non-perennial streams predicted to be in good, intermediate, or poor condition in California and Arizona, based on the episodic CRAM module (for L2 assessments) or terrestrial phase indicators (for L3 assessments).

		Percent of stream-length in:		
		Good condition	Intermediate condition	Poor condition
L2 assessments				
California	Episodic	84.8%	11.3%	3.9%
Arizona		75.5%	23.4%	1.1%
L3 assessments				
California	Ephemeral	93.4%	4.6%	2.1%
	Intermittent	53.8%	32.2%	14.1%
Arizona	Ephemeral	97.6%	2.0%	0.4%
	Intermittent	99.6%	0.4%	0.03%

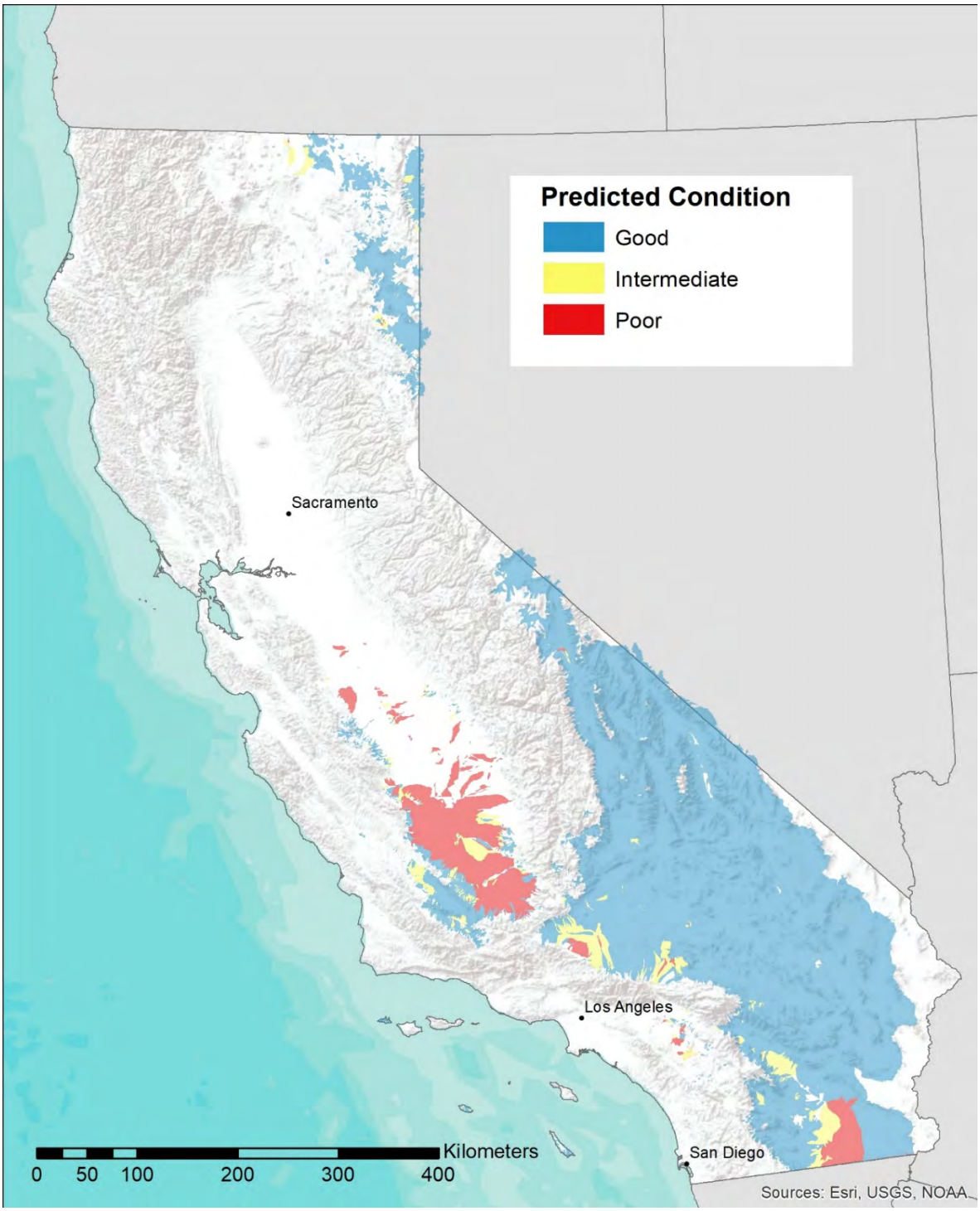


Figure 29. Estimated stress levels on California episodic catchments based on the total percent of watershed NLCD covers: urban (high, medium, low, open) and agriculture (crop, hay). Good condition: < 3.6% development in the watershed. Intermediate condition: 3.6 to 17.4% development in the watershed. Poor condition: > 17.4% development in the watershed.

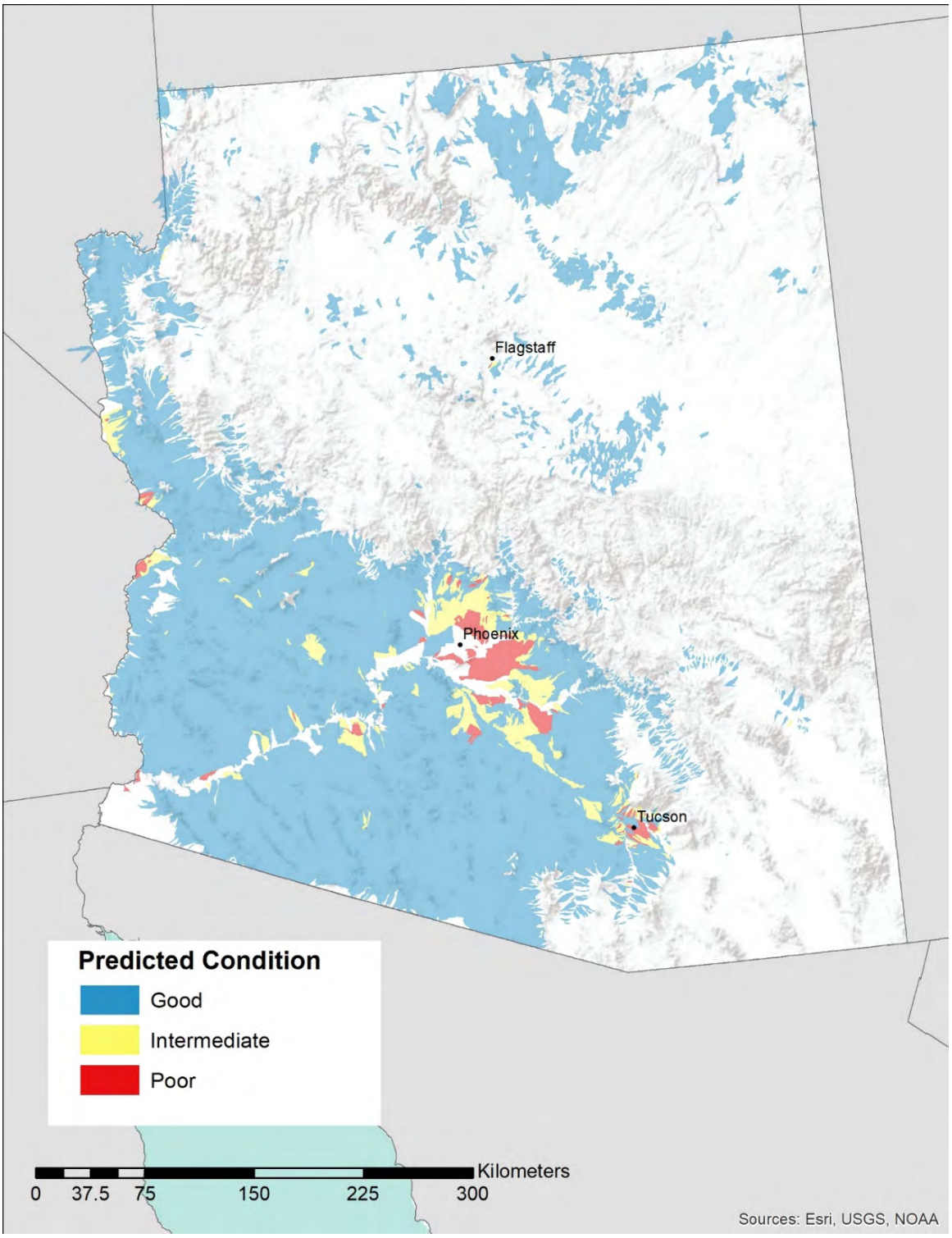


Figure 30. Estimated stress levels on Arizona episodic catchments based on the total percent of watershed covered by urban (high, medium, low, open) and agriculture (crop, hay) land uses. Good condition: < 1.2% development in the watershed. Intermediate condition: 1.2 to 53.5% development in the watershed. Poor condition: > 53.5% development in the watershed.

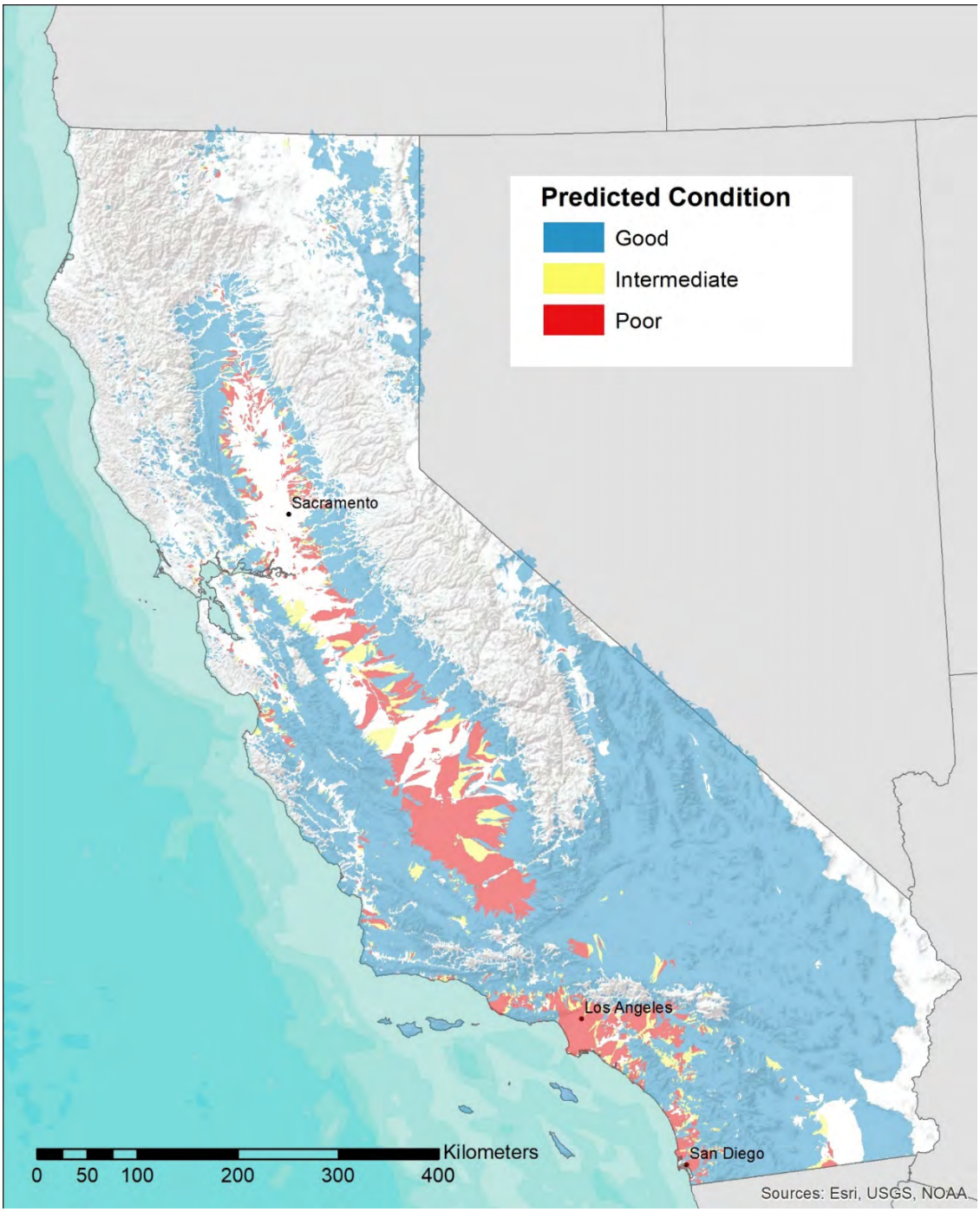


Figure 31. Estimated stress levels on California intermittent and ephemeral streams based on the total percent of watershed covered by urban (high, medium, low, open) and agriculture (crop, hay) land uses. Conditions are predicted based on terrestrial indicators, but presented for both intermittent streams (where both terrestrial and aquatic indicators are needed) and ephemeral streams (where only terrestrial indicators are needed. Good condition: < 22.9% development in the watershed. Intermediate condition: 22.9 to 59.0% development in the watershed. Poor condition: > 59.0% development in the watershed.

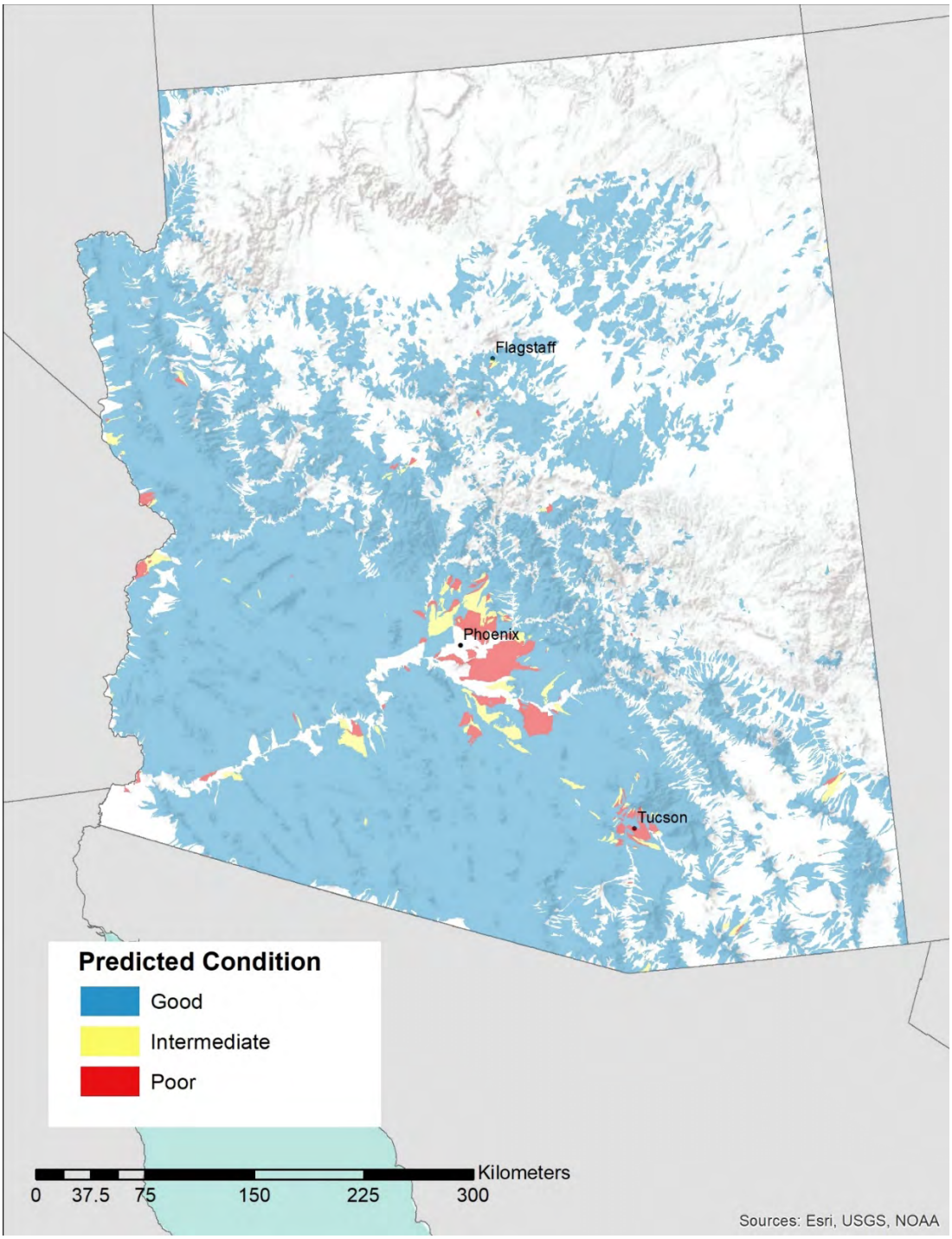


Figure 32. Estimated stress levels on Arizona intermittent and ephemeral streams based on the total percent of watershed covered by urban (high, medium, low, open) and agriculture (crop, hay) land uses. Conditions are predicted based on terrestrial indicators, but presented for both intermittent streams (where both terrestrial and aquatic indicators are needed) and ephemeral streams (where only terrestrial indicators are needed. Good condition: < 25.6% development in the watershed. Intermediate condition: 25.6 to 75.6% development in the watershed. Poor condition: > 75.6% development in the watershed.

Discussion

Conditions of nonperennial streams in California and Arizona

Our L1 assessments show that ephemeral and intermittent streams are widespread in the Southwest, and largely in good biological condition, although large extents of degraded intermittent streams are evident in parts of California, where agricultural and urban development are more extensive. The maps created as part of this project allow managers to prioritize protection or restoration efforts in their jurisdictions, depending on the prevalence of streams likely to be in good or poor condition. Moreover, they can use the maps to more effectively plan their monitoring programs by knowing which L2 or L3 assessment tools are appropriate.

It may be possible to improve the accuracy of models (both those that predict habitat, and those that predict likely ecological condition). These models made extensive use of the StreamCat data set (Hill et al. 2019), which generally emphasizes large-scale watershed characteristics, rather than local features. Although watershed characteristics are demonstrated to correlate with reach-scale conditions in dry streambeds, the relationships appear to be less strong than for intermittent or perennial streams (Mazor et al. 2014; Mazor et al. 2019a). Inclusion of more local factors as predictors may improve model performance.

The benefits of modeling sampleability rather than hydrology to predict habitat

Predicting the locations of intermittent or ephemeral streams has often relied on hydrologic models (e.g., Sengupta et al. 2018). But predicting hydrologic regimes is challenging, particularly in arid regions with highly variable precipitation patterns and complex geology, and even more so when predicting low-flow conditions (Miller et al. 2018; Sengupta et al. 2018). But perhaps the biggest shortcoming of these efforts to model hydrologic regimes, from the perspective of natural resource management and environmental monitoring, is that predictions of discharge or hydrologic metrics is generally insufficient to predict how flow interacts with geomorphology to create the habitats needed to support ecological communities. A hydrologic model may predict discharge or characteristics of a flow regime for a given site, but that is not enough to know if a reach can support certain taxa or if a field crew will be able to collect samples at a site. By directly modeling the outcome relevant to monitoring programs (i.e., sampleability), our maps offer a useful guide for practitioners to plan their sampling efforts. Recent approaches using functional flows make efforts to bridge the gap between strict hydrology and ecologically relevant outcomes (Yarnell et al. 2015), and may also provide useful guidance for monitoring programs.

Using the L1 predictions to select L2 and L3 assessment tools

Geodatabases containing the shapefiles used to produce Figure 24 to Figure 27 can be downloaded by links in the Supplemental Material (specifically, the shapefiles named “L2 predicted habitat” and “L3 predicted habitat”). The L2 file for California provides greater spatial resolution than the current CRAM Episodic Module field book provides (which is based on more coarse scale climate patterns than the models used here). Consulting the predictions in the geodatabases will allow practitioners to anticipate which tool they will likely need. The specific field conditions observed at a site supersede these predictions, as described in the CRAM field book (CWMW 2013).

Supplemental Material

Geodatabases containing shapefiles for the two states may be downloaded below:

California:

ftp://ftp.sccwrp.org/pub/download/PROJECTS/EPA_EphemeralStreams/CAgdb.zip

Arizona:

ftp://ftp.sccwrp.org/pub/download/PROJECTS/EPA_EphemeralStreams/AZgdb.zip

Each geodatabase contains the following shapefiles:

- L2 predicted habitat. A shapefile of catchment polygons with the following columns:
    - COMID: Unique identifier of the catchment. May be used to link data with the NHD Plus or StreamCat data (Hill et al. 2016).
    - Habitat: Habitat classification of the catchment as Episodic or Riverine
- L2 predicted condition. A shapefile of episodic catchment polygons with the following columns:
    - COMID: Unique identifier of the catchment. May be used to link data with the NHD Plus or StreamCat data (Hill et al. 2016).
    - PctDeveloped: Sum of urban, agricultural and developed open space in the watershed. Derived from StreamCat (Hill et al. 2016).
    - Condition: Predicted condition, based on PctDeveloped:
        - Good: PctDeveloped is less than 17.8%
        - Intermediate: PctDeveloped is between 17.8 and 55.3%
        - Poor: PctDeveloped is greater than 55.3%
- L3 predicted habitat. A shapefile of catchment polygons with the following columns:
    - COMID: Unique identifier of the catchment. May be used to link data with the NHD Plus or StreamCat data (Hill et al. 2016).
    - Habitat: Habitat classification of the catchment as perennial (i.e., where aquatic indicators are needed), intermittent (i.e., where aquatic and terrestrial indicators are needed), or ephemeral (i.e., where terrestrial indicators are needed).
- L3 predicted condition. A shapefile of intermittent or ephemeral catchment polygons with the following columns:
    - COMID: Unique identifier of the catchment. May be used to link data with the NHD Plus or StreamCat data (Hill et al. 2016).

- PctDeveloped: Sum of urban, agricultural and developed open space in the watershed. Derived from StreamCat (Hill et al. 2016).
- Condition: Predicted condition, based on PctDeveloped:
    - Good: PctDeveloped is less than 29.6%
    - Intermediate: PctDeveloped is between 29.6 and 47.8%
    - Poor: PctDeveloped is greater than 47.8%

REFERENCES

Arizona Department of Environmental Quality (ADEQ). 2015. Implementation procedures for the narrative biocriteria standard. Phoenix, AZ. Available from: https://legacy.azdeq.gov/environ/water/standards/download/draft_bio.pdf.

Armitage, P.D. and Bass, J.A.B. 2013. Long-term resilience and short-term vulnerability of South Winterbourne macroinvertebrates. Proceedings of the Dorset Natural History and Archaeological Society. 134.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99 002. U.S. Environmental Protection Agency; Office of Water; Washington, D.C.

Bulmer, M.G. 1979. Principle of Statistics. Dover Publications, Inc. New York.

California Wetlands Monitoring Workgroup (CWMW). 2016. Data Quality Assurance Plan: California Rapid Assessment for Wetlands. Sacramento, CA https://www.cramwetlands.org/sites/default/files/CRAM%20data%20QA%20plan%20v7-2018.10.pdf (accessed 12/14/2020)

California Wetland Monitoring Workgroup (CWMW). 2013. California Rapid Assessment Method (CRAM) for Wetlands User's Manual, Version 6.1 pp. 67. http://www.cramwetlands.org/sites/default/files/2013-04-22_CRAM_manual_6.1%20all.pdf (accessed 03/26/2020)

California Wetland Monitoring Workgroup (CWMW). 2011. Steps to Develop a Module for the California Rapid Assessment Method for Wetlands. Sacramento, CA. https://www.cramwetlands.org/sites/default/files/Steps%20to%20Develop%20a%20Statewide%20CRAM%20Module_03212011_final.pdf (accessed 03/26/2020)

Chiu, M.C., C. Leigh, R.D. Mazor, N. Cid, V. Resh. 2017. Anthropogenic Threats to Intermittent Rivers and Ephemeral Streams. in: T. Datry, N. Bonada, A. Boulton (eds.), Intermittent Rivers and Ephemeral Streams: Ecology and Management pp. 433-454. Academic Press. London, UK.

Corti R, Datry T. 2015. Terrestrial and aquatic invertebrates in the riverbed of an intermittent river: parallels and contrasts in community organisation. Freshwater Biology 61:1308–1320.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, and K. T. Hess. 2007. Random forests for classification in ecology. Ecology 88:2783–2792

Datry, T., S.T. Larned, and K. Tockner. 2014. Intermittent Rivers: A Challenge for Freshwater Ecology. BioScience 64(3); 229-235

Datry, T., H. Pella, C. Leigh, N. Bonada, and B. Hugueny. 2016. A landscape approach to advance intermittent river ecology. Freshwater Biology 61(8); 1200-1213.

Dodge, Y. 2010. The Concise Encyclopedia of Statistics. New York (NY): Springer. pp 502-505.

Febria CM, Hosen JD, Crump BC, Palmer MA, Williams DD. 2015. Microbial responses to changes in flow status in temporary headwater streams: a cross-system comparison. Frontiers in microbiology 6:522.Fox, E., R. A. Hill, S. G. Leibowitz, A. R. Olsen, D. J. Thornbrugh, and M. H. Weber. 2017. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environmental Monitoring and Assessment 189:316. DOI: 10.1007/s10661-017-6025-0.

Hamada, Y., B. O'Connor, A. Orr, and K. Wuthrich. 2016. Mapping Ephemeral Stream Networks in Desert Environments Using Very-High-Spatial-Resolution Multispectral Remote Sensing. Journal of Arid Environments. V 130: pp 40-48

Hawkins, C.P., R.H. Norris, J.M. Hogue, and J.W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10 (5), 1456-1477.

Herbst, D. and E. Silldorff. 2006. Comparison of the performance of different bioassessment methods: similar evaluations of biotic integrity from separate programs and procedures. Journal of North American Benthological Society 25(2); 513-530.

Herbst, D., and E. Silldorff. 2009. Development of a benthic macroinvertebrate Index of Biotic Integrity (IBI) for stream assessment in the eastern Sierra Nevada of Caliofrnia. Surface Water Ambient Monitoring Program. Sacramento, CA.

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 Association (JAWRA) 52:120-128. DOI: 10.1111/1752-1688.12372.

Holway, D.A. 1998. Effect of Argentine ant invasions on ground-dwelling arthropods in northern California riparian woodlands. Oecologia 116(1-2): 252-258.

Kerezsy A, Gido K, Magalhães MF, Skelton PH. 2017. The biota of intermittent rivers and streams: Fishes. In: Intermittent rivers and ephemeral streams: Ecology and management. Elservier. p. 271-298.

Krueper, D. 1996. In: Shaw, Douglas W.; Finch, Deborah M., tech coords. Desired future conditions for Southwestern riparian ecosystems: Bringing interests and concerns together. 1995 Sept. 18-22, 1995; Albuquerque, NM. General Technical Report RM-GTR-272. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. p. 281-301.

Levick, L. 2010. Ecological and Hydrological Significance of Episodic Streams. Workshop Presentation at the Southern California Coastal Water Research Project. Costa Mesa, CA.

Levick, L., J. Fonseca, D. Goodrich, M. Hernandez, D. Semmens, J. Stromberg, R. Leidy, M. Scianni, D. P. Guertin, M. Tluczek, and W. Kepner. 2008. The Ecological and Hydrological Significance of Ephemeral and Intermittent Streams in the Arid and Semi-arid American Southwest. U.S. Environmental Protection Agency and USDA/ARS Southwest Watershed Research Center, EPA/600/R-08/134, ARS/233046, 116 pp.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18–22.

Mazor, R.D., K. Schiff, P.R. Ode, and E.D. Stein. 2012. Final report on bioassessment in nonperennial streams. Southern California Coastal Water Research Project Technical Report 695. Costa Mesa, CA.

Mazor, R.D., E.D. Stein, P.R. Ode, and K. Schiff. 2014. Integrating intermittent streams into watershed assessments: applicability of an index of biotic integrity. Freshwater Science 33(2):459–474.

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);249-271.

Mazor, R., J. Olson, M. Robinson, A. Caudillo, and J. Brown. 2019a. Assessing the biological condition of dry ephemeral and intermittent streams. Southern California Coastal Water Research Project Technical Report #1089. Costa Mesa, CA.

Mazor, R.D., V.H. Resh, and D.M. Rosenberg. 2019b. Use of Aquatic Insects in Bioassessment. in: R.W. Merritt, K.W. Cummins, M.B. Berg (eds.), An Introduction to the Aquatic Insects of North America pp. 141-164. Kendall Hunt Publishing Company. Dubuque, IA.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Reah. 2012. NHDPlus Version 2: User guide.

Menke, S.B., Fisher, R.N, Jetz, W., and Holway, D.A. 2007. Biotic and abiotic controls of argentine ant invasion success at local and landscape scales. Ecology 88(12): 3164-3173.

Micacchion, M. and B. Gara. 2008. An ecological and functional assessment of urban wetlands in central Ohio. Volume 3: Comparisons of the Amphibian Communities of Urban and Reference Wetlands Using Level 1, 2 and 3 Assessment Tools. Ohio EPA Technical Report WET/2008-1. Ohio Environmental Protection Agency, Wetland Ecology Group, Division of Surface Water, Columbus, Ohio.

Miller, M.P., D.M. Carlisle, D.M. Wolock, and M. Wieczorek. 2018. A database of natural monthly streamflow estimates from 1950 to 2015 for the Conterminous United States. Journal of the American Water Resources Association 54(6). https://doi.org/10.1111/1752-1688.12685.

Newmaster, S.G., R.J. Belland, A. Arsenault, D.H. Vitt, and T.R. Stephens. 2005. The ones we left behind: comparing plot sampling and floristic habitat sampling for estimating bryophyte diversity. Diversity and Distributions 11(1): 57-72.

National Hydrography Dataset (NHD). 2020. United States Geological Survey. https://www.usgs.gov/core-science-systems/ngp/national-hydrography (accessed 4/7/2020).

Ode, P. R., A. C. Rehn, R. D. Mazor, K. C. Schiff, E. D. Stein, J. T. May, L. R. Brown, D. B. Herbst, D. Gillett, K. Lunde, and C. P. Hawkins. 2016. Evaluating the adequacy of a reference-pool site for ecological assessments in environmentally complex regions. Freshwater Science 35:237–248.

Patrick, L.B. and A. Hansen. 2013. Comparing ramp and pitfall traps for capturing wandering spiders. Journal of Arachnology 41:404-6.

Pearce, J.L., D. Schuurman, K.N. Barber, M. Larrivée, L.A. Venier, J. McKee, and D. McKenney. 2005. Pitfall trap designs to maximize invertebrate captures and minimize captures of nontarget vertebrates. The Canadian Entomologist 137:233-50.

Pima County. 2000. Biological Stress Assessment, and Overview Discussion of Issues and Concerns. Sonoran Desert Conservation Plan website. http://www.pima.gov/CMO/SDCP/reports/d9/008BIO.PDF

Rehn, A., R. Mazor, and P. Ode. 2015. The California Stream Condition Index (CSCI): A New Statewide Biological Scoring Tool for Assessing the Health of Freshwater Streams. SWAMP Technical Memorandum 2015-0002.

Robinson, M.D., J.R. Olson, and R.D. Mazor. 2016. Development of biological indicators for the dry phase of non-perennial rivers and streams. Master's Thesis.

Romaní AM, Chauvet E, Febri C, Mora-Gómez J, Risse-Buhl U, Timoner X, Weitere M, Zeglin L. 2017. The biota of intermittent rivers and ephemeral streams: Prokaryotes, fungi, and protozoans. In: In: Intermittent rivers and ephemeral streams: Ecology and management. Elservier. p. 161-188.

Roth, N.E., Allan, J.D. and Erickson, D.L.:1996, 'Landscape influences on stream biotic integrity assessed at multiple spatial scales', Landscape Ecol. 11, 141–156.

Sabater, S., Tockner, K. 2009. Effects of hydrologic alterations on the ecological quality of river ecosystems. In *Water scarcity in the Mediterranean* (pp. 15-39). Springer, Berlin, Heidelberg

Sabater S, Timoner X, Bornette G, De Wilde M, Stromberg J, Stella JC. 2017. The biota of intermittent rivers and intermittent streams: Algae and vascular plants. In: Intermittent rivers and ephemeral streams: Ecology and management. Elservier. p. 189-216.

Sánchez-Montoya MM, Moleón M, Sánchez-Zapata JA, Tockner K. 2016a. Dry riverbeds: corridors for terrestrial vertebrates. Ecosphere 7(10).

Sánchez-Montoya MM, Von Schiller D, Ruhí A, Pechar GS, Proia L, Miñano J, Vidal-Abarca MR, Suárez ML, Tockner K. 2016b. Responses of ground-dwelling arthropods to surface flow drying in channels and adjacent habitats along Mediterranean streams. Ecohydrology 9(7):1376-1387.

Sánchez-Montoya MM, Moleón M, Sánchez-Zapata JA, Escoriza D. 2017. The biota of intermittent and ephemeral rivers: Amphibians, reptiles, birds, and mammals. In: Intermittent rivers and ephemeral streams: Ecology and management. Elservier. p. 299-322.

SAS Institute Inc. 2011. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc.

Shaw, J. and Cooper, D.J. 2008. Linkages among watersheds, stream reaches, and riparian vegetation in dryland ephemeral stream networks. Journal of Hydrology 350: 68-82.

Southern California Coastal Water Research Project (SCCWRP). 2018. Quality Assurance Project Plan for Developing and Validating Assessment Tools for Ephemeral (Dry) Streams in California and Arizona. Funding Number: CD-99T65301-0. Costa Mesa, CA.

Sengupta, A., S.K. Adams, B.P. Bledsoe, E.D. Stein, K. McCune, and R.D. Mazor. 2018. Tools for managing hydrologic alteration on a regional scale: Estimating changes in flow characteristics at ungauged sites. Freshwater Biology 63(8): 769-785. DOI:10.1111/fwb.13074.

Shaw, J. and D. Cooper. 2008. Linkages among Watersheds, Stream Reaches, and Riparian Vegetation in Dryland Ephemeral Stream Networks. Journal of Hydrology v 350, pp 68-82.

Stein, E.D., A.E. Fetscher, R.P. Clark, A. Wiskind, J.L. Grenier, M. Sutula, J.N. Collins, and C. Grosso. 2009. Validation of a wetland rapid assessment method: Use of EPA's level 1-2-3 framework for method testing and refinement. Wetlands 29:648-665.

Stein, E.D., K. Vyverberg, G.M. Kondolf, and K. Janes 2011. Episodic Stream Channels: Imperatives for Assessment and Environmental Planning in California: Proceedings of a Special Technical Workshop. November 8-10, 2010. Southern California Coastal Water Research Project Technical Report # 645.

Steward, A.L., D. von Schiller, K. Tockner, J.C. Marshall, and S.E. Bunn. 2012. When the river runs dry: human and ecological values of dry riverbeds. Frontiers in Ecology and the Environment 10: 202–209

Steward, A.L., P. Negus, J.C. Marshall, S.E. Clifford, and C. Dent. 2018. Assessing the ecological health of rivers when they are dry. Ecological Indicators. 85. 537-547.

Steward, A.L., S.D. Langhans, R. Corti, and T. Datry. 2017. Chapter 4.4: The biota of intermittent rivers and ephemeral streams – Terrestrial and semi-aquatic invertebrates. In: Datry, T., Bonada, N., Boulton, A. (Eds), Intermittent Rivers and Ephemeral Streams: Ecology and Management, Elsevier. pp. 245–271.

Stoddard, J.T., D.P. Larsen, C.P. Hawkins, R.K. Johnson, and R.H. Norris. 2006. Setting Expectations for the ecological condition of streams: the concept of reference condition. Ecological Indicators 16(4); 1267-1276.

Stubbington R, Datry T. 2013. The macroinvertebrate seedbank promotes community persistence in temporary rivers across climate zones. Freshwater Biology 58(6):1202-1220.

Stubbington R, England J, Wood PJ, Sefton CEM. 2017. Temporary streams in temperate zones: recognizing, monitoring and restoring transitional aquatic-terrestrial ecosystems. Wiley Interdisciplinary Reviews: Water 4(4):e1223.

Stubbington, R., Chadd, R., Cid, N., Csabai, Z., Miliša, Morais, M., Munné, A., Pařil, P., Pešic, V., Tziortzis, I., Verdonschot, R.C.M., and Datry, T. 2018. Biomonitoring of intermittent rivers and ephemeral streams in Europe: Current practice and priorities to enhance ecological status assessments. Science of the Total Environment 618. https://doi.org/10.1016/j.scitotenv.2017.09.137

Sutula, M.A., E.D. Stein, J.N. Collins, A.E. Fetscher, and R. Clark. 2006. A practical guide for the development of a wetland assessment method: The California experience. Journal of the American Water Resources Association 42:157-175.

Taka, M., J. Aalto, J. Virkanen, and M. Luoto. 2016. The direct and indirect effects of watershed land use and soil type on stream water metal concentrations. Water Resources Research 52(10): pp 7711-7725.

Theroux, S., R.D. Mazor, M.W. Beck, P.R. Ode, E.D. Stein, and M. Sutula. 2020. Predictive Biological Indices for Algae Populations in Diverse Landscapes. Ecological Indicators. 119" 106421. DOI: https://doi.org/10.1016/j.ecolind.2020.106421

Tornés E, Ruhí A. 2013. Flow intermittency decreases nestedness and specialisation of diatom communities in Mediterranean rivers. Freshwater Biology 58(12):2555-2566.

Vander Laan, J.J. and C.P. Hawkins. 2014. Enhancing the Performance and interpretation of freshwater biological indices: an application in arid zone systems. Ecological Indicators 36; 470-482.

von Schiller, D., V. Acuña, D. Graeber, E. Martí, M. Ribot, S. Sabater, X. Timoner, and K. Tockner. 2011. Contraction, fragmentation and expansion dynamics determine nutrient availability in a Mediterranean forest stream. Aquatic Sciences 73: 485–497.

Vyvlerberg, K. 2010. A review of stream processes and forms in dryland watersheds. California Department of Fish and Game. Sacramento, California. Available from: https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=25779

Wood PJ, Boulton AJ, Little S, Stubbington R. 2010. Is the hyporheic zone a refugium for macroinvertebrates during severe low flow conditions? Fundamental and Applied Limnology/Archiv für Hydrobiologie 176(4):377-390.

Yarnell, S.M., G.E. Petts, J.C. Schmidt, A.A. Whipple, E.E. Beller, C.M. Dahm, P. Goodwin, and J.H. Viers. 2015. Functional flows in modified riverscapes: Hydrographs, habitats, and opportunities.