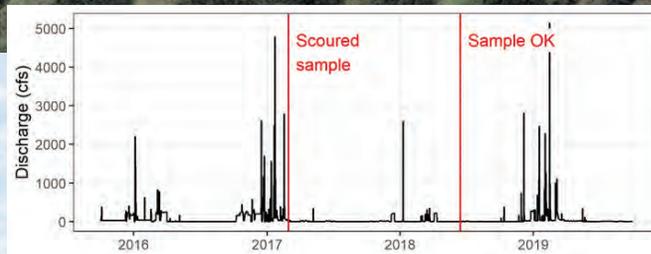
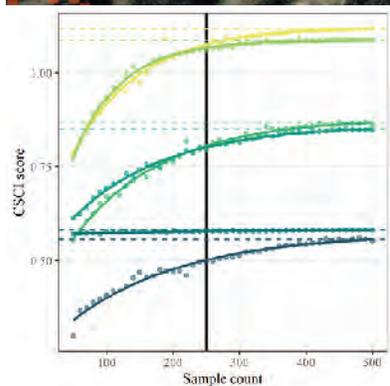
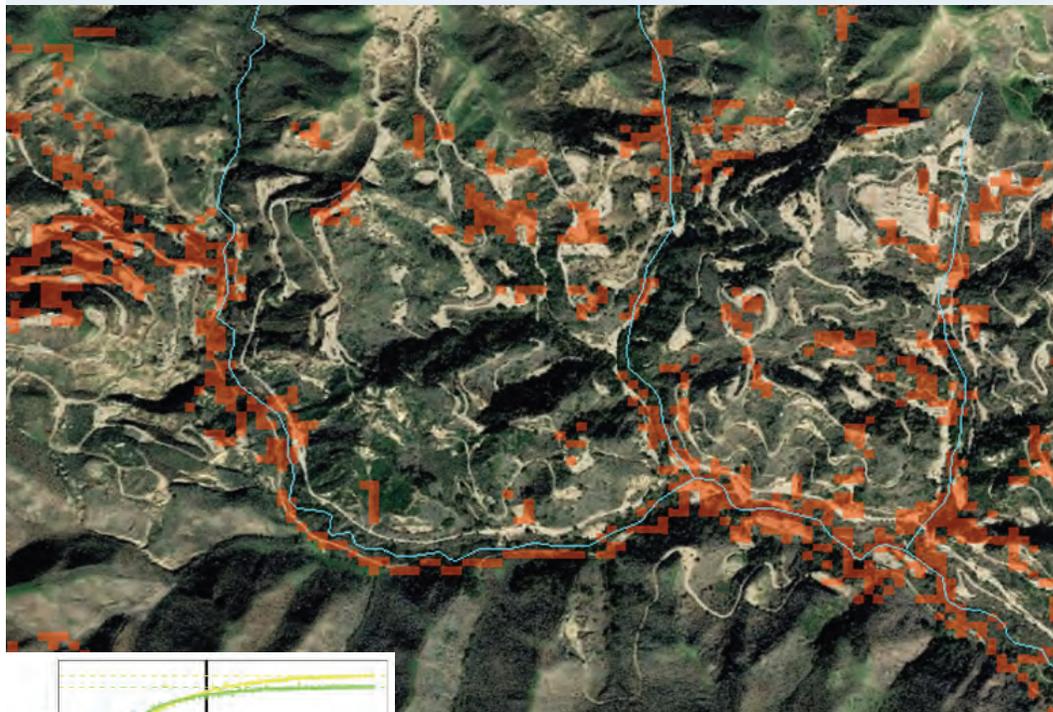
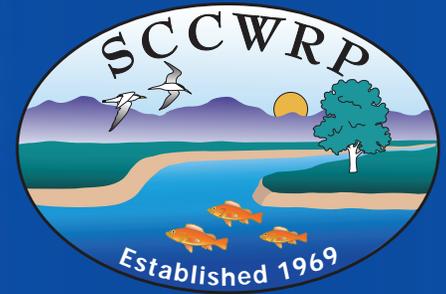


# A decision framework for evaluating bioassessment samples and landscape models



Marcus Beck  
Raphael Mazor

*Southern California Coastal Water Research Project*

SCCWRP Technical Report #1115

# **A decision framework for evaluating bioassessment samples and landscape models**

Marcus Beck and Raphael Mazor

*Southern California Coastal Water Research Project, Costa Mesa, CA*

**April 2020**  
Technical Report 1115

## **ACKNOWLEDGEMENTS**

This report was prepared for the San Gabriel River Regional Monitoring Program.

This document should be cited as:

Beck, M. and Mazon, R. 2020. A decision framework for evaluating bioassessment samples and landscape models. Technical Report 1115. Southern California Coastal Water Research Project. Costa Mesa, CA.

# TABLE OF CONTENTS

Acknowledgements.....	i
Table of Figures.....	vi
Background.....	1
Description of bioassessment tools.....	1
The California Stream Condition Index (CSCI).....	1
The Stream Classification and Prioritization Explorer (SCAPE).....	1
Validation and data evaluation.....	2
Evaluating the adequacy of BMI data for calculating CSCI scores.....	2
General guidance on evaluating substitute data.....	2
Is the CSCI score close to a key threshold?.....	3
Why is this a problem?.....	3
Where do you find an answer?.....	3
How do you evaluate the answer?.....	4
What can you do about it?.....	4
Was the BMI sample collected with an appropriate method?.....	4
Why is this a problem?.....	4
Where do you find an answer?.....	4
How do you evaluate the answer?.....	4
Were the data collected by an adequately trained, intercalibrated, and audited crew?.....	5
Why is this a problem?.....	5
Where do you find an answer?.....	5
How do you evaluate the answer?.....	6
Was the sample collected outside the typical index period?.....	6
Why is this a problem?.....	6
Where do you find an answer?.....	7
How do you evaluate the answer?.....	7
Are there enough organisms in the sample?.....	7
Why is this a problem?.....	7
Where do you find an answer?.....	8
What can you do about it?.....	8
How do you evaluate the answer?.....	8
What can you do about it?.....	9
Are there too many ambiguous identifications in the sample?.....	10
Why is this a problem?.....	10
Where do you find an answer?.....	10

How do you evaluate the answer?.....	11
What can you do about it? .....	12
Was the sample influenced by scouring flows? .....	13
Where do you find answers and how are they evaluated? .....	13
What can you do about it? .....	15
Was the sample influenced by drying? .....	15
Why is this a problem? .....	15
Where do you find an answer? .....	15
How do you evaluate the answer?.....	15
What can you do about it? .....	16
Was the sample influenced by drought?.....	16
Why is this a problem? .....	16
Where do you find an answer? .....	16
How do you evaluate the answer?.....	17
What can you do about it? .....	17
Was the sample affected by fire? .....	17
Why is this a problem? .....	17
Where do you find an answer? .....	17
How do you evaluate the answer?.....	18
What can you do about it? .....	18
Was the sample influenced by vegetation management or debris removal?.....	18
Why is this a problem? .....	18
Where do you find an answer? .....	18
How do you evaluate the answer?.....	19
What can you do about it? .....	19
Was the sample influenced by vector control activities? .....	19
Why is this a problem? .....	19
Where do you find an answer? .....	19
How do you evaluate the answer?.....	19
What can you do about it? .....	19
Was the sample influenced by tides or naturally high salinity? .....	19
Why is this a problem? .....	19
Where to you find an answer? .....	20
How do you evaluate the answer?.....	20
What can you do about it? .....	20
Was the sample collected from a setting where the CSCI is suspected to give low scores?..	21

Why is this a problem? .....	21
Where do you find an answer, and how do you evaluate the information? .....	21
What can you do about it? .....	22
Is the watershed delineation used to calculate CSCI predictors appropriate? .....	22
Why is this a problem? .....	22
How do you evaluate the answer?.....	23
What can you do about it? .....	23
Are GIS predictors correct? .....	24
Why is this a problem? .....	24
Where do you find an answer? .....	25
How do you evaluate the answer?.....	25
What can you do about it? .....	25
Evaluating the adequacy of the SCAPE landscape model .....	25
Is the sampling reach atypical of the channel's overall constraint class? .....	26
Why is this a problem? .....	26
Where do you find an answer? .....	26
How do you evaluate the answer?.....	26
What can you do about it? .....	27
Has land cover changed since the SCAPE model was calibrated? .....	27
Why is this a problem? .....	27
Where do you find an answer? .....	27
How do you evaluate the answer?.....	28
What can you do about it? .....	28
Could the site be influenced by constraints not captured by the model?.....	29
Why is this a problem? .....	29
Where do you find an answer? .....	30
How do you evaluate the answer?.....	30
What can you do about it? .....	30
Does the NHD Plus hydrography accurately represent the stream channel?.....	30
Why is this a problem? .....	30
Where do you find an answer? .....	30
How do you evaluate the answer?.....	30
What can you do about it? .....	32
Are the results of the SCAPE model close to a key threshold?.....	32
Why is this a problem? .....	32
Where do you find an answer? .....	32

How do you evaluate the answer?.....	32
What can you do about it? .....	32
What to do if the SCAPE model outputs are invalidated .....	32
Manual substitution .....	32
Applying model averages for stream-types.....	33
Applying constraints from empirical distributions (e.g., engineered channels in Southern California).....	34
Applying constrains based on habitat condition .....	35
Helpful resources .....	36
Cited literature.....	37
Appendices .....	39
Appendix 1 .....	39
Evaluating sample-count effects on CSCI scores .....	39
Appendix 2 .....	40
Evaluating the effects of ambiguous identifications on CSCI scores.....	40
Appendix 3.....	41
Typical values of CSCI predictors in the South Coast region.....	41
Appendix 4.....	42
Ranges of expected CSCI scores for different stream types in regions of California.....	42
Appendix 5.....	44
Relationship between CSCI scores and habitat condition.....	44

## TABLE OF FIGURES

Figure 1. Probability of a site having a true mean CSCI score above a threshold (e.g., $\geq 0.79$ ), based on a single observation and assuming a within-site standard deviation of 0.11 (as reported in Mazor et al. 2016). .....	3
Figure 2. CSCI scores at three southern California reference sites sampled before, during, and after the typical index period (gray) in 2009 (Bear Canyon and South Fork) or 2010 (Cedar Canyon). Apart from a late-summer sample collected from Cedar Canyon, all samples scored well above the threshold of 0.79 (dashed line). .....	7
Figure 3. The CSCI core report includes metadata that can be used to evaluate the BMI sample. The total number of organisms is indicated in the "Count" column, highlighted in red. ....	8
Figure 4. Effects of systematically reducing sample count on CSCI scores for six sites. The dashed lines indicate the true CSCI score at each site, and the vertical black line indicates a minimum sample size of 250. ....	9
Figure 5. CSCI metadata that can be evaluated from the standard results. The second sample returns an invalid CSCI score because of many ambiguous individuals and taxa (in red). .....	11
Figure 6. Effects of systematically increasing the proportion of ambiguous taxa in a sample on CSCI scores for six sites. The dashed lines indicate the true CSCI score at each site and the vertical black line indicates a recommended maximum of 50% ambiguous taxa. ....	12
Figure 7. Flow record for four years at the San Gabriel mainstem. ....	14
Figure 8. Watershed delineation examples showing snapping challenges. The pink dot is the recorded site latitude/longitude and the green dot is the snapped location. The watershed boundaries resulting in each case are outlined in red. ....	23
Figure 9. Changes in CSCI scores (left) and E (number of expected taxa, right) when correct predictors are used versus inflated precipitation. The red line indicates a sample where changes were large enough to drop the score below a threshold of 0.79 (dotted line, left). ....	24
Figure 10. An unexpected stream class is validated by examining the land use/land cover base layer. ....	26
Figure 11. An example showing an unconstrained stream in Orange County that has undergone extensive land cover change. The image on the left is from 2007, and the image on the right is from 2019. ....	28
Figure 12. Data in the SCAPE model may underestimate landscape alteration, as seen in Cañada del Diablo in Ventura County. Oil extraction activity has greatly altered the landscape (evident as clearings in the satellite imagery), yet only a fraction of these disturbed areas is recorded as non-natural land types (red pixels) in the landcover data in StreamCat. ....	29
Figure 13. Four examples of spatial errors in the NHD Plus. Light blue flowlines represent possibly unconstrained streams, and orange flowlines represent possibly constrained streams. ....	31
Figure 14. An example where manual substitution may provide appropriate estimates of likely ranges of CSCI scores. ....	33
Figure 15. CSCI scores compared to IPI and CRAM physical habitat measures. The blue lines are quantiles regression results for the 90th percentile and the red dotted line is CSCI at 0.79. IPI: Index of Physical Habitat Integrity. Ev_FlowHab: Evenness of flow habitats. H_SubNat: Diversity of natural substrate types. PCT_SAFN: Percent sands, fines, and concrete in the substrate. XCMG: Mean riparian vegetation cover. CRAM: California Rapid Assessment Method. Blc: CRAM buffer and landscape condition attribute. Bs: CRAM biotic structure attribute. Hy: CRAM hydrologic structure attribute. Ps: CRAM physical structure attribute. ....	35

Figure 16. Effects of reducing sample counts on CSCI scores .....39  
Figure 17. Effects of ambiguous identifications on CSCI scores.....40

## **BACKGROUND**

Bioassessment indices and related tools are increasingly used to support management decisions about aquatic resources, such as identifying which sites should be prioritized for restoration or protection. As the adoption of bioassessment data grows, questions may arise about how to interpret index scores. For example, managers need to determine if a poor index score reflects ecological degradation related to human activity, or if other factors (like low numbers of organisms in the sample) should be considered. This document is intended to help managers (particularly those new to working with bioassessment data) develop expertise in evaluating the usability of bioassessment data. Our hope is that this guidance will improve transparency and facilitate communication among stakeholders who must make decisions about watershed management based on bioassessment data.

This document is focused on two major bioassessment tools: the California Stream Condition Index (CSCI; Mazor et al. 2016), and the Stream Classification and Prioritization Explorer tool (SCAPE, Beck et al. 2019). Herein, we describe situations that may affect the interpretation of these tools, including some circumstances that may invalidate their outputs. Although written specifically for the CSCI and SCAPE tools, some of this guidance may apply to other bioassessment tools, such as the Algal Stream Condition Index (Theroux et al. in review).

### **Description of bioassessment tools**

The California Stream Condition Index (CSCI) is a tool for evaluating the biological integrity of wadeable streams in California using aquatic insect as indicators of stream health and is applicable to most of the diverse conditions found in the state (Mazor et al. 2016). The CSCI is used in routine assessments conducted by the San Gabriel River Regional Monitoring Program (SGRRMP), a cooperative organization that manages aquatic resources in the watershed. CSCI data collected by the SGRRMP are the most comprehensive sources of information on biological condition of streams in the region. Established field sampling protocols are in place to ensure that the data are sufficient to provide an accurate representation of biological communities to use the CSCI. Questions about the interpretability of a CSCI score arise when unusual circumstances affect a sampling event or sample analysis, such as low numbers of organisms in a sample, or scouring events a few days prior to sample collection.

The Stream Classification and Prioritization Explorer (SCAPE) is a statistical model that calculates expected ranges of CSCI scores at nearly all stream-segments in California, based on measures of landscape alteration, such as urban land cover and road density (Beck et al. 2019). Watershed groups like the SGRRMP can use these predicted ranges and compare them to observed CSCI scores to identify streams that are scoring within, above, or below their expected range, and set management priorities accordingly. For example, the SGRRMP has prioritized sites with scores below their expected range for further investigation, which may lead to causal assessment and restoration. The underlying data that inform the model comes from the National Hydrography Dataset Plus (NHD Plus; McKay et al. 2012), as well as the StreamCat database (Hill et al. 2016), which is a collection of landscape-level metrics calculated for each stream segment in the NHD Plus. Questions about the SCAPE tool's appropriateness arise when these two data sources poorly characterize the sampling location (e.g., extensive land use change has occurred between the development of the StreamCat data set and biological sample collection).

## Validation and data evaluation

This document describes a set of questions to consider when evaluating outputs from the CSCI or SCAPE model. For certain intended uses of these tools, answers to these questions may invalidate the data. There are only a few circumstances where invalidation is categorically recommended for all intended uses of the data (e.g., benthic macroinvertebrates [BMI] collected with inappropriate methods should not be used to calculate CSCI scores); more often, these questions provide additional context to help interpret data and make informed decisions.

In this document, the term “validation” refers to one of the following two processes:

- Determining if the CSCI score from a sample is likely to correctly represent the biological conditions of a site at the time of sampling.
- Determining if the outputs from the SCAPE model are likely to correctly represent the environmental setting of the sampling site.

Several activities described in this document do not fall under the validation processes described above but may still be good standard practices to follow when making decisions based on CSCI scores or SCAPE model outputs. For example, this document describes a process to determine if a wildfire may have influenced CSCI scores; evidence of the influence of wildfire does not invalidate a CSCI score, yet it provides useful information when considering possible causes of low CSCI scores.

## EVALUATING THE ADEQUACY OF BMI DATA FOR CALCULATING CSCI SCORES

This document presents several questions to consider when evaluating the adequacy of BMI data for calculating CSCI scores. Several questions pertain to the circumstances of sample collection (e.g., was the sampling event influenced by scour?) or to data production (e.g., was taxonomic resolution sufficiently high?). For each question, we provide an explanation for why the issue may be a problem for interpreting CSCI scores, where to look for information to help you answer the question, and a recommended threshold or framework for interpreting that information. In some cases, we also recommend steps to address the problem (beyond collecting new data).

### General guidance on evaluating substitute data

In many cases when a BMI sample is determined to be unsuitable for CSCI score calculation, the analyst has two options: collect additional samples, or find additional samples that represent the site in question and use them as *substitutes* in analysis. Samples collected from the same site at different times are often the best substitutes, but in some circumstances, samples collected at different sites may be good alternatives. Here, we provide general guidelines on determining if data collected elsewhere is suitable for this purpose. Ideally, the new site should be:

- Hydrologically connected (i.e., upstream or downstream) from the original site
- Ideally, within 300 m
- No major intervening tributaries or discharges
- Same stream-order
- Similar land use and geophysical properties at the reach

- Similar temporal conditions (e.g., collected in the same season)

### Is the CSCI score close to a key threshold?

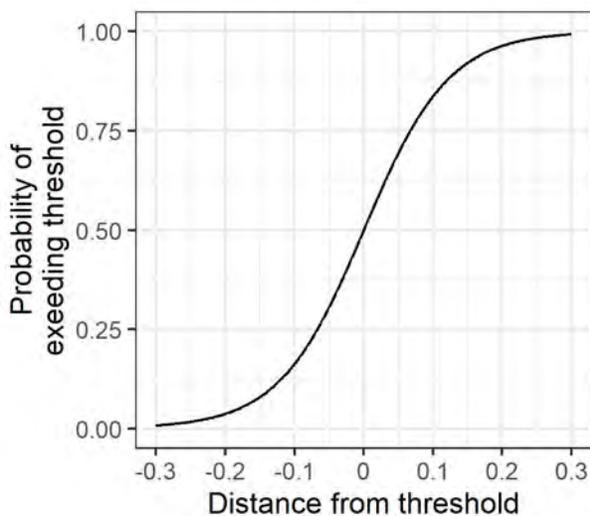
Why is this a problem?

CSCI scores may vary at a site due to a number of factors, such as sampling variability and the patchy distribution of BMI in within a sampling reach. Analysis of replicate sample in Mazor et al. (2016) indicate that within-site sampling standard deviation can be as much as 0.11 points. At the same time, many management decisions rely on bright-line thresholds. For example, the SGRRMP identifies CSCI scores above 0.79 as meeting management goals, with lower scores indicating that goals aren't met.

Where do you find an answer?

Scores from additional samples can indicate the level of variability at a site, providing a measure of confidence that the score is truly above or below a threshold.

If only a single sample was collected, confidence may be estimated by assuming a normal distribution of CSCI scores at a site (Figure 1). Note that for some applications (e.g., identifying impairments), a score from a single sample may be insufficient regardless of how close or far a score is from a threshold.



**Figure 1. Probability of a site having a true mean CSCI score above a threshold (e.g.,  $\geq 0.79$ ), based on a single observation and assuming a within-site standard deviation of 0.11 (as reported in Mazor et al. 2016).**

How do you evaluate the answer?

The importance of determining with confidence that a CSCI score is below (or above) a threshold depends on the type of management question being asked. For example, assessing compliance with a permit may require a high level of confidence that a CSCI score is low, and state guidance requires multiple samples to identify if beneficial uses are impaired. In contrast, prioritizing sites for future sampling (e.g., sites that exceed expected ranges of scores in the San Gabriel watershed) may not require a comparable level of certainty.

What can you do about it?

If the results of the above analyses are ambiguous, and managers are unable to make decisions without more confidence, additional sample collection is recommended.

### **Was the BMI sample collected with an appropriate method?**

Why is this a problem?

Many different collection methods have been used to collect BMI in California, but the CSCI was calibrated for use with the methods described in standard SWAMP protocols (Ode et al. 2016). Other methods may not produce data that can yield a valid CSCI score.

Valid methods all involve collection with a D-frame net from a fixed number of locations along a reach. Many of these methods also prescribe a fixed number of organisms to sort from each sample in the lab. Valid methods fall into one of two general types: those that target the richest microhabitats within a reach (e.g., riffles), and those that sample microhabitats in proportion to their relative abundance within a reach (e.g., reachwide methods). CSCI scores from samples collected from other methods are usually considered invalid.

Where do you find an answer?

This information is not typically stored within the CSCI scores, although it may be incorporated into the SampleID. The SWAMP, SMC, and CEDEN databases store this information as part of taxonomic data under "CollectionMethodCode".

How do you evaluate the answer?

Valid and invalid method codes are described in Table 1.

**Table 1. Valid and invalid collection codes for BMI**

Valid methods		Invalid methods (partial list)	
<b>BMI_RWB</b>	Reach-Wide Benthos collection method for freshwater BMI samples.	<b>BMI_ArtSub</b>	Artificial Substrate collections for BMI samples
<b>BMI_RWB_MCM</b>	Margin-Center-Margin collection method for freshwater BMI samples.	<b>Lentic_CSBP</b>	CSBP collection method for lentic samples
<b>BMI_SNARL</b>	Sierra Nevada Aquatic Research Lab collection method for freshwater BMI samples; collected from 5 randomly selected riffles.  Not in widespread use.	<b>TerInvt_T_DS</b>	Trap collection method for dry stream terrestrial invertebrate samples
<b>BMI_TRC</b>	Targeted Riffle Composite collection method for freshwater BMI samples.  Not in widespread use.	<b>TerInvt_V_DS</b>	Vegetative bag collection method for dry stream terrestrial invertebrate samples
<b>BMI_CSBP_Comp</b>	CSBP collection method for composited (i.e., samples from distinct transects were combined) freshwater BMI samples.  Not in widespread use.	<b>MI_RWB</b>	Reach-Wide Benthos collection method for freshwater macroinvertebrate samples in depressional wetlands
<b>BMI_CSBP_Trans</b>	CSBP collection method for non-composited (i.e., samples from distinct transects were processed separately) freshwater BMI samples.  Not in widespread use.		

Note that database errors where incorrect collection method codes are used may occur. If a sample has an invalid collection method code indicated (especially if a non-BMI collection method is indicated, such as an algae collection method), we recommend you track down additional information to verify which collection method was used.

In general, we do not recommend using the CSCI with data collected from novel methods (i.e., methods not identified in Table 1). If necessary, consult the SWAMP Bioassessment Workgroup for guidance on interpreting the resulting scores.

**Were the data collected by an adequately trained, intercalibrated, and audited crew?**

Why is this a problem?

Bioassessment data is perhaps more difficult to collect than many other kinds of monitoring data. Untrained crews may not be familiar with field practices required to generate comparable data. As described later in this document, well trained and experienced field crews often provide crucial data that can provide insight into the validity of a CSCI score.

Where do you find an answer?

There are several ways to assess the adequacy of a field crew:

- **Training:** Have crews received standard training in SWAMP protocols, such as those provided by the Water Boards Training Academy's College of Bioassessment?
- **Intercalibration:** Have crews participated in intercalibration activities with other field crews working in the region, such as those provided by the Stormwater Monitoring Coalition (SMC)?
- **Auditing:** Have crews been audited by a qualified and independent practitioner? Have corrective actions (if any are recommended) been carried out and documented?
- **Experience:** Have crews implemented the protocol at a sufficient number of sites and under a sufficient variety of conditions (e.g., wet years and dry years, urban sites and forested sites)?

In general, this information is maintained by each field crew. Information on audits and intercalibration events conducted by SWAMP can be obtained from Shawn McBride ([Shawn.McBride@wildlife.ca.gov](mailto:Shawn.McBride@wildlife.ca.gov)), and information on events conducted by the SMC can be obtained from Jeff Brown ([jeffb@sccwrp.org](mailto:jeffb@sccwrp.org)).

Note that data collected by community science groups (also known as citizen-science groups) are considered comparable to data collected by professionals, provided that these crews have undergone the same training and other quality assurance requirements.

How do you evaluate the answer?

In general, data collected by untrained crews should not be used for most decision-making purposes. If there is reason to suspect the data collected by a trained crew, information provided by intercalibration, audits, and documented experience may prove whether those suspicions are warranted.

### **Was the sample collected outside the typical index period?**

Why is this a problem?

The SOP guidelines for field sampling of macroinvertebrates (Ode et al. 2016) states the typical index period as being from May through September to characterize base flow conditions. This period depends on the region, such that sampling can occur towards the earlier end of this range in southern California (typically May 15 to July 15), and later in this range for higher latitudes. Sampling that occurs outside of this range could produce a sample that is not representative of the macroinvertebrate community for which the CSCI is calculated.

Sampling outside this index period could affect a CSCI score in a few ways:

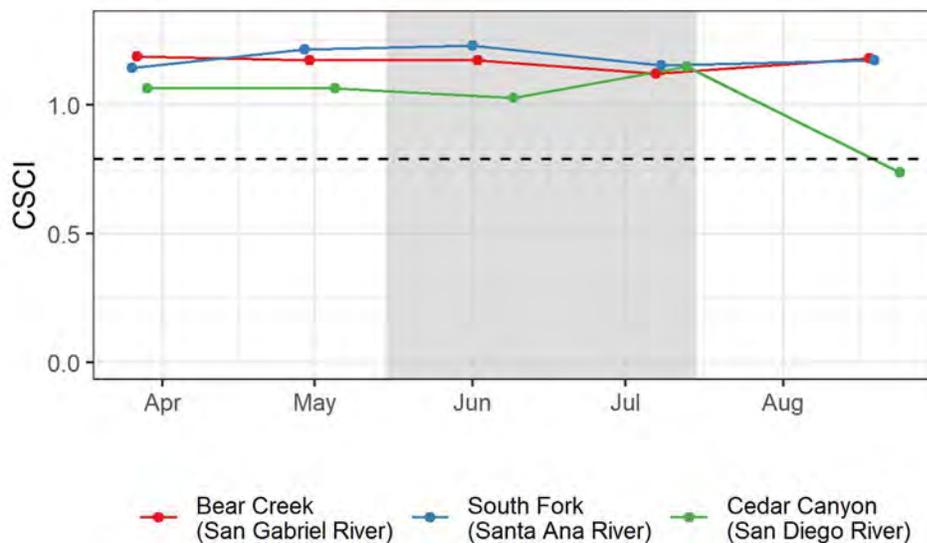
- Samples collected in the winter are more likely to be affected by scour from storms
- Samples collected before the index period may be dominated by immature specimens that cannot be identified to the desired level of taxonomic resolution.
- Samples collected after the index period may lack species that have matured into terrestrial forms and are largely absent from the water.
- Samples collected after the index period may be experiencing stresses related to drying, particularly during extreme droughts.

Where do you find an answer?

The date of sampling is recorded with other sample metadata.

How do you evaluate the answer?

In general, the CSCI is robust and can correctly score samples collected well outside the typical index period (Figure 2). Nonetheless, additional data evaluation may be warranted, as described elsewhere in this document. For samples collected prior to the index period, we recommend evaluating the influence of *scour*, as well as *ambiguous taxa* (which are more common when immature specimens dominate a sample). For samples that could be affected by drying (e.g., the August sample from Cedar Canyon in Figure 2), see the section on *drought*.



**Figure 2. CSCI scores at three southern California reference sites sampled before, during, and after the typical index period (gray) in 2009 (Bear Canyon and South Fork) or 2010 (Cedar Canyon). Apart from a late-summer sample collected from Cedar Canyon, all samples scored well above the threshold of 0.79 (dashed line).**

## Are there enough organisms in the sample?

Why is this a problem?

The CSCI was calibrated with samples containing 600 organisms. Smaller counts may yield inaccurate (and likely lower) CSCI scores because smaller samples may not provide a complete picture of the community that was present during sampling. Sample counts could be low for several reasons, including but not limited to sampling failure (e.g., loss of insects from net

failure), poor timing of sampling (e.g., outside of index period), or protocols from the field or lab manuals were not followed.

Where do you find an answer?

The report that is generated by the CSCI calculator (<https://sccwrp.github.io/CSCI>) includes a “Count” column which indicates the number of organisms present in the original sample (Figure 3).

What can you do about it?

In general, samples collected with inappropriate collection methods cannot be used to calculate CSCI scores. The only recourse is to identify substitute samples or collect new ones.

```
> results$score
```

	StationCode	SampleID	Count	Number_of_MMI_Iterations	Number_of_OE_Iterations			
1	site3	BadSample1	100		1			1
2	site3	BadSample2	600		20			1
3	site1	Sample1	556		20			20
4	site2	Sample2	826		20			20
5	site3	Sample3	607		20			20
6	site3	Sample4	513		20			1
	Pcnt_Ambiguous_Individuals	Pcnt_Ambiguous_Taxa		E	Mean_O		OoverE	
1	0.0000000	0.000000	10.248486	1.00	0.09757538			
2	83.3333333	50.000000	10.248486	1.00	0.09757538			
3	0.5395683	2.631579	7.544418	8.95	1.18630752			
4	0.9685230	1.666667	12.953853	11.25	0.86846749			
5	9.7199341	6.250000	10.248486	12.90	1.25872246			
6	37.6218324	41.025641	10.248486	9.00	0.87817846			
	OoverE_Percentile	MMI	MMI_Percentile	CSCI	CSCI_Percentile			
1	0.00	0.1638082	0.00	0.1306918	0.00			
2	0.00	0.3488195	0.00	0.2231975	0.00			
3	0.84	0.8323170	0.17	1.0093122	0.52			
4	0.24	0.8512077	0.20	0.8598376	0.19			
5	0.91	1.2045220	0.87	1.2316222	0.93			
6	0.26	0.9044572	0.30	0.8913178	0.25			

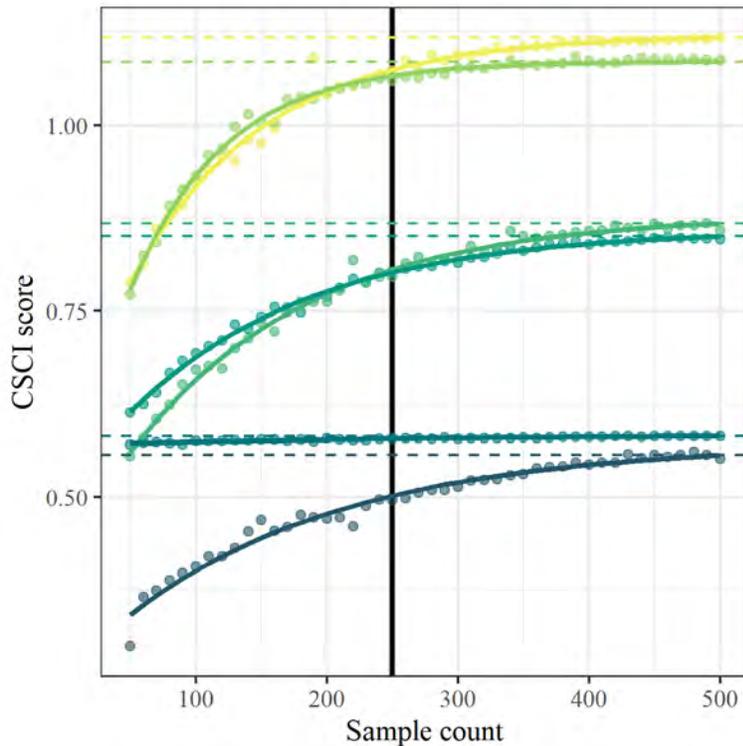
Figure 3. The CSCI core report includes metadata that can be used to evaluate the BMI sample. The total number of organisms is indicated in the "Count" column, highlighted in red.

How do you evaluate the answer?

We recommend a **minimum of 250 organisms** in a sample for CSCI scores to be considered valid.

This recommendation is based on simulation analyses (Figure 4; details provided in Appendix 1) where sample counts were reduced at random, and scores were compared to the original (true) score based on the full 600-count sample. As long as counts were above 250, the estimated score was typically within 10% of the true score, with relatively little variation across simulations. In

general, low count samples had biased, low-scoring CSCI scores (although biased high-scoring samples were also observed in some iterations). Sites with low richness and high evenness were more robust to small counts than were more diverse samples. Additionally, low-scoring samples tend to be more robust than high-scoring samples.



**Figure 4. Effects of systematically reducing sample count on CSCI scores for six sites. The dashed lines indicate the true CSCI score at each site, and the vertical black line indicates a minimum sample size of 250.**

What can you do about it?

If a sample count is below 250, it is best to collect or seek out additional samples from the site. If higher-count samples are available, the lower-count sample should be disregarded, as it likely reflects transient conditions or sampling error. However, if repeated sampling consistently yields low-count samples, this pattern may indicate severe, long-term disturbance (e.g., repeated drying and scouring of flood-control channels) or natural conditions (e.g., certain bedrock-dominated streams) that are intrinsically depauperate in BMI. Further investigation of the site will determine whether the consistently low counts are due to natural or anthropogenic factors. If the low counts are unambiguously attributable to natural circumstances, the low count invalidates the CSCI score; otherwise, the CSCI scores are likely valid.

## Are there too many ambiguous identifications in the sample?

Why is this a problem?

Ambiguous individuals or taxa cannot be used to calculate the O/E component of the CSCI, and may distort calculations of some metrics in the MMI component, likely leading to an underestimate of the CSCI score. Ambiguous taxa may be found in low numbers in any bioassessment sample, but they can create a problem when they comprise a large portion of the organisms in a sample. High proportions of ambiguous taxa may occur if a sample isn't identified to the CSCI's standard level of taxonomic effort (SAFIT1a), or if the sample is dominated by immature or hard-to-identify taxa (e.g., early instar stoneflies).

Ambiguous taxa are excluded from calculating many components of the CSCI, leading to inaccurate scores for the same reasons as described above. Additionally, ambiguous taxa often lack trait information used to calculate certain metrics (e.g., tolerance value, functional feeding group). If all the ambiguous taxa belong to a certain group (e.g., stoneflies), excluding them from metric calculation will mischaracterize the composition of the sample. Therefore, the presence of a high proportion of ambiguous taxa should be considered a separate problem from low counts in a sample.

Where do you find an answer?

The taxonomic identifications for macroinvertebrate samples used to calculate the CSCI are compared against SAFIT's *standard taxonomic effort*. The CSCI output returns information on the percentage of a sample that does not conform to the SAFIT taxonomy, both as the percentage of individuals from the total count that are ambiguous and the percentage of taxa that are ambiguous. Although no maximum number has been established by SWAMP, samples with high percentages of ambiguous taxa may have invalid CSCI scores. Figure 5 shows output from the CSCI calculate that reports the percentage of ambiguous individuals and taxa. The second sample for site 3 has many ambiguous observations.

```

> results$score
  StationCode  SampleID Count Number_of_MMI_Iterations Number_of_OE_Iterations
1      Site3 BadSample1   100                      1                      1
2      Site3 BadSample2   600                      20                     1
3      Site1  Sample1    556                      20                     20
4      Site2  Sample2    826                      20                     20
5      Site3  Sample3    607                      20                     20
6      Site3  Sample4    513                      20                     1
Pcnt_Ambiguous_Individuals Pcnt_Ambiguous_Taxa      E Mean_O      OoverE
1                0.0000000      0.000000 10.248486      1.00 0.09757538
2                83.3333333      50.000000 10.248486      1.00 0.09757538
3                0.5395683      2.631579   7.544418      8.95 1.18630752
4                0.9685230      1.666667 12.953853     11.25 0.86846749
5                9.7199341      6.250000 10.248486     12.90 1.25872246
6                37.6218324     41.025641 10.248486      9.00 0.87817846
OoverE_Percentile      MMI MMI_Percentile      CSCI CSCI_Percentile
1                0.00 0.1638082      0.00 0.1306918      0.00
2                0.00 0.3488195      0.00 0.2231975      0.00
3                0.84 0.8323170      0.17 1.0093122      0.52
4                0.24 0.8512077      0.20 0.8598376      0.19
5                0.91 1.2045220      0.87 1.2316222      0.93
6                0.26 0.9044572      0.30 0.8913178      0.25

```

Figure 5. CSCI metadata that can be evaluated from the standard results. The second sample returns an invalid CSCI score because of many ambiguous individuals and taxa (in red).

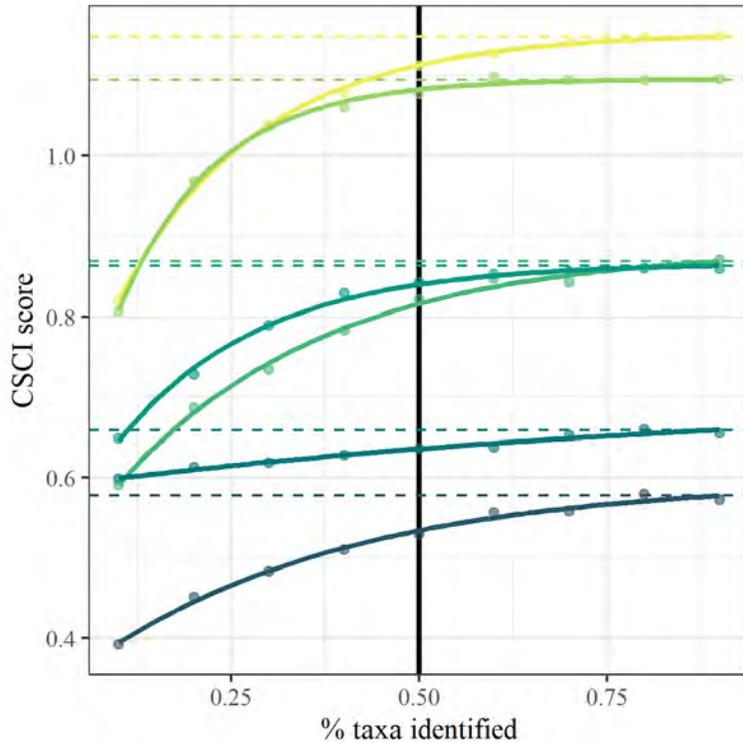
There are several possible causes for a high proportion of ambiguous taxa in a sample:

- Samples were collected very early, well before the normal index period (which for southern California is May 15 to July 15). Samples collected mid-winter (e.g., December to February) tend to have many early-instar taxa, which are difficult to identify to the desired levels.
- The sample was poorly preserved, and specimens were in poor condition. Notes from the taxonomy lab should indicate if this was the case. Field crews should be notified so that they can improve sample preservation practices.
- The taxonomy lab did not apply a level of effort (i.e., SAFIT 1a or 2) required for CSCI calculation. Work orders and chains-of-custody should indicate the level of effort the lab strove for.

Regardless of the cause of ambiguous taxa, it is always better to evaluate the overall severity of the problem before deciding that a sample should be used for CSCI calculation.

How do you evaluate the answer?

We recommend a **maximum of 50% ambiguous taxa or individuals** in a sample for CSCI scores to be considered valid. Higher proportions invalidate a sample.



**Figure 6. Effects of systematically increasing the proportion of ambiguous taxa in a sample on CSCI scores for six sites. The dashed lines indicate the true CSCI score at each site and the vertical black line indicates a recommended maximum of 50% ambiguous taxa.**

This recommendation is based on simulation analyses presented Appendix 2 and in Figure 6, where ambiguous identifications were introduced in to sample at random, and scores were compared to the original (true) score based on the original sample. As long as the proportion of ambiguous taxa or individuals were above below 50%, the estimated score was typically within 10% of the true score, with relatively little variation across simulations.

What can you do about it?

If ambiguous taxa are largely midges (the typical scenario when the taxonomy lab used SAFIT Level 1 as a standardized taxonomic effort level), rescore the sample using the `MissingMidges()` function in the CSCI package. This function assumes that all undetected midge subfamilies are present, providing a defensible upper-end estimate of CSCI scores. This range can then be used to determine if reanalysis of vouchered specimens may be worthwhile.

If ambiguous taxa are largely non-midges, or if the range of possible CSCI scores identified through the `MissingMidges()` function is too large, additional sample collection may be necessary.

## Was the sample influenced by scouring flows?

High flow conditions can scour stream beds and temporarily lower biodiversity at a site. This can produce samples that have lower CSCI scores than would be expected for what the site supports under baseline conditions. By restricting sampling during the index period, bioassessment programs can avoid scour from most natural storms. However, atypical storms or human activities (e.g., dam releases) can create scouring events that could affect a bioassessment sample.

Scour resulting from human activities should be interpreted as an impact. In such cases, the CSCI scores are expected to reflect the impact of an anthropogenic stressor and the sample should be considered valid. Scour from natural storms within 4 to 6 weeks of sample collection may also be considered an impact if they are exacerbated by watershed or channel alteration (e.g., increased runoff).

## Where do you find answers and how are they evaluated?

There are three typical sources of information about scour (in order of decreasing importance for validation): Field observation, stream flow data, and precipitation data.

1. Notes from field crews should indicate if they observed evidence of recent scour (e.g., minimal algae growth, evidence of recent bed movement, etc.). Site photos collected by field crews may provide further evidence.
  - Notes from well trained, experienced, and regularly audited/intercalibrated field crews should be considered **evidence that scour** may have influenced a bioassessment sample.
2. Continuous flow data from a stream gauge (e.g., USGS flow stations). In general, gauges should be close to the sampling location (ideally, the areas of the watershed contributing to the gauge and the watershed contributing to the sampling location should vary by no more than 10%). These data can be used to produce hydrographs, indicating the timing of peak flows, relative to the timing of sample collection. For sites downstream of dams (within 5 km), data on dam releases may provide similar evidence. Figure 7 shows an example of retrieving stream gauge data using the USGS dataRetrieval package for R (De Cicco et al. 2018).
  - Hydrographs that indicate the **sudden onset of flow within 2 weeks of the sampling date** should be considered evidence that scour may have influenced a bioassessment sample, particularly if they are consistent with field observations.

```

library(dataRetrieval)

siteNumber <- "11087020" # San Gabriel mainstem
parameterCd <- "00060" # Discharge
startDate <- "2015-10-01"
endDate <- "2019-09-30"

discharge <- readNWISdv(siteNumber, parameterCd, startDate, endDate)

plot(X_00060_00003 ~ Date, type = 'l', data = discharge, ylab = 'Discharge (ft3/s)')

```

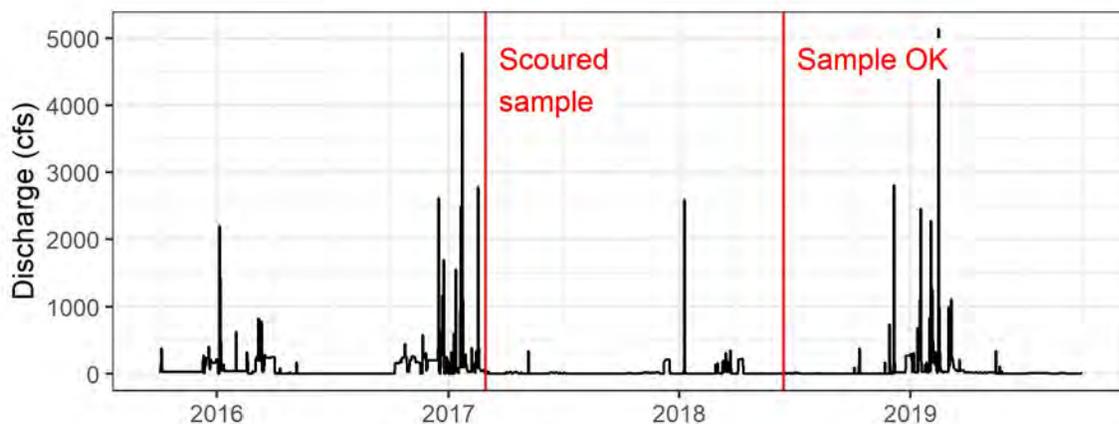


Figure 7. Flow record for four years at the San Gabriel mainstem.

3. Rainfall data (e.g., from NOAA stations) may also provide evidence of scour. However, a small amount of rainfall can lead to extensive scour at some sites, while large storms may have minimal impact on others. Data from nearby gauges (within 10 km in low elevations, or 5 km in mountainous regions) can be used to identify the timing and magnitude of storms, which can then be used to infer if scouring might have occurred.
  - Evidence of **major storms occurring within 2 weeks of the sampling date** should be considered evidence that scour may have influenced a bioassessment sample. As a general rule for the San Gabriel watershed, a major storm is one with **more than 1 inch of precipitation in 24 hours**. However, best professional judgment and local expertise may support different criteria for identifying major storms in some settings. For example, large early winter storms in certain watersheds may not lead to any runoff due to rapid local infiltration, whereas small storms elsewhere may cause substantial scour due to watershed imperviousness or soil hydrophobicity.

What can you do about it?

Additional samples may need to be collected if a sample was affected by scour. Samples collected at different times may provide an appropriate substitute. Nearby sites are unlikely to provide substitutes as they are typically affected by the same scouring events that affected the original site. Be sure to evaluate the number of organisms in these samples (described [here](#)) if scour is suspected.

## **Was the sample influenced by drying?**

Why is this a problem?

Drying events results in the death or aestivation of most aquatic organism, so bioassessment samples collected shortly after a drying event look very different from those collected before drying. This is particularly true in stream reaches lacking aquatic refugia (such as concrete channels), as sensitive organisms may take time to recolonize the reach.

Samples collected from a reach that is partially dry (e.g., surface water is discontinuous and/or stagnant) or shortly after a short-term drying event are known to yield low CSCI scores at reference sites. Therefore, CSCI scores from samples affected by drying are not typically considered valid. However, if the drying is related to human activity (e.g., groundwater pumping, diversions), then the low CSCI scores correctly reflect the impact of this activity.

Note that intermittency alone does not invalidate a CSCI score. Studies from Southern California reference streams have shown that the CSCI and other indices performs in these streams (e.g., Mazor et al. 2014, Loflen 2019). However, there have been too few studies in other parts of the state to determine how the CSCI performs in intermittent streams elsewhere.

Where do you find an answer?

Normal practices defined under the stream sampling SOP (Ode et al. 2016) require that sampling be conducted under baseflow conditions. Field notes should be consulted to determine if flow was abnormally low, possibly as a result of drought or diversions. For example, sampling transects may have been skipped if stream flow was discontinuous on the sampling reach. The field notes should indicate if any deviation occurred from the normal protocol. Sites where the normal sampling protocol was altered may not produce accurate CSCI scores.

Short-term drying events are not always easy to detect. In streams overlying coarse alluvial substrates, streams may dry and re-wet on a diurnal cycle associated with changes in evapotranspiration by riparian vegetation; typically, flows increase overnight, when respiration is lowest, and diminish or cease by mid-afternoon, when respiration rates are highest. Field crews may note evidence of a recent rewetting, such as dried yet submerged algal mats. However, continuous data loggers or data from nearby stream gauges (see section on [accessing gauge data](#)) may provide more conclusive evidence that a stream was dry prior to sampling.

How do you evaluate the answer?

Field notes may indicate if the sampling protocol was modified to accommodate drying. These modifications may invalidate a CSCI score:

- The entire sampling reach length was at least **100 m**
- There were **no dry transects** within the reach
- There was some **evidence of flow** at one or more transect (that is, the entire reach wasn't stagnating).

For example, the August sample from Cedar Canyon shown in Figure 2 is accompanied by the following comment, entered under “SampleComments” in the SWAMP database:

*“Dry transects at C,G and GH. The entire reach was small pools with heavy tule growth. Could not measure flow. This site was entered on a field laptop.”*

Based on this comment, the CSCI score should be considered invalid.

If data from a stream gauge or a logger indicates that a drying event occurred within **1 month** of sample collection, and the drying is entirely due to natural causes (e.g., no diversions), then the CSCI score from the sample should not be considered valid. Similarly, field notes indicating that a stream has recently resumed flow (e.g., observations of submerged dead algal mats) should also be considered as potentially invalidating a CSCI score.

Note that these issues are likely to depress CSCI scores, and the observation of high scores could be taken as evidence that the sample wasn't greatly affected by low flows related to drought.

What can you do about it?

Additional samples may need to be collected if a sample was affected by drying.

### **Was the sample influenced by drought?**

Why is this a problem?

Drought conditions (and the management response to droughts, such as diversions or groundwater extraction) can stress stream communities in several ways, primarily by reducing baseflow conditions. In extreme cases, flow may cease altogether (see section on *drying*). Flow reduction can alter the physical and chemical conditions in the stream, which can adversely impact biological communities (Herbst et al. 2019). For example, reduced flow may lead to lower dissolved oxygen, increased stream temperatures, encroachment of riparian vegetation, concentration of pollutants, and saltwater intrusion in coastal streams. Natural streams have some resilience to drought, particularly those in semi-arid climates such as southern California.

The influence of drought does not invalidate a CSCI score. As described above, drought may exacerbate changes in physical habitat quality or the impacts of anthropogenic stressors. However, determining whether drought has influenced a CSCI score can be useful for assessing overall stream condition.

Where do you find an answer?

There are numerous ways to measure the severity of a drought. Although it is not yet clear which of these indices are most relevant to stream ecology, the Palmer Drought Severity Index (*PDSI*) is one of the more widely used indices. Calculating the index for a given site or sampling date

requires a substantial effort (but is possible with *open source software*, if data on precipitation and evapotranspiration are available), but *weekly drought maps* may provide sufficient information on conditions at the general time and location of sampling.

If sites are visited under multiple years, comparing field notes, site sketches, photos, wetted width measurements, or water quality parameters may indicate if a stream is responding to drought.

How do you evaluate the answer?

Drought conditions should be evaluated relative to the magnitude (how dry) and duration (how long). Drought conditions that are more severe and that persist for longer will have a larger effect on stream health. Negative PDSI scores indicate drought conditions, whereas positive values indicate moist conditions. PDSI scores below -3 indicate severe drought, and PDSI scores below -4 indicate extreme drought. At this time, we cannot identify a threshold minimum PDSI score to determine if drought has influenced a CSCI score.

What can you do about it?

Additional samples may need to be collected if a sample was affected by drought conditions.

### **Was the sample affected by fire?**

Why is this a problem?

Fire events that occur in the watershed or riparian area can dramatically affect CSCI scores. Fire can alter soil chemistry and water runoff characteristics, which in turn affects stream conditions. Sites impacted by fire typically have increases in fine sediment and chemical changes from burned debris or litter that flows downstream. Riparian conditions may also change if vegetation in or around the stream is removed by fire, which can reduce shading and increase stream temperature. As a result, biological integrity is reduced post-fire and may not recover until several years after the fire. An evaluation of fire impacted sites in the Lake Tahoe basin showed that communities did not recover until two years after a fire event (Oliver et al. 2012). In southern California, bioassessment scores at reference sites did not recover until three to four years post-fire, with burned sites having reduced taxonomic diversity and characterized by rapid colonizers and pollution-tolerant taxa (e.g., black flies and minnow mayflies) (Rehn, Ode, and Harrington 2011).

Although fire is a natural phenomenon, an increase in the severity and magnitude may result from climate change or may otherwise be exacerbated by drought conditions. Moreover, evidence of the influence of wildfire does not invalidate a CSCI score, but it may provide useful information for understanding why scores may be low.

Where do you find an answer?

As always, field notes should indicate if conditions at the time of sampling are affected by fire. Additionally, geospatial data can be consulted to view fire perimeter maps (*Cal Fire Hub*) that may have occurred in the watershed for a sampling site. Overlaying the fire perimeter over the watershed boundary can provide an indication of what percentage of the watershed was burned.

There are no clear boundaries for how much of the watershed is burned to determine if the stream sampling site is adversely affected. However, impacts are more likely to be observed if a larger percentage of the watershed was burned and the fire perimeter is closer to the sampling site (e.g., as opposed to higher up in the watershed).

How do you evaluate the answer?

Although the severity of fire impacts can vary greatly from site to site, these guidelines provide a general indication if a sample is likely to be impacted:

- The **sampling reach** was burned within the past **5 years**.
- More than **10% of the contributing catchment within 5 km** of the sampling location was burned within the past **5 years**.
- More than **25% of the contributing catchment** (any distance upstream of the sampling location) was burned within the past **5 years**.

What can you do about it?

Previous studies in southern California show that many burned reference sites have scores that return to pre-burned values in ~3 years (Rehn, Ode, and Harrington 2011). Sites surveyed in the SGR watershed were observed to return to baseline conditions 2.5 to 3 years post-fire. Therefore, we recommend resampling no sooner than three years post burn to characterize baseline stream condition.

### **Was the sample influenced by vegetation management or debris removal?**

Why is this a problem?

Vegetation management, debris removal, and stream regrading can be common management activities in urban streams to improve flood control and speed the passage of stormwater through a reach. However, these activities can have acute, short-term impacts on benthic communities, either through direct habitat removal (as for vegetation removal) or promotion of downstream drift (as for regrading). Herbicides that could be used for vegetation removal may also be non-specific and can harm other stream biota. Although low CSCI scores correctly reflect these impacts, we may want to identify samples that are particularly affected by them, as opposed to samples that reflect baseline conditions. Evidence of these activities does not invalidate a CSCI score.

Where do you find an answer?

Flood control maintenance records provided by public works departments can provide information on the location and types of maintenance activities that could have occurred at a sampling site. Field notes may also indicate conditions that suggest recent maintenance operations have occurred, such as a stream bed that has been recently regraded or habitat conditions that otherwise differ from those that were observed at the previous visit (e.g., vegetation present previous year but absent in the current).

How do you evaluate the answer?

If vegetation or debris removal occurred within **4 weeks** of the sampling event, the CSCI score is likely to be influenced by this activity.

What can you do about it?

Additional samples may need to be collected at a later date after which biological communities are not affected by the vegetation removal.

### **Was the sample influenced by vector control activities?**

Why is this a problem?

Vector control activities are also common in urban streams to control nuisance species that may impact public well-being and health. Pesticides may be applied in some cases, whereas biological controls could be used in others (BTI applications or mosquitofish introductions). In more extreme cases, waterbodies may be diverted or drained to eliminate a water source that acts as a biological vector. In all cases, vector control activities can negatively affect the natural macroinvertebrate community as controls are usually non-specific, causing lower CSCI scores. Evidence of these activities does not invalidate a CSCI score.

Where do you find an answer?

Contact the local vector control district to determine when and where vector control activities occur. Fish surveys or field notes could indicate if mosquitofish are present at a site. Water and sediment chemistry data may also indicate if pesticides are present.

How do you evaluate the answer?

Any evidence of pesticide application or biological control of disease vectors that has occurred **prior to sampling within the sampling season** could potentially influence a CSCI score. A subsequent causal assessment is needed to determine if vector control is the likely cause of a low score.

What can you do about it?

If possible, collect additional samples prior to the implementation of vector control activities. Note that this may not be feasible in most urban settings.

### **Was the sample influenced by tides or naturally high salinity?**

Why is this a problem?

Although evaluations to date suggest that the CSCI works in a wide range of streams with different levels of solute concentrations, it was calibrated for freshwater systems, and should not be used in brackish or saline environments, such as estuaries or inland saline rivers.

Streams in close proximity to coastal areas may be tidally-influenced through groundwater or direct exchange through tidal inlets. Although macroinvertebrates communities can thrive in

tidally-influenced streams, the CSCI was not calibrated for these locations. Many of the taxa expected to occur in wadeable streams cannot withstand the stresses posed by the high and fluctuating salinity levels these waterbodies exhibit, and so even an unstressed stream would likely have a low CSCI score.

Evidence of tidal influence typically invalidates a CSCI score, except when the tidal influence is due to human activity (e.g., excessive groundwater pumping leading to saltwater intrusion). Other benthic indices may be more appropriate for evaluating these samples.

Where to you find an answer?

There are several ways to determine if tidal influence affects a site. Note that most of these methods are only valid near the coast, and they may indicate non-tidal saline influence in waterbodies further inland.

1. In some coastal watersheds tidal influence is controlled by drop structures or tidal gates (e.g., the drop structures near the 405 freeway limit the extent of tidal influence on the San Gabriel River). Samples collected downstream of these controls should be considered tidally influenced.
2. Site is located within *mapped estuarine region*, immediately contiguous with saline waters, and at elevations within twice the spring tide height for the region.
3. Field notes or other local expertise suggesting bi-directional flows from tidal exchange or dominance of halophytic vegetation (e.g., pickleweed, marsh cordgrass, or black needle rush/*Juncus* spp. as a low salinity indicator).
4. The sample has a high abundance of marine or brackish invertebrates, such as polychaete worms (such as *Ficopomatus* or *Nereis*), or certain amphipod genera (e.g., *Americorophium* or *Rammelogrammarus*).
5. Direct measurements of salinity at a sample site can provide an indication of relative tidal influence. Note that tidal conditions vary daily and with precipitation patterns, so a single measure of salinity may not reflect the gradient of conditions at a site. A complete survey of salinity covering the tidal cycle (e.g., over 12 hours) or, at a minimum, one sample occurring at mean high water, will provide a more complete description of the salinity regime.

How do you evaluate the answer?

Any indication that the sample site is tidally-influenced is evidence that the sample is not valid for calculating the CSCI, unless as noted above, the tidal-influence is the result of human activity. Specific conductivity measurements in excess of 10,000 uS/cm could be considered tidally-influenced.

What can you do about it?

Collect additional samples well upstream where the location is not tidally-influenced. Use different sampling protocols and indices.

## Was the sample collected from a setting where the CSCI is suspected to give low scores?

Why is this a problem?

There are some settings in California where CSCI scores have been hypothesized to be unreliable (e.g., watershed underlain by the Monterey formation or similar recent marine sediments), although there are few studies that have thoroughly explored this issue. In these settings, scores may be depressed because the biological community may be naturally low in diversity and the statewide reference pools do not account for these localized exceptions. For example, the geological setting may be uncharacteristic of the region (e.g., unusual geology types with limited extent, Campbell, McCulloh, and Vedder 2009). This setting can influence the physical and chemical characteristics of the stream that structure the diversity of the biological community. This may confound the ability of the CSCI to distinguish between natural and anthropogenic variation, possibly resulting in unreliable scores. Samples collected from such settings may invalidate a CSCI score or require adjustments for correct interpretation.

Where do you find an answer, and how do you evaluate the information?

Evaluating the potential that natural settings introduce bias to the CSCI or other bioassessment indices requires rigorous study, typically following these steps:

1. Review the literature to determine if factors related to the setting are known to influence the abundance or distribution of benthic macroinvertebrates. The more relevant the study (e.g., in California or on taxa found in California), the better.
2. Evaluate scores at reference sites in the unusual setting. If scores are high (e.g., mean close to 1, 90% of scores > 0.79), you have confidence that the CSCI is valid in this setting. Look in *CEDEN* or the *SMC Data Portal* as a preliminary step in finding these sites. Additionally, consult the development data that comes with the CSCI package. Specifically, the `loadRefBugData()` function can be used to view reference taxonomy data or the `loadRefData()` function can be used to view reference site data that were used to build the CSCI. However, keep in mind that if the unusual setting is represented in the CSCI development data set, there is little reason to doubt the applicability of the index in that setting.
3. If reference sites are unavailable to characterize the unusual setting, the least stressed sites should be identified. If the CSCI scores are high at these sites (e.g., mean close to 1, 90% of scores > 0.79), there is confidence that the CSCI is valid in this setting. However, if these scores are lower, the CSCI may or may not be valid at identifying reference conditions in this setting. Regardless, the scores may still be valid for evaluating relative condition. Consider evaluating if CSCI scores respond negatively to stressor gradients within this setting. If there is large correlation (e.g.,  $R^2 > 0.2$ ), the CSCI is likely responding to stress in this setting. Note that if the “least stressed” sites representing the setting are still severely stressed, it may be hard to interpret results from this study with certainty.

What can you do about it?

In some cases, the studies recommended above may suggest a correction or numerical adjustment to apply to a CSCI score at sites in these settings. However, we recommend that any such adjustments be done in consultation with the SWAMP bioassessment workgroup, particularly for regulatory applications of CSCI scores.

## **Is the watershed delineation used to calculate CSCI predictors appropriate?**

Why is this a problem?

The CSCI requires data describing landscape characteristics of the watershed for a site. These data are used to develop a prediction of the macroinvertebrate community that could be expected at the site under reference conditions. A watershed delineation is required for a site to obtain these landscape data for the CSCI predictions. The CSCI interim instructions (Mazor et al. 2018) describe in detail how these delineations can be created. In short, a digital elevation model is used with the site's longitude/latitude to identify the area of land where all elevations are increasing and higher than the starting elevation of the site. This watershed is then used to calculate landscape-level data needed for the CSCI, such as the total elevation range, average precipitation, and various soil characteristics. An inaccurate representation of the watershed can produce inaccurate estimates of the landscape data used to calculate the CSCI.

The watershed delineation process is partially automated using standard geospatial software, with some intervention and manual inputs from the analyst. In general, delineations will accurately represent the watershed at the site if:

- The actual site location is spatially co-located with a stream reach line in a GIS, and vice versa.
- The actual drainage area has sufficient topographic relief to support elevation-based delineation processes.

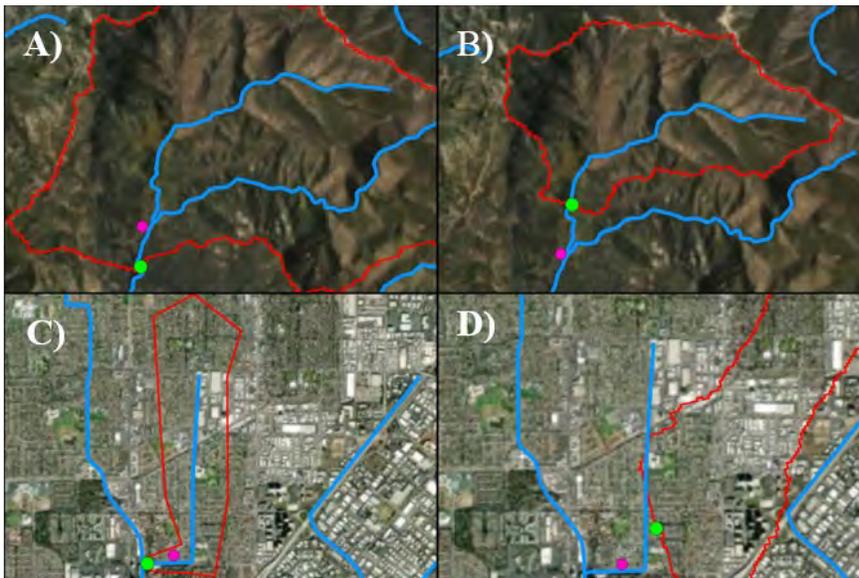
For the first scenario, the site location is typically referenced by longitude/latitude coordinates. For delineation, these coordinates must be spatially linked to a stream reach in a GIS. Stream reaches are usually represented by the NHD-Plus dataset (McKay et al. 2012), which is a national-level product describing stream hydrography for the entire United States. The first step in the delineation is to “snap” the site location to the nearest stream reach. If the site location is imprecise or was entered incorrectly, the snapping distance can be large. Conversely, the stream reach in the NHD-Plus dataset may not accurately portray the true channel. In either case, the resulting watershed will originate from a location that does not represent reality. Visual assessment of the site location, the segment that was used for the delineation, and the snapping distance can provide clues about the quality of the delineation.

For the second scenario, topographical characteristics of the landscape around a site can also affect the quality of the delineation. In general, watershed boundaries are more easily identified at high gradient sites in hilly or mountainous areas where topographical variation is more pronounced. Conversely, low gradient streams may have less accurate watershed delineations because it is more difficult to identify clear elevation differences that define drainage patterns. The latter scenario is more common in coastal plains, plateaus, or other low-topography areas.

Overlaying the watershed delineation on aerial photos can provide clues if the delineation is well-represented by topography.

How do you evaluate the answer?

Figure 8 provides some examples of how snapping and quality of the delineation can produce inaccurate watershed polygons. In all four plots, the pink dot represents the latitude/longitude of the recorded site location and the green dot represents the location where the site was snapped to the stream flow-line for the delineation. In Figure 8a, we see a snap that produced a likely realistic representation of the watershed that drained to the site. However, in Figure 8b, we see a problematic snap where the location was shifted upstream to a tributary. In this case, the watershed is an under-representation of what drains to the site. In Figure 8c, we see a site that was manually snapped to a location and the watershed was manually delineated. In Figure 8d, we see the same site but the snap location is likely incorrect and the resulting watershed is likely inaccurate. These final two examples represent the challenges of watershed delineation in developed settings, where manual changes may be needed to create a more realistic interpretation of the watershed.



**Figure 8. Watershed delineation examples showing snapping challenges. The pink dot is the recorded site latitude/longitude and the green dot is the snapped location. The watershed boundaries resulting in each case are outlined in red.**

What can you do about it?

Manual editing of the polygon may be needed if the delineation is inadequate. Alternatively, follow the procedures in the CSCI interim instructions (Mazor et al. 2018) to delineate a new watershed. In both cases, the GIS predictors obtained from the delineation must be calculated again prior to calculating a new CSCI score.

## Are GIS predictors correct?

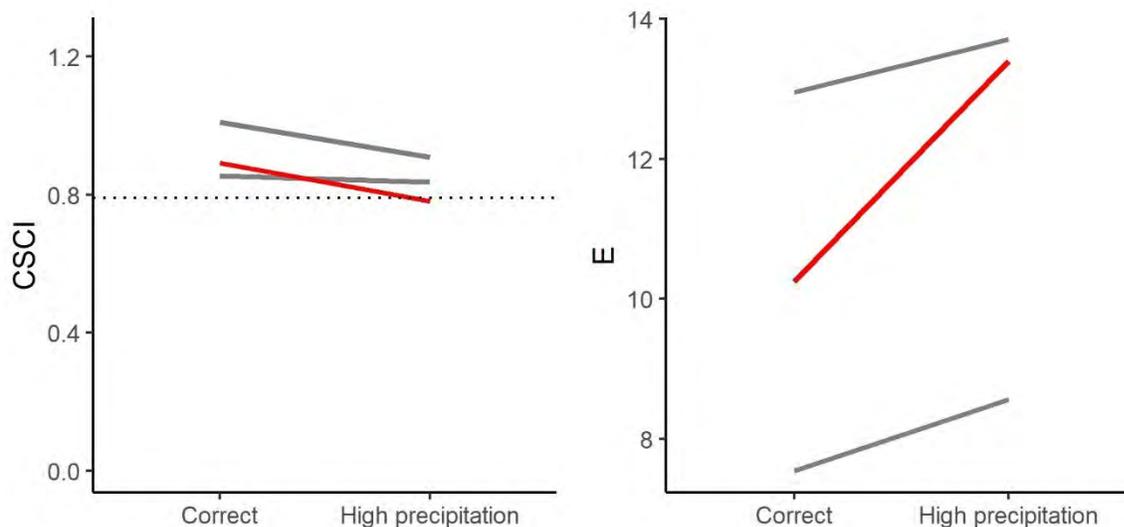
Why is this a problem?

GIS predictors are required for calculating CSCI scores. The CSCI interim instructions (Mazor et al. 2018) provide information on which GIS metrics are used and how they can be calculated.

These predictors include:

- site latitude/longitude
- site elevation
- elevation range
- watershed area
- average precipitation
- average temperature
- mean June to September monthly precipitation
- average bulk soil density
- average soil erodibility factor
- average phosphorus geology

There's little chance for errors to arise in calculating predictors, as long as the instructions are followed. Yet errors could greatly alter expected values used in the CSCI, leading to incorrect scores. For example, increasing precipitation by a factor of 10 in the example data set built in the CSCI package resulted in lower CSCI scores across the board, largely driven by inflated expectations of higher diversity (Figure 9).



**Figure 9. Changes in CSCI scores (left) and E (number of expected taxa, right) when correct predictors are used versus inflated precipitation. The red line indicates a sample where changes were large enough to drop the score below a threshold of 0.79 (dotted line, left).**

Where do you find an answer?

The predictor data used to calculate the CSCI score isn't provided with standard metadata outputs, they but can be accessed from the [SMC Data Portal](#). Currently, the SWAMP and CEDEN databases do not store this information.

How do you evaluate the answer?

As mentioned above, there are few opportunities to introduce errors to these predictors. However, comparing predictor data at your site to typical values for a region may indicate if errors should be suspected. Appendix 3 provides typical ranges of predictors found in the South Coast region.

What can you do about it?

If you suspect that predictor data used to calculate CSCI scores are incorrect, we recommend recalculating these metrics following the instructions in Mazor et al. (2018) and re-scoring samples.

## **EVALUATING THE ADEQUACY OF THE SCAPE LANDSCAPE MODEL**

As with the CSCI, the SCAPE landscape model underpins a number of decisions in monitoring and management programs like the SGRAMP. Therefore, evaluating its appropriateness on a site-by-site basis can increase the confidence in these decisions.

The SCAPE model calculates the likely range of CSCI scores, based on the extent of alterations (e.g., urban or agricultural land cover, road density) in the watershed as well as in the local catchment. By comparing these ranges to a decision-point threshold (e.g., a CSCI threshold of 0.79), stream channels may be classified as “constrained” (i.e., streams that are unlikely to meet management goals due to large-scale alteration) or “unconstrained (i.e., streams that are likely to meet management goals. For the SGRAMP, four classes of streams, as described below:

- Likely constrained: 90% of the range of likely scores is below 0.79.
- Possibly constrained: 50% of the range of likely scores is below 0.79
- Possibly unconstrained: 50% of the range of likely scores is above 0.79
- Likely unconstrained: 90% of the range of likely scores is above 0.79

The questions described in this section may result in the invalidation of the SCAPE model's predictions for a given site; in contrast, invalidation is rarely an outcome of the questions concerning CSCI scores, described above. Reasons for concern about the SCAPE model arise when the underlying data used in the model (specifically, the NHD Plus and StreamCat) inaccurately characterize local conditions. Although these data sources are correct for the vast majority of streams in California, inaccuracies are known, and can lead the SCAPE model to produce incorrect predictions about the likely range of CSCI scores.

## Is the sampling reach atypical of the channel's overall constraint class?

Why is this a problem?

Typically, stream reaches of the same class cluster in space. In other words, a constrained reach is typically surrounded by other constrained reaches, and an unconstrained reach is typically surrounded by other unconstrained reaches. Occasionally, a stream segment will have a constraint class that differs or is otherwise unexpected based on the classes for reaches nearby. For example, an unconstrained reach may be found in an urban setting where reaches upstream and downstream are constrained. This could reflect a real phenomenon or could result from inaccuracies in the land use data. In these cases, the constraint class should be investigated.

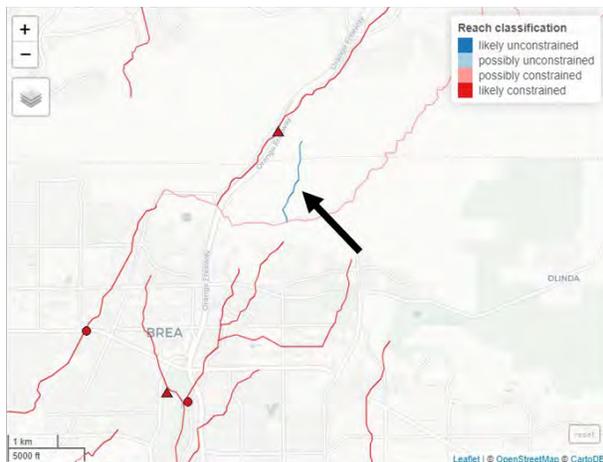
Where do you find an answer?

Viewing an aerial image of land use for a site is the easiest way to assess the validity of an unexpected stream class. An online *SCAPE application* provides this information for the San Gabriel River watershed.

How do you evaluate the answer?

As an example, Figure 10 was taken from SCAPE and shows a stream reach that is assigned a class of likely unconstrained. All of the surrounding reaches are possibly or likely constrained. Without looking at the land use, we might assume that this constraint class is invalid (Figure 10a). We can toggle the base layer to show a satellite image of the location to get a better idea of the landscape (Figure 10b). From the satellite image we can see that this reach drains a small undeveloped, hilly area upstream of the housing units. With this information we can assume that the constraint class is valid because it accurately reflects land use in the watershed.

(a) An unexpected constraint class



(b) Changing the base layer to see land use/land cover



Source: <https://sccwrp.shinyapps.io/SCAPE>

**Figure 10. An unexpected stream class is validated by examining the land use/land cover base layer.**

What can you do about it?

See the section titled “*What to do if the SCAPE model outputs are invalidated*” below.

## Has land cover changed since the SCAPE model was calibrated?

Why is this a problem?

The SCAPE model provides a range of likely CSCI scores based on the landscape characteristics of the watershed upstream of a site. The landscape characteristics are based on national-level, geospatial data products that characterize the relative extent human development in the watershed. Specifically, the landscape model is based on StreamCat data (Hill et al. 2016) that provide estimates of canal/ditch density, imperviousness, road density/crossings, and urban/agricultural land use for each site. Within StreamCat, many of these estimates were derived from primary data products, such as the National Land Cover Database for 2006 and 2011 (Table 2). Because some of the primary products relate to a specific year, the associated constraint classes from the landscape model may not accurately reflect present-day conditions.

**Table 2. Land use variables used to develop the landscape model. All variables were obtained from StreamCat (Hill et al. 2016) and applied to stream segments in the National Hydrography Dataset Plus (NHD-plus) (McKay et al. 2012). Ws: watershed. Cat: local catchment. Ws + Rp100: 100-m riparian buffer at the watershed scale. Cat + Rp100: 100-m riparian buffer at the local catchment scale (See Hill et al. 2016 for explanations of these spatial scales).**

Variable	Scale	Description	Unit
CanalDens	Cat, Ws	Density of NHDPlus line features classified as canal, ditch, or pipeline	km/sq km
PctImp2006	Cat, Ws, Cat + Rp100, Ws + Rp100	Mean imperviousness of anthropogenic surfaces (NLCD 2006)	%
TotUrb2011	Cat, Ws, Cat + Rp100, Ws + Rp100	Total urban land use as sum of developed open, low, medium, and high intensity (NLCD 2011)	%
TotAg2011	Cat, Ws, Cat + Rp100, Ws + Rp100	Total agricultural land use as sum of hay and crops (NLCD 2011)	%
RdDens	Cat, Ws, Cat + Rp100, Ws + Rp100	Density of roads (2010 Census Tiger Lines)	km/sq km
RdCrs	Cat, Ws	Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines)	crossings/sq km

Where do you find an answer?

Historical imagery is a great way to assess changes in land cover. *Google Earth*'s time slider provides a convenient way to view this imagery. The slider can be used to view a current image

and any of a number of images of land use and cover for the past twenty years. Figure 11 shows a possibly unconstrained stream in Orange County. The image on the left, taken in 2007, shows a moderate amount of landscape alteration associated with agriculture and low-density residential development. The image on the right, taken in 2019, shows extensive land clearing associated with planned high-density development. It is possible that updated land cover data would lead to a different classification for this stream segment.



**Figure 11. An example showing an unconstrained stream in Orange County that has undergone extensive land cover change. The image on the left is from 2007, and the image on the right is from 2019.**

How do you evaluate the answer?

There is no quantitative approach to verify if the constraint class accurately reflects the current landscape. The constraint class is typically an accurate representation of the current landscape because land use changes that affect stream biology usually occur over time scales much longer than would be expected between present day and the data used to create the model, and that much of the landscape alteration in California occurred before the data in StreamCat was compiled. However, in some cases, local alteration of the landscape can occur rapidly and at a scale sufficient to affect stream condition. For example, construction of a parking lot adjacent to a stream channel could alter drainage patterns sufficiently to affect stream health. If there is sufficient evidence that recent changes may be affecting biology and that the current constraint class is not an accurate representation of biological expectations, additional data may be consulted, or an alternative classification could be used.

What can you do about it?

See the section titled “*What to do if the SCAPE model outputs are invalidated*” below.

## Could the site be influenced by constraints not captured by the model?

Why is this a problem?

The SCAPE model is based on a relatively small set of factors that characterize alterations of the landscape with the potential for long-lasting impacts to stream condition (Table 2). Other factors that could impose similar long-term constraints are not explicitly included in the model (such as dams, or legacy impacts of mining operations or timber harvesting). Because the vast majority of these impacts are closely correlated with factors included in the model (especially road density), their exclusion is rarely a cause for concern. However, it's worth investigating potential constraints, even at sites considered by the SCAPE model to be unconstrained.

One example of such a stream where the model doesn't reflect likely constraints is Cañada del Diablo, a small tributary of the Ventura River (Figure 12). Much of the watershed has been disturbed by oil extraction activities, yet because the overall road density is low, the stream is classified as "possibly unconstrained".



**Figure 12. Data in the SCAPE model may underestimate landscape alteration, as seen in Cañada del Diablo in Ventura County. Oil extraction activity has greatly altered the landscape (evident as clearings in the satellite imagery), yet only a fraction of these disturbed areas is recorded as non-natural land types (red pixels) in the landcover data in StreamCat.**

Where do you find an answer?

Consult aerial maps to evaluate current land use and potential stressors that could be affecting the site. For legacy impacts, evaluate historical data to identify stressors that may have had long-term effects on biological condition.

How do you evaluate the answer?

Determining that a human activity or alteration to the landscape constitutes a constraining factor is a decision that requires extensive local knowledge as well as experience with bioassessment data. We recommend close consultation with local stakeholders as well as the SWAMP Bioassessment Workgroup when determining if a factor should be considered a constraint.

What can you do about it?

See the section titled “*What to do if the SCAPE model outputs are invalidated*” below.

### **Does the NHD Plus hydrography accurately represent the stream channel?**

Why is this a problem?

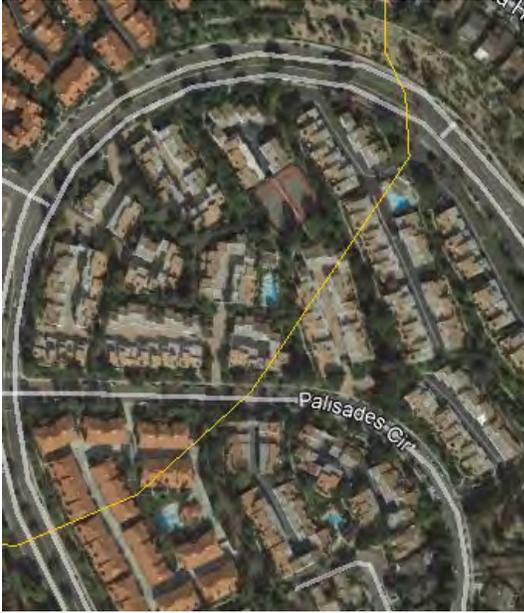
The landscape model assigns a constraint classification to every NHD Plus stream segment where StreamCat data are available. Consequently, the SCAPE model outputs are only as good as the spatial representation of stream locations in the hydrography dataset. Although most inaccuracies in the NHD Plus are relatively minor, a few may be consequential.

Where do you find an answer?

Comparing NHD Plus flowlines to aerial imagery is the best way to evaluate the accuracy of the spatial data underpinning the SCAPE model.

How do you evaluate the answer?

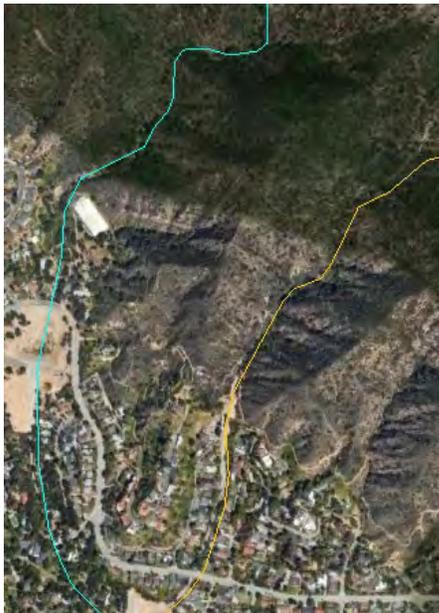
A few common inaccuracies in the NHD Plus are illustrated in Figure 13. Some errors may be consequential (e.g., the bottom two photos).



There is no channel near the NHD Plus flowline. It is likely that the stream has been moved or put into a pipeline beneath the development. The inaccuracy invalidates the SCAPE model output, because the stream no longer exists.



The stream has migrated away from the flow line. In this case, the inaccuracy is not significant, as the flowline characterizes a similar setting to the stream's true location. Constraint classifications won't be affected.



The stream segment crosses a boundary between developed and undeveloped areas. The streams in the above example likely experience few constraints in the upstream portion and more constraints in the lower portion.



The stream segment is part of a braided complex. In these cases, the outer segments correctly reflect the impact of surrounding development, whereas the inner segments inaccurately appear to be surrounded by natural land cover.

**Figure 13. Four examples of spatial errors in the NHD Plus. Light blue flowlines represent possibly unconstrained streams, and orange flowlines represent possibly constrained streams.**

What can you do about it?

See the section titled “*What to do if the SCAPE model outputs are invalidated*” below.

### **Are the results of the SCAPE model close to a key threshold?**

Why is this a problem?

As described *above* for the CSCI, many management decisions rely on bright-line thresholds between classes. If the SCAPE model outputs are very close to the thresholds, uncertainty may arise about the accuracy of the classification. Although the SCAPE model is not affected by sampling variability in the way that CSCI scores are, it can be influenced by relatively minor inaccuracies in underlying landcover and hydrographic data.

Where do you find an answer?

In the example of the SGRAMP, key decisions are based on the 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> percentile outputs of the SCAPE model. The closeness of these distribution-points to the threshold (e.g., 0.79) should be evaluated.

How do you evaluate the answer?

If the distribution point is within **0.01** points of the threshold (e.g., the 90<sup>th</sup> percentile CSCI score for a stream segment is predicted to be 0.78), you should consider the classification to be uncertain and look for additional evidence to support or refute it.

What can you do about it?

See the section titled “*What to do if the SCAPE model outputs are invalidated*” below.

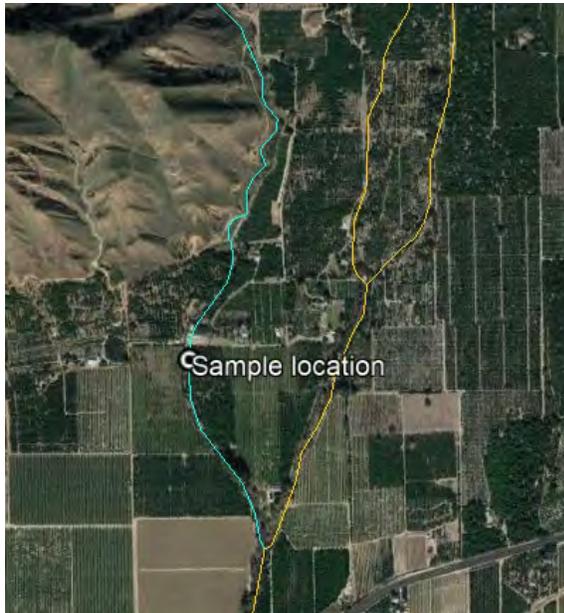
### **What to do if the SCAPE model outputs are invalidated**

Based on the evaluations described above, the SCAPE model outputs may be determined to be invalid for a given stream segment. It is also possible that, due to deficiencies in the NHD Plus or StreamCat, the SCAPE model provides no estimates of constraints for a segment where these estimates are needed (e.g., stream segments missing from NHD Plus). Approximately 15% of the segments in the San Gabriel River watershed were unclassified by the SCAPE model, requiring an alternative approach for estimating constraints. Several options are described below.

#### **Manual substitution**

The SCAPE model may provide appropriate estimates for adjacent stream segments. For example, a downstream segment may more accurately reflect landscape conditions than the segment that was actually sampled. These segments generally become obvious during the data evaluation process. Figure 17 shows an example of a sample collected from an agricultural stream. The classification of “possibly unconstrained” was determined to be invalid because it did not reflect the local conditions appropriately. Constraints estimated from the adjacent

tributaries provide a more appropriate range of likely CSCI scores and could be used for making management decisions about the site in question.



**Figure 14. An example where manual substitution may provide appropriate estimates of likely ranges of CSCI scores.**

In some cases, estimating constraints from the average of multiple nearby reaches may be more appropriate than substituting estimates from a single reach.

### **Applying model averages for stream-types**

We assigned biological expectations to unclassified segments in typically urban or agricultural segments by estimating the range of expectations for segments with similar land use. This analysis was conducted statewide and stratified by major regions to account for statewide variation in land use. The approximate range of CSCI scores in unclassifiable segments were defined for three different groups: segments dominated by either 1) urban, 2) agricultural, or 3) open (i.e., lack of urban or agriculture) land use. The three groups were identified using k-means clustering of percentage land use estimates that were available across segments (MacQueen 1967). This created groups of segments with similar land use types, where membership of a segment within a particular group was based on the minimum difference in land use estimates for a segment from the group average for each land use type (within-group centroid). The two groups that were dominated by agricultural or urban land use were identified based on the largest centroid average of the clusters for each land use type. The third “open” group that was defined by a lack of urban and agricultural land use was identified by the minimum sum of the centroid values for the two land use types. The expected range of CSCI scores for the three groups were based on averages from the landscape model for segments with available predictions.

Ranges of expected CSCI scores for typical segments in urban, agricultural, and undeveloped catchments in Southern California are shown in Table 3 (other regions are presented in Appendix 4). Where appropriate, these values can be used to substitute missing or invalid ranges derived from the SCAPE model. For example, the site shown in Figure 14 may be better characterized by the range of scores for agricultural streams shown in Table 3 (e.g., 0.41 – 1.01 for high-certainty estimate) than the ranges estimated for the actual segment where the site is located (e.g., 0.46 – 1.09).

**Table 3. Ranges of expected CSCI scores for sites that are typically urban, agricultural, or open (neither urban nor agricultural) land uses in the South Coast region of California.**

Land use	High certainty (10th - 90th)	Moderate (25th - 75th)	Low certainty (40th - 60th)
Urban	0.30 - 0.76	0.40 - 0.66	0.48 - 0.57
Agricultural	0.41 - 1.01	0.53 - 0.90	0.63 - 0.78
Open	0.83 - 1.15	0.93 - 1.08	0.98 - 1.04

### **Applying constraints from empirical distributions (e.g., engineered channels in Southern California)**

In certain cases, it may be possible to estimate ranges of likely CSCI scores by evaluating available data. For example, ranges of scores in engineered channels in Southern California were evaluated by the Stormwater Monitoring Coalition (SMC), and found that fully hardened channels invariably had scores below 0.79 (Stormwater Monitoring Coalition 2017). The distribution points reported in that study could be used to estimate ranges at other engineered channels (Table 4). Based on the SGRRMP classification system described above, fully hardened channels would be considered likely constrained, whereas partially earthen channels would be considered possibly constrained. As with the ranges shown in Table 3, these ranges may be used to estimate constraints where SCAPE model outputs are unavailable or invalid. However, we do not recommend using them outside Southern California without additional analysis.

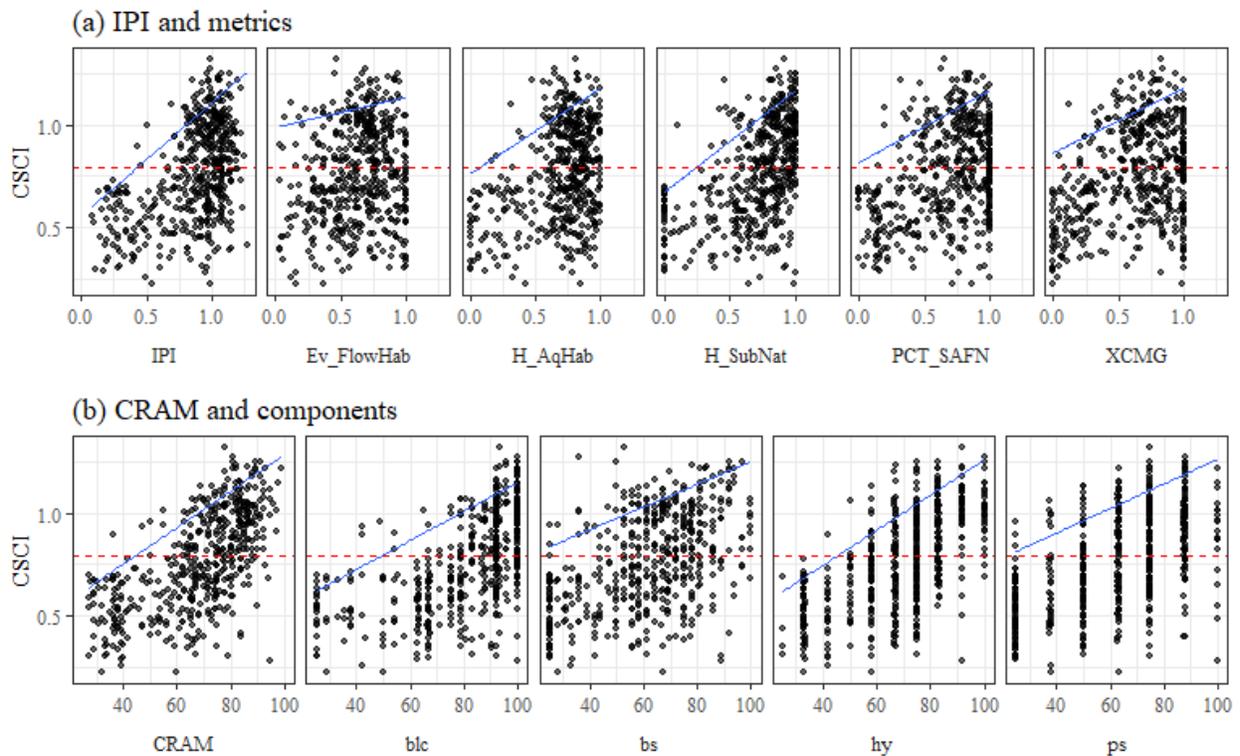
**Table 4. Ranges of CSCI scores reported for Southern California engineered channels.**

Type	High certainty (10th - 90th)	Moderate (25th - 75th)	Low certainty (40th - 60th)
Fully hardened	0.28 - 0.66	0.35 - 0.60	0.44 - 0.53
Partially earthen	0.41 - 0.84	0.49 - 0.70	0.57 - 0.64

## Applying constrains based on habitat condition

Physical habitat information is typically collected alongside other bioassessment data. These data could include Index of Physical Habitat Integrity scores (IPI; Rehn et al. 2018) or California Rapid Assessment Method scores (CRAM; California Wetlands Monitoring Workgroup 2013), as well as their individual metrics or components, respectively. Although habitat quality alone should not be considered a constraint without a full understanding of the factors that have led to habitat degradation at a site, it may be used to provide additional evidence about likely ranges of CSCI scores at a site, which may provide additional information to outputs from the SCAPE model.

An analysis of approximately 500 sites from southern California shows that habitat quality may limit the likelihood of observing high CSCI scores (Appendix 5). A comparison of physical habitat data with CSCI scores below provides some indication of when physical condition may be sufficiently poor and when an impacted CSCI score may be observed (e.g., below CSCI = 0.79, Figure 15). The blue lines are the quantile regression estimates for the 90th percentile. Where these regressions intersect the dotted line could be an indication of when CSCI scores are well below 0.79 for the corresponding habitat measure.



**Figure 15.** CSCI scores compared to IPI and CRAM physical habitat measures. The blue lines are quantiles regression results for the 90th percentile and the red dotted line is CSCI at 0.79. IPI: Index of Physical Habitat Integrity. Ev\_FlowHab: Evenness of flow habitats. H\_SubNat: Diversity of natural substrate types. PCT\_SAFN: Percent sands, fines, and concrete in the substrate. XCMG: Mean riparian vegetation cover. CRAM: California Rapid Assessment Method. Blc: CRAM buffer and landscape condition attribute. Bs: CRAM biotic structure attribute. Hy: CRAM hydrologic structure attribute. Ps: CRAM physical structure attribute.

The regression lines in Figure 15 could be used to estimate an upper limit of likely CSCI scores, using the equations shown in Appendix 5.

## HELPFUL RESOURCES

The following is a list of resources that can provide information to address the questions described in this document, with a special focus on the San Gabriel River watershed.

- CSCI metadata- consult CSCI SOP in (Mazor et al. 2018) and package documentation on [GitHub](#).
- [San Gabriel SCAPE](#) website
- Reference site information - check the `loadRefData()` or `loadRefBugData()` in the CSCI package
- USGS stream gauge data: can be downloaded using the [dataRetrieval](#) R package (De Cicco et al. 2018) or manually from the [NWIS](#) website
- Open-source [scripts](#) to calculate drought severity indices
- GIS data sources:
  - The [SWAMP bioassessment geodatabase](#) includes a wealth of layers that characterize both natural and anthropogenic factors. This database is used for a number of bioassessment applications, including calculating CSCI predictors and screening reference sites.
  - [StreamCat](#) watershed data
  - [NHD Plus](#) hydrography layers
  - Hydrologic data from [StreamStats](#)
  - GIS metrics in the [SMC Data Portal](#)
  - [Google Earth](#) aerial imagery and time slider
- Field Data
  - [SMC Data Portal](#)
  - [CEDEN](#) (includes SWAMP data)
- Other data sets
  - [Weather conditions](#) from NOAA
  - [Stream gauges](#) maintained by the USGS
  - [Drought maps](#)
  - [Fire perimeters](#)
  - [Mining data](#)
  - [Timber harvesting](#) data from USDA Forestry Service

## CITED LITERATURE

Beck, M.W., R.D. Mazor, S. Johnson, K. Wisenbaker, J. Westfall, P.R. Ode, R. Hill, C. Loflen, M. Sutula, and E.D. Stein. 2019. "Prioritizing Management Goals for Stream Biological Integrity Within the Developed Landscape Context." *Freshwater Science* 38 (4).

California Wetlands Monitoring Workgroup. 2013. *California Rapid Assessment Method (CRAM). Riverine User's Manual and Field Book*.

Campbell, R.H., T.H. McCulloh, and J.G. Vedder. 2009. "The Miocene Topanga Group of Southern California - a 100-Year History of Changes in Stratigraphic Nomenclature." Open-File Report 2007-1385, v. 1.1. Reston, Virginia: US Geological Survey.  
<https://pubs.usgs.gov/of/2007/1385/>.

De Cicco, L.A., D. Lorenz, R.M. Hirsch, and W. Watkins. 2018. *DataRetrieval: R Packages for Discovering and Retrieving Water Data Available from U.S. Federal Hydrologic Web Services (version 2.7.5)*. Reston, VA: U.S. Geological Survey; U.S. Geological Survey.  
<https://doi.org/10.5066/P9X4L3GE>.

Herbst, D.B., S.D. Cooper, R.B. Medhurst, S.W. Wiseman, and C.T. Hunsaker. 2019. Drought ecohydrology alters the structure and function of invertebrate communities in mountain streams. *Freshwater Biology*. 64: 1-17. DOI: 10.1111/fwb.13270

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* 52: 120–28.  
<https://doi.org/10.1111/1752-1688.12372>.

Loflen, C. 2019. *Biological Objectives for the San Diego Region. Draft report to the Regional Water Quality Control Board – San Diego Region*. San Diego, CA.  
[https://www.waterboards.ca.gov/sandiego/water\\_issues/programs/basin\\_plan/bio\\_objectives/doc/R9\\_Biological\\_Objectives\\_Staff\\_Report\\_Feb2019.pdf](https://www.waterboards.ca.gov/sandiego/water_issues/programs/basin_plan/bio_objectives/doc/R9_Biological_Objectives_Staff_Report_Feb2019.pdf)

MacQueen, J. 1967. "Some Methods for Classification and Analysis of Multivariate Observations." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, edited by L. M. Le Cam and J. Neyman, 281–97. Berkeley, California, USA: University of California Press.

Mazor, R.D., P.R. Ode, A.C. Rehn, M. Engeln, M.W. Beck, T. Boyle, E. Fintel, and C. Yang. 2018. "The California Stream Condition Index (CSCI): Interim Instructions for Calculating Scores Using GIS and R." SWAMP-SOP-SB-2015-0004, Revision date August 1, 2018. Sacramento, California: California State Water Resources Control Board Surface Water Ambient Monitoring Program.  
[https://www.waterboards.ca.gov/water\\_issues/programs/swamp/bioassessment/docs/CSCI\\_Instructions\\_08\\_01\\_18.pdf](https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/CSCI_Instructions_08_01_18.pdf).

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

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

Ode, P.R., A.E. Fetscher, and L.B. Busse. 2016. “Standard Operating Procedures (SOP) for the Collection of Field Data for Bioassessments of California Wadeable Streams: Benthic Macroinvertebrates, Algae, and Physical Habitat.” SWAMP-SOP-SB-2016-0001. Sacramento, California: California State Water Resources Control Board Surface Water Ambient Monitoring Program.

[https://www.waterboards.ca.gov/water\\_issues/programs/swamp/bioassessment/sops.html](https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/sops.html).

Oliver, A.A., M.T. Bogan, D.B. Herbst, and R.A. Dahlgren. 2012. “Short-Term Changes in-Stream Macroinvertebrate Communities Following a Severe Fire in the Lake Tahoe Basin, California.” *Hydrobiologia* 694 (1): 117–30. <https://doi.org/10.1007/s10750-012-1136-7>.

Rehn, A.C., P.R. Ode, and J.M. Harrington. 2011. “The Effects of Wildfire on Benthic Macroinvertebrates in Southern California Streams.” San Diego, California: San Diego Regional Water Quality Control Board.

[https://www.waterboards.ca.gov/water\\_issues/programs/swamp/docs/reglrpts/rb9postfire\\_tech.pdf](https://www.waterboards.ca.gov/water_issues/programs/swamp/docs/reglrpts/rb9postfire_tech.pdf).

Rehn, A.C., R.D. Mazor, and P.R. Ode. 2018. An index to measure the quality of physical habitat in California wadeable streams. SWAMP Technical Memorandum SWAMP-TM-2018-0005. Sacramento, California.

[https://www.waterboards.ca.gov/water\\_issues/programs/swamp/bioassessment/docs/physical\\_habitat\\_index\\_technical\\_memo.pdf](https://www.waterboards.ca.gov/water_issues/programs/swamp/bioassessment/docs/physical_habitat_index_technical_memo.pdf)

Stormwater Monitoring Coalition. 2017. 2015 Report on the Stormwater Monitoring Coalition Regional Stream Survey. SCCWRP Technical Report 963. Southern California Coastal Water Research Project. Costa Mesa, CA.

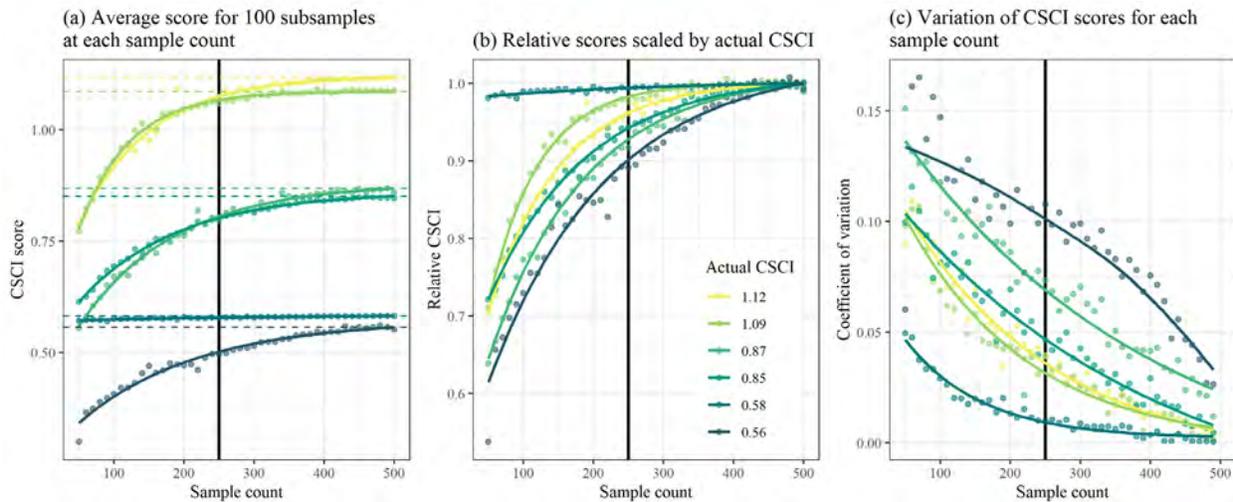
[http://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/963\\_2015\\_SMC\\_Report\\_EnginChannels.pdf](http://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/963_2015_SMC_Report_EnginChannels.pdf)

## APPENDICES

### Appendix 1

#### Evaluating sample-count effects on CSCI scores

An analysis of the effect on CSCI scores of systematically reducing the sample count well below 600 individuals for several sites is provided in the appendix. Figure 16 was created by taking subsamples of the total sample size for six different sites with a range of CSCI scores (horizontal dashed lines in Figure 16a). For each sample count, 100 subsamples were randomly selected from the total and CSCI scores were summarized by the average and coefficient of variation. Overall, reducing the sample size caused reductions in the CSCI scores, with the reductions increasing more quickly with smaller sample sizes. Figure 16b shows the relative change as a proportion from the actual CSCI score. The CSCI score is within ten percent of the actual score with sample counts of around 250 or more. CSCI scores were reduced by greater than ten percent of the actual score with lower sample counts, the exception being a site with very low diversity. The variation of CSCI scores for each sample count also increases with lower sample counts (Figure 16c), although variation did not exceed ten percent until very low sample sizes (e.g., 150 or less).



**Figure 16. Effects of reducing sample counts on CSCI scores**

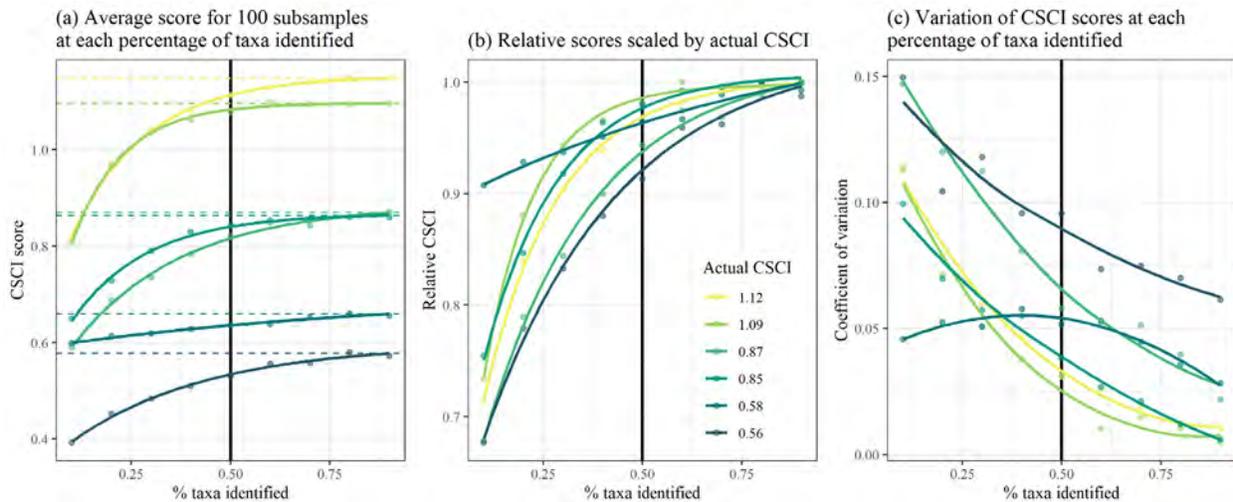
Based on the above analyses, we recommend a minimum sample size of 250 for a valid sample. Detailed recommendations are as follows:

- CSCI scores are generally within ten percent of the actual with sample counts of 250 or more
- Changes in CSCI score with lower samples are similar for high or low quality sites, however;
- Sites with very low scores and very low richness are minimally affected by changes in sample counts.
- Precision decreases with lower sample size, although variation is typically less than 10% of the true mean with sample sizes of 200 or more.

## Appendix 2

### Evaluating the effects of ambiguous identifications on CSCI scores

We simulated the effects of high numbers of ambiguous taxa by replacing individuals that had known identifications with the taxonomic Order (using the same samples evaluated in Appendix 1). By doing so, species were combined into larger groups at the Order level and discarded from the CSCI sample if the Order could not be resolved for any metric calculations. An increasing number of ambiguous identifications was evaluated ranging from 10% (right side of plots) to 90% ambiguous (left side of plots). For each level of percent ambiguity (or percent taxa identified), 100 samples were evaluated where a different set of individuals were randomly selected to replace with the Order. As before, the results in the plots represent the average CSCI score for the 100 random samples (Figure 17a, b) and the coefficient of variation associated with the 100 random samples (Figure 17c).



**Figure 17. Effects of ambiguous identifications on CSCI scores.**

Based on the above analyses, we recommend a maximum percentage of ambiguous taxa not to exceed 50% (i.e., percent of identified taxa not to fall below 50%). Details include:

- Increasing ambiguity caused a decrease in CSCI scores from the true estimates
- CSCI scores are generally within ten percent of the actual if the ambiguous taxa are less than 50-60% of the total sample
- Precision decreases with more ambiguous taxa, although variation is typically less than 10% of the true mean if at least 30-40% of the sample contains unambiguous taxa.

### Appendix 3

#### Typical values of CSCI predictors in the South Coast region

Ranges of values for GIS predictors in the South Coast region, as well as Mountain and Xeric sub-regions, based on 1306 sites in the South Coast region. See Mazor et al. (2016) for details on about predictors. Min: Minimum observed value; q25, q50, and q75: 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> quantile of observed values. Max: Maximum observed value.

Variable	Min	q25	q50	q75	Max
Latitude (New_Lat)	32.563	33.658	34.068	34.240	34.727
Mountain	32.753	33.811	34.197	34.309	34.727
Xeric	32.563	33.515	33.955	34.157	34.471
Longitude (New_Long)	-119.406	-118.632	-117.662	-117.139	-116.435
Mountain	-119.406	-118.253	-117.447	-116.823	-116.435
Xeric	-119.382	-118.700	-117.760	-117.247	-116.467
Elevation (SITE_ELEV)	1	113	293	721	2447
Mountain	259	670	983	1392	2447
Xeric	1	62	172	289	1082
Elevation range (ELEV_RANGE)	4	615	1081	1775	3430
Mountain	70	884	1243	1704	2792
Xeric	4	500	920	1807	3430
Watershed area (AREA_SQKM)	0.12	18.2	61	220	6021
Mountain	0.12	11.0	34	107	952
Xeric	0.13	23.6	101	367	6021
Precipitation (PPT_00_09)	19435	30361	39941	50237	90057
Mountain	31117	48281	54437	64608	90057
Xeric	19435	28254	33403	40643	72396
Summer precipitation (SumAve_P)	220	437	670	1428	5574
Mountain	363	825	1254	3525	5574
Xeric	220	392	528	934	3702
Air temperature (TEMP_00_09)	1446	2218	2402	2527	2740
Mountain	1446	2003	2175	2310	2650
Xeric	2034	2379	2478	2562	2740
Soil bulk density (BDH_AVE)	1.38	1.54	1.57	1.59	1.69
Mountain	1.48	1.54	1.55	1.58	1.69
Xeric	1.38	1.55	1.57	1.59	1.69
Soil erodibility (KFCT_AVE)	0.09	0.19	0.24	0.27	0.34
Mountain	0.09	0.13	0.17	0.22	0.30
Xeric	0.10	0.23	0.26	0.28	0.34
Phosphorous geology (P_MEAN)	0.11	0.13	0.14	0.16	0.30
Mountain	0.12	0.13	0.15	0.16	0.19
Xeric	0.11	0.12	0.14	0.15	0.30

## Appendix 4

### Ranges of expected CSCI scores for different stream types in regions of California

Beck et al. (2019) provides estimates of likely ranges of CSCI scores for nearly all streams in California. Typical urban, agricultural, and open (i.e., undeveloped) stream segments were identified by evaluating land use in the catchment. Typical urban segments were identified where catchment-scale urban land use exceeded 50%, agricultural segments were identified where catchment-scale agricultural land-use exceeded 50%, and open segments were identified where catchment-scale undeveloped land uses exceeded 90%. The median value of each distribution point was then calculated and reported in the table below.

These typical values are shown for more to less certainty (wide to narrow range) in the landscape model predictions. Among regions, the variation in expected scores also provides context for landscape pressures that differ by location. For example, the expected range of scores in regions with heavy urban development (e.g., South Coast) is much smaller than streams that are neither urban nor agricultural. The North Coast region in contrast has an expected range of scores in urban streams that is similar to streams that are open. The range of scores in urban and agricultural streams were similar in the Central Valley where agriculture is the dominant land use.

Region	Land use	High certainty (10th - 90th)	Moderate (25th - 75th)	Low certainty (40th - 60th)
Statewide	Urban	0.34 - 0.83	0.45 - 0.72	0.53 - 0.64
	Ag	0.38 - 0.93	0.47 - 0.77	0.54 - 0.66
	Open	0.80 - 1.15	0.91 - 1.08	0.97 - 1.03
Desert-Modoc	Urban	0.53 - 1.07	0.68 - 0.98	0.78 - 0.89
	Ag	0.39 - 0.96	0.48 - 0.78	0.56 - 0.67
	Open	0.79 - 1.15	0.90 - 1.08	0.96 - 1.03
Sierra Nevada	Urban	0.51 - 1.07	0.65 - 0.97	0.76 - 0.88
	Ag	0.41 - 1.03	0.53 - 0.87	0.62 - 0.75
	Open	0.80 - 1.16	0.92 - 1.09	0.98 - 1.04
North Coast	Urban	0.72 - 1.17	0.87 - 1.10	0.94 - 1.04
	Ag	0.41 - 1.04	0.51 - 0.86	0.60 - 0.72
	Open	0.82 - 1.14	0.92 - 1.07	0.97 - 1.03
Chaparral	Urban	0.32 - 0.80	0.42 - 0.69	0.50 - 0.60
	Ag	0.40 - 0.98	0.51 - 0.84	0.60 - 0.72
	Open	0.80 - 1.15	0.91 - 1.08	0.97 - 1.03
Central Valley	Urban	0.39 - 0.90	0.51 - 0.79	0.60 - 0.71
	Ag	0.36 - 0.89	0.45 - 0.73	0.52 - 0.63
	Open	0.67 - 1.11	0.80 - 1.02	0.87 - 0.96
South Coast	Urban	0.30 - 0.76	0.40 - 0.66	0.48 - 0.57
	Ag	0.41 - 1.01	0.53 - 0.90	0.63 - 0.78
	Open	0.83 - 1.15	0.93 - 1.08	0.98 - 1.04

## Appendix 5

### Relationship between CSCI scores and habitat condition

Physical habitat information is typically collected alongside other bioassessment data. These data could include Index of Physical Habitat Integrity scores (IPI; Rehn et al. 2018) or California Rapid Assessment Method scores (CRAM; California Wetlands Monitoring Workgroup 2013), as well as their individual metrics or components, respectively. Although habitat quality alone should not be considered a constraint without a full understanding of the factors that have led to habitat degradation at a site, it may be used to provide additional evidence about likely ranges of CSCI scores at a site, which may provide additional information to outputs from the SCAPE model.

An analysis of approximately 500 sites from southern California shows that habitat quality may limit the likelihood of observing high CSCI scores. A comparison of physical habitat data with CSCI scores below provides some indication of when physical condition may be sufficiently poor and when an impacted CSCI score may be observed (e.g., below CSCI = 0.79, Figure 15). The blue lines are the quantile regression estimates for the 90th percentile. Where these regressions intersect the dotted line could be an indication of when CSCI scores are well below 0.79 for the corresponding habitat measure.

These regressions could be used to estimate ranges of expected CSCI scores where the landscape model is invalidated. For example, if a site has an IPI score of 0.45, the formula in the table below shows that the expected upper range of CSCI scores would be 0.81 (i.e.,  $0.56 + 0.55 * 0.45$ ).

**Equations from quantile regression models of CSCI scores against each habitat measure for the 90th percentile response. The model parameters are shown (intercept and slope) and the estimate of the habitat variable where CSCI is at 0.79. Negative estimates should be ignored.**

Habitat measure	Quantile model	Estimate at CSCI = 0.79
IPI	$0.56 + 0.55 * IPI$	0.42
Ev_FlowHab	$0.99 + 0.15 * Ev\_FlowHab$	-1.36
H_AqHab	$0.76 + 0.42 * H\_AqHab$	0.07
H_SubNat	$0.67 + 0.49 * H\_SubNat$	0.23
PCT_SAFN	$0.82 + 0.35 * PCT\_SAFN$	-0.08
XCMG	$0.87 + 0.32 * XCMG$	-0.24
CRAM	$0.4 + 0.01 * CRAM$	44.13
CRAM buffer and landscape condition (blc)	$0.44 + 0.01 * blc$	48.91
CRAM biotic structure (bs)	$0.7 + 0.01 * bs$	16.42
CRAM hydrologic structure (hy)	$0.41 + 0.01 * hy$	44.97
CRAM physical structure (ps)	$0.66 + 0.01 * ps$	20.96