Decision framework for validating CSCI scores and landscape model classifications

Marcus Beck

Raphael Mazor

27 August, 2019

Table of Contents

# Background

The California Stream Condition Index or CSCI is a robust tool for evaluating the biological integrity of streams in California using aquatic insect as indicators of stream health (Mazor et al. 2016). The CSCI is used in routine assessments conducted by the San Gabriel River (SGR) Regional Monitoring Program, a copperative non-governmental organization that manages aquatic resources in the watershed. CSCI data collected by the SGR Regional Monitoring Program 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 evaluate with the CSCI.

In many places, SGR managers lack the context for a biological condition assessment to determine the likelihood of a site for achieving biological integrity. For example, managers may expect a mountain stream in the upper watershed to have greater biological potential than an urban stream in the lower watershed. Many stream reaches with low CSCI scores occur in developed urban and agricultural landscapes, where restoration can be costly and it may be difficult to achieve regional reference-like conditions. In these situations, identifying a context for what CSCI scores are expected under current landscape conditions could be helpful for identifying where management actions are most likely to achieve intended outcomes, or where landscape alteration could limit success in achieving biological integrity. Conversely, identifying locations where biological condition exceeds the expectation could be prioritized for conservation.

The landscape model was developed to predict an expected range of CSCI scores relative to the level of landscape alteration that occurs in a watershed (Beck et al., n.d.). This model was applied to monitoring sites in the SGR waterhsed to identify stream reaches that are constrained for achieving a potential biological objective for the CSCI. Local managers are already using the landscape model with CSCI scores to prioritize sites for different management actions. For example, sites where observed scores are outside the predicted ranges from the landscape model could receive a high priority for follow-up actions, such as additional monitoring, casual assessment, or restoration.

Before additional actions are pursued at high priority sites, a reasonable level of confidence is needed in the sample used to estimate the CSCI and the stream classification from the landscape model. Samples may not be reliable if standard sampling procedures are not followed or abnormal environmental events affect a sample, producing a CSCI score that does not represent baseline conditions. Similarly, landscape model results may not be reliable if the spatial data used to create the classification do not represent current conditions at a site. A *validation process* is needed at high priority sites to verify the following:

* Is the biological sample used to calculate the CSCI reliable and within the standard protocol for estimating a site score?
* Does the constraint class defined by the landscape model reflect the actual landscape context for these sites?
* What data can we use and what questions can we ask to assess the validity of a bioassessment sample and stream class?

This document provides a set of questions to consider when evaluating the validity of the CSCI or landscape model results at high priority sites. The validation process is organized hierarchically from low to high effort, starting with the simplest questions that are easy to answer and should be evaluated with any assessment. High effort questions are those that may require additional data that may be more difficult to acquire, or the certainty of the conclusions may require some level of professional judgment.

Most importantly, this document is not a validation of the CSCI as an index or the SCAPE model; rather it is a validation of the usefulness of the sample used to estimate the CSCI score, or the SCAPE model outputs at the site in question. There are no policy recommendations herein for what makes a sample and/or score valid; validation should be part of normal QA/QC. Lastly, the document does not provide specific recommendations on what management actions are pursued after validation occurs.

# Validation

## Workflow description

Validation is the process of confirming validity of a CSCI score and stream channel classification from the landscape model. A “valid” CSCI score or classification is determined from professional judgment after reviewing appropriate data as being a sufficient description of the biological condition or constraint class, respectively, within the limits of the methods used to estimate each component. Validation is also within the larger framework of comparing CSCI scores to results form the landscape model to develop priorities, such that high priority sites identified through this process should undergo validation prior to follow-up actions. The validation process ends with a determination on whether or not the CSCI score and/or landscape model classification are sufficiently accurate to warrant follow-up management actions.

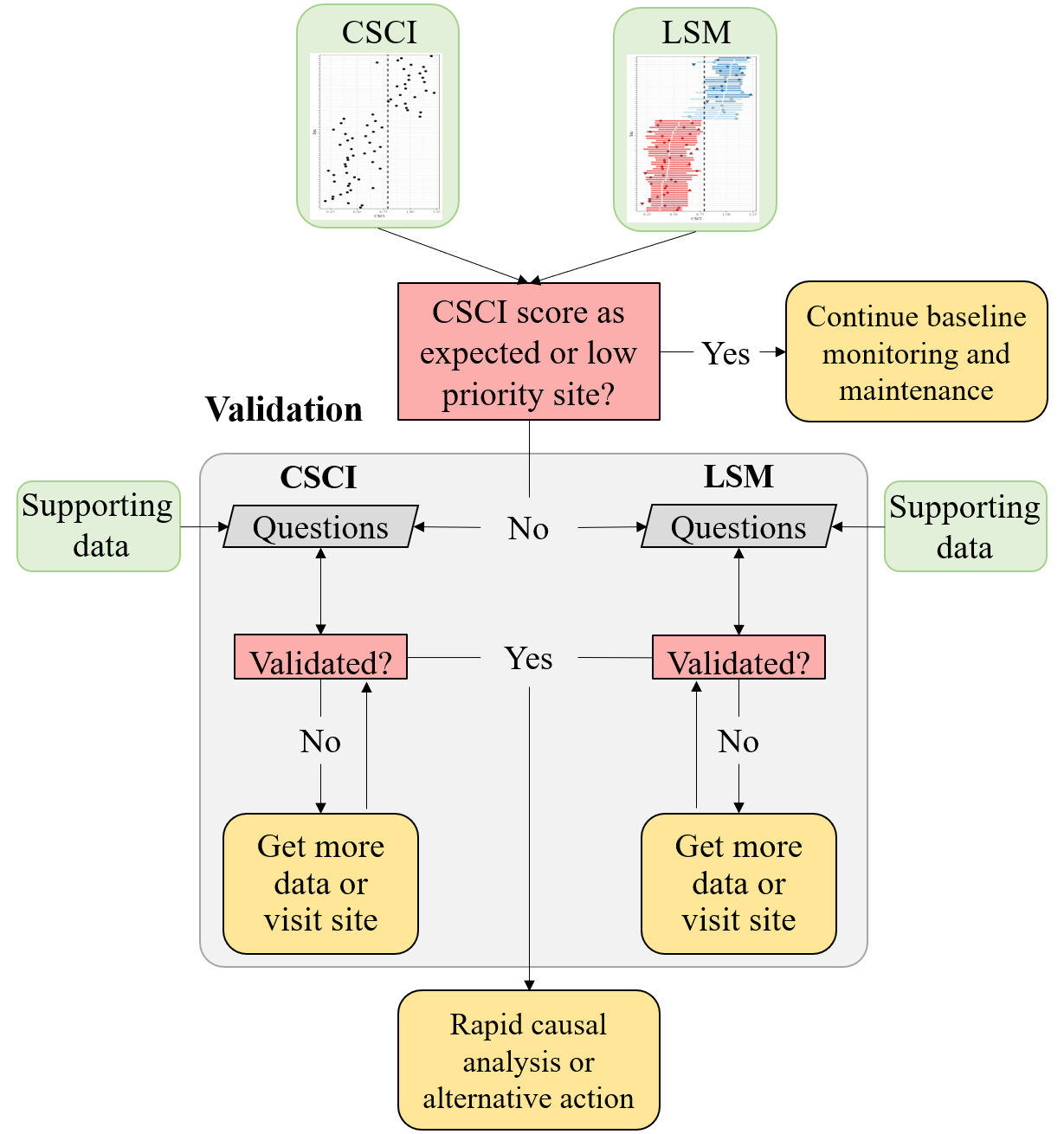


Figure 1: A simplified framework for validating CSCI and landscape model (LSM) information.

The conceptual framework for the validation process is shown in Figure 1. High priority sites are first identified by comparing CSCI scores to expectations from the landscape model. Validation can occur if a site is considered high priority, which may be a site that has a CSCI score that is not within the expected range. If the site is otherwise low priority, then baseline monitoring and maintenance should be continued. For high priority sites, the validation process is shown in the light grey box.

Separate but parallel validation processes occur for the CSCI score and landscape model classification. In both validation processes, questions vary in the level of difficulty to answer regarding validity of the CSCI sample or SCAPE ouput. Difficulty can very either because the data to address the question may or may not be readily available or the certainty of the conclusion can vary because of incomlete knowledge about how much the issue affects the outcome. Questions are generally organized in the section below from simple/easy to answer vs more difficult questions that may require some level of professional judgement.

A CSCI score or landscape model classification can be invalidated at any step in the validation process. A final conclusion is that one or both of the results are invalidated and that more data may be needed, which can include a site visit to collect information or to verify existing data. Once new data are obtained or existing data are confirmed, the validation process is reinitiated. The process occurs iteratively until both the CSCI score and landscape model result are sufficiently validated, i.e., the analyst has confidence the results of both are sufficient for pursuing follow-up actions. If both are validated, the validation process is finished and additional actions can be pursued, which may include rapid causal analysis to determine why biological condition differs from the expectation. If it is not possible to validate the information with additional data, the analyst can also decide to proceed with follow-up actions regardless of the results. In other words, the analyst may decide to trust the results based on their professional judgment, regardless of what the validation process provided.

It is also important to note that the CSCI is sensitive to stressors that may be caused by unusual events and that the score is valid in these instances, although it is not representative of “normal” conditions. Whether or not a CSCI score is valid after a temporary disturbance depends on the objective of the sampling. If the sampling is meant to capture the effects of these events, then the score is more than likely valid. However, if the objective is to evaluate “normal”, long-term conditions at a site, then the score may not represent these baseline conditions. In the latter case, field crews may or may not be aware of temporary events that can depress CSCI scores. If there is reason to question a score based on an unusual sampling condition, external data must be consulted. Field notes may be the best source of information.

## CSCI validation

The following is a list of questions to consider when validating a CSCI score. Each question focuses on a specific issue that may influence a CSCI score outside of the standard operating procedure for the index. These questions correspond to the left track of the validation box in Figure 1. For each question, a description is provided for why the the question or issue is a problem, followed by a description of what data do you need or where to you look to determine if the problem applies to your sample. A CSCI score could be invalidated for one to any of the questions and it is up to the analyst to determine when to stop considering additional questions.

### Is the sample count sufficient?

*Why is this a problem?*

Low sample counts 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 sampling protocols from the field manuals were not followed. Macroinvertebrate data used to calculate the CSCI are based on a representative subsample of 600 individuals from the total sample for each site. In general, sites where CSCI scores are based on substantially less than 600 individuals may be suspect.

*Where do you find an answer?*

The index output that is generated by the CSCI calculator (<https://sccwrp.github.io/CSCI>) provides information that can be used to evaluate the sample count. Specifically the “core” output contains a column for each CSCI sample for the sample count. In the first row, we see that the first sample was based on only 100 organisms. Therefore, we can assume that CSCI scores from the first sample are likely invalid.

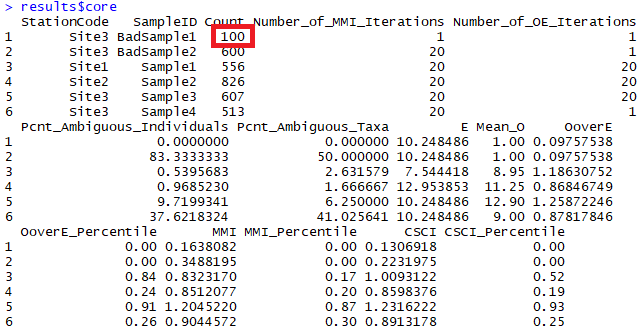


Figure 2: CSCI metadata that can be evaluated from the standard results. The first sample returns an invalid CSCI score because of a low sample count (in red).

The SWAMP program does not provide guidance on how many samples are required. However, an analysis of the effect of systematically reducing the sample count well below 600 individuals for several sites is provided in the appendix. Figure 10 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 10a). 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 10b 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 10c), although variation did not exceed ten percent until very low sample sizes (e.g., 150 or less).

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.

### Are there many ambiguous individuals or taxa?

*Why is this a problem?*

Ambiguous individuals or taxa cannot be used for O/E calculations in the CSCI, and may distort calculations of some metrics. This might occur if, for example, 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) . In these cases, a lower sample count of unambiguous individuals is used to calculate CSCI scores, which may not reflect true site condition.

Although the effect of ambiguous individuals or taxa on the CSCI may be similar to the effects of low sample count, the underlying reason for depressed CSCI scores is different. In the former case, taxa used to calculate the CSCI sample lack specificity and certain traits that are diagnostic for specific metrics may be impossible to quantify. Even if the sample is truly diverse and has a sufficiently high sample count, designations at higher taxonomic levels may not provide enough detail to fully characterize biological condition.

*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 (available at <https://safit.org/ste.html>). 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 (fig:coreex2) 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.

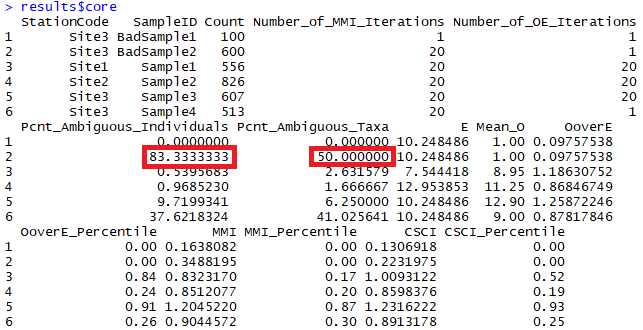


Figure 3: CSCI metadata that can be evaluated from the standard results. The second sample returns an invalid CSCI score because of many ambiguous individuals and taxa (in red).

The effect of systematically increasing the amount of ambiguous taxa on CSCI scores is shown in the appendix. Figure 11 was created by increasing the amount of ambiguous taxa that were included in each sample. Ambiguous taxa were introduced by replacing individuals in the sample that had known identifications with the taxonomic Order. 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 ambiguity 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 (Figures 11a,b) and the coefficient of variation associated with the 100 random samples (Figures 11c).

Based on the above analyses, we recommend a maximum percentage of ambiguous taxa not to exceed 40-50% (i.e., percent of identified taxa not to fall below 50-60%). 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.

### Was the sample affected by abnormal flow?

*Why is this a problem?*

Abnormal flow events or high flow conditions can scour stream beds and lower biodiversity as a site. This can produce samples that have lower macroinveretebreate diversity than would be expected for what the site supports under baseline conditions. Although high flow events can occur during and after storms, scouring resulting from human activities (e.g., dam releases) 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. For baseline assessment, scour from natural events and the effect on the sample are the primary concern for validation.

*Where do you find an answer?*

Field notes should provide indications of what scour looks like at a site, relative to normal conditions. It is up to the analyst to determine what constitutes rejectable scour. Consulting data from nearby weather stations could provide some information about precipitation events that could cause scour ([NOAA stations](https://www.ncdc.noaa.gov/cdo-web/datatools/findstation). USGS [flow stations](https://waterdata.usgs.gov/nwis/rt) also provide direct measures of flow in continuous records. Dam operations that could affect flow could be directly obtained from the operator to identify periods when water was released.

### Was the sample affected by drought?

*Why is this a problem?*

Drought conditions can stress stream communities in several ways, primarily by reducing flow below baseline. Flow reduction can alter the physical and chemical conditions in the stream, which can adversely impact biological communities. Changes from reduced flow may include reduced dissolve oxygen, increased stream temperatures, encroachment of riparian vegation, 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. As such, CSCI scores in streams where drought conditions are not uncommon do provide an accurate indication of biological integrity. However, water diversions as part of dam operations or municipal water use can reduce community resilience to drought. In these cases, abnormally low CSCI scores that do not reflect the actual biological condition are the expected outcome.

*Where do you find an answer?*

Normal protocols defined under the stream sampling SOP (Ode, Fetscher, and Busse 2016) require that sampling be conducted under baseline 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 do not produce accurate CSCI scores. Moreover, a [weekly drought map](https://www.climate.gov/maps-data/dataset/weekly-drought-map) can be consulted to provide additional information on conditions at the time of sampling.

### 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 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). Although fire is a natural phenomenen, an increase in the severity and magnitude may result from climate change or may otherwise be exacerbated by drought conditions.

*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](https://hub.arcgis.com/datasets/653647b20bc74480b335e31d6d81a52f)) 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 perecentage 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).

### Was the sample affected by vegetation management and/or debris removal?

*Why is this a problem?*

Vegetation management, debris removal, and stream regrading can be common management activities in urban streams. These activities help remove in-stream barriers that may obstruct normal flow operations or otherwise assist in maintaining flood control. 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.

*Where do you find an answer?*

Flood control maintenance records provided by public works departments or sanitation districts 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).

### Was the sample affected 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.

*Where do you find an answer?*

Records in existing databases, such as the SMC database, can indicate if biological control vectors are present. For example, fish surveys or field notes could indicate if mosquitofish are present at a site. Chemistry data may also indicate if pesticides are present. If none of the above provide clues, municipal agencies that conduct vector control operations could be contacted.

### Was the sample outside of the typical index period?

*Why is this a problem?*

The SOP guidelines for field sampling of macroinvertebrates (Ode, Fetscher, and Busse 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, 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.

*Where do you find an answer?*

The date of sampling should coincide with the guidelines outlined above. However, in southern California, these recommended sampling dates are contracted due to more pronounced seasonal flow cycles in this more arid climate. If sampling is well past July or well before April, the likelihood increases of not sampling the community under baseline conditions. Also, see the guidelines above on scour and drought as this often relates directly to the sample period.

### Does the site have a bad watershed delineation?

*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 is well-represented by topography.

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 elevation areas. Developed landscapes also complicate watershed delineations because the flow of water may have been significantly altered from natural patterns. In these cases, water may not follow strict topographical boundaries due to channelization or diversions. Overlaying the watershed delineation on aerial photos can provide clues about the effect of land use alteration on flow patterns and if the delineation is well-represented by topography.

*Where do you find an answer?*

The original GIS file for the watershed delineation should be viewed to assess the relative quality. If this file is unavailable, the [Streamstats](https://streamstats.usgs.gov/ss/) website can provide some clues about the watershed delineation. Figures 4 provides examples of two extremes when watershed delineations can be produced with differing quality. Figure 4a shows a high gradient stream in a mountainous, undeveloped area. In this scenario, the topography is sufficiently pronounced to easily identify the watershed boundaries and the hydrological flow follows the landscape. Conversely, figure 4b shows a low gradient stream in an urban setting. The watershed boundaries are not easily identify from elevation gradients and the digitized stream channels probably do not accurately reflect the flow of water in this altered landscape.

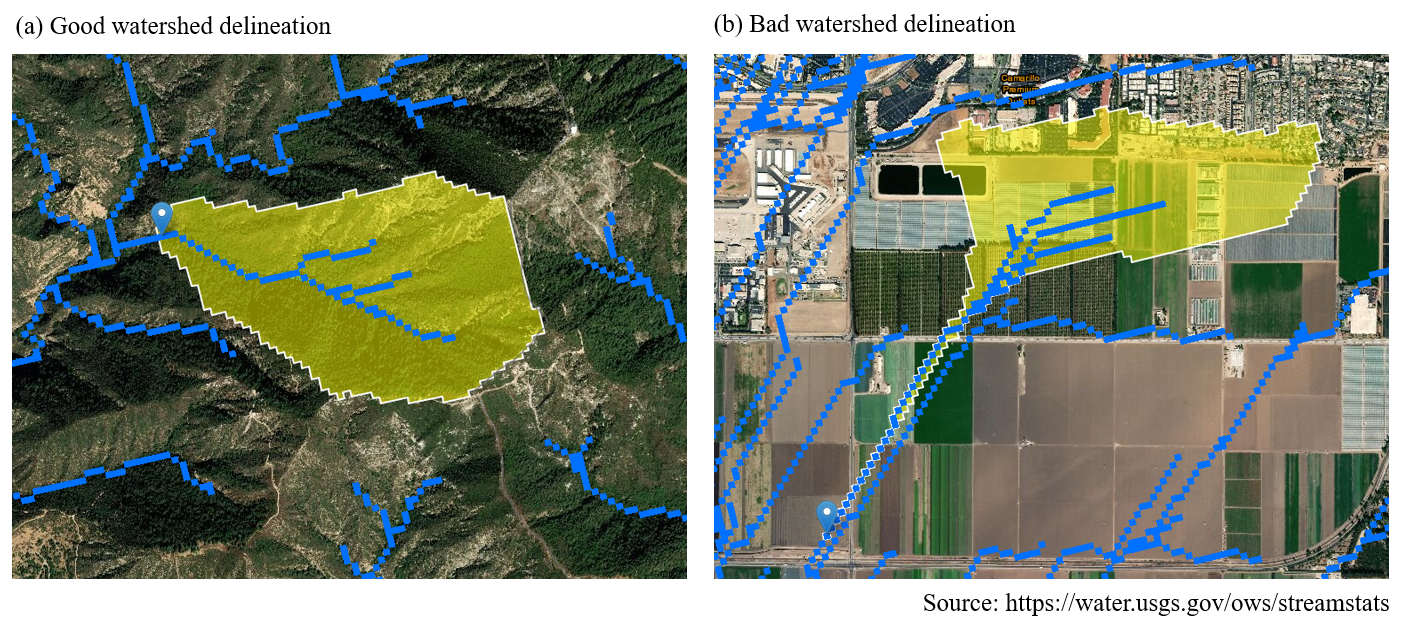


Figure 4: Watershed delineation examples.

Figure 5 provides some additional examples of how snapping can produce inaccurate watershed delineations. 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 hydroline for the delineation. In Figure 5a, we see a snap that produced a likely realistic representation of the watershed that drained to the site. However, in Figure 5b, 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 5c, we see a site that was manually snapped to a location and the watershed was manually delineated. In Figure 5d, 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.

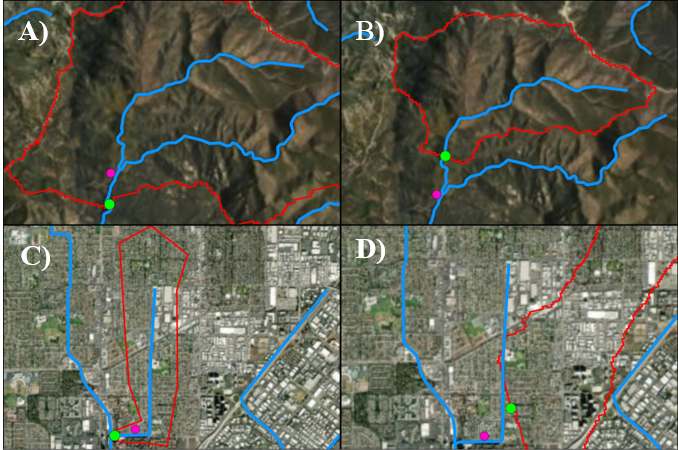


Figure 5: Watershed delineation examples showing snapping challenges. The pink dot is the recorded site latitude/longitude and the green dot is the snapped location. The watershed boundaries resulting in each case are outlined in red.

### Are there unusual settings where the CSCI is known to give low scores?

*Why is this a problem?*

There are a few locations in California where CSCI scores may be unreliable, although this is only theoretical and not shown in any literature. In these locations, scores may be depressed because the biological community may be naturally low in diversity and the 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 can influence the physical and chemical characteristics of the stream that structure the diversity of the biological community. This confounds the ability of the CSCI to distinguish between natural and anthropogenic variation, resulting in unreliable scores.

*Where do you find an answer?*

Local site knowledge from field crews could provide information on abnormal locations where the biological expectations are naturally different from other streams in the region. In most cases, field crews will likely already be aware of locations where standard bioassessment methods produce odd or unexpected results compared to other locations. Detailed geological maps may also provide clues regarding localized environmental differences that could influence the score.

Proving that natural settings introduce bias to the CSCI or other bioassessment indices is challenging, but can be done with the following 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](http://www.ceden.org/) or the SMC data portal as a preliminary step in finding these sites. Additionally, consult the development data that comes with the [CSCI](https://github.com/SCCWRP/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.
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 substantial correlation (e.g., R >0.2), the CSCI is likely responding to stress in this setting.

### Are the GIS predictors for the CSCI adequate?

*Why is this a problem?*

GIS predictors are required for calculating CSCI scores. The CSCI interim intstructions (Mazor et al. 2018) provide information on which GIS metrics are used and how they can be estimated. In addition to site latitude/longitude, the predictors include site elevation, elevation range, watershed area, average precipitation, average temperature, mean June to September monthly precipitation, average bulk soil density, average soil erodibility factor, and average phosphorus geology. This information is used in the predictive model of the CSCI to help identify an expected macroinvertebrate community based on watershed or other landscape characteristics. If the predictors are inaccurate, then the expected community will no incorrectly predicted and the overall CSCI score will be invalid. Predictors could be inaccurate because of an inaccurate watershed delineation or the predictors do not describe landscape conditions at the time of sampling.

*Where do you find an answer?*

The wateshed delineation should first be evaluated to determine accuracy, see the section above. Any inaccuracies in the delineation will produce inaccurate GIS predictors. Second, consult the section below on “has land cover changed recently?” for the landscape model. This section describes how to validate issues of currency in GIS data.

## Landscape model validation

The landscape model validation questions correspond to the right track of the grey validation box in Figure 1.

### Is the sampling reach atypical of a channel’s overall class (e.g., unconstrained class surrounded by constrained classes)?

*Why is this a problem?*

The landscape model assigns a constraint class to digitized reaches in the NHD-Plus stream hydrography layer. A reach can be described as one of four classes: likely constrained, possibly constrained, possibly unconstrained, likely unconstrained. Typically, stream reaches of the same class co-occur in space such that it is more likely that a constrained reach is surrounded by other constrained reaches and an unconstrained reach is 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 be an artifact of the 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. As an example, the online [SCAPE](https://sccwrp.shinyapps.io/scape) application provides this information for the San Gabriel River Watershed. Figure 6 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 6a). We can toggle the base layer to show a satellite image of the location to get a better idea of the landscape (Figure 6b). 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.

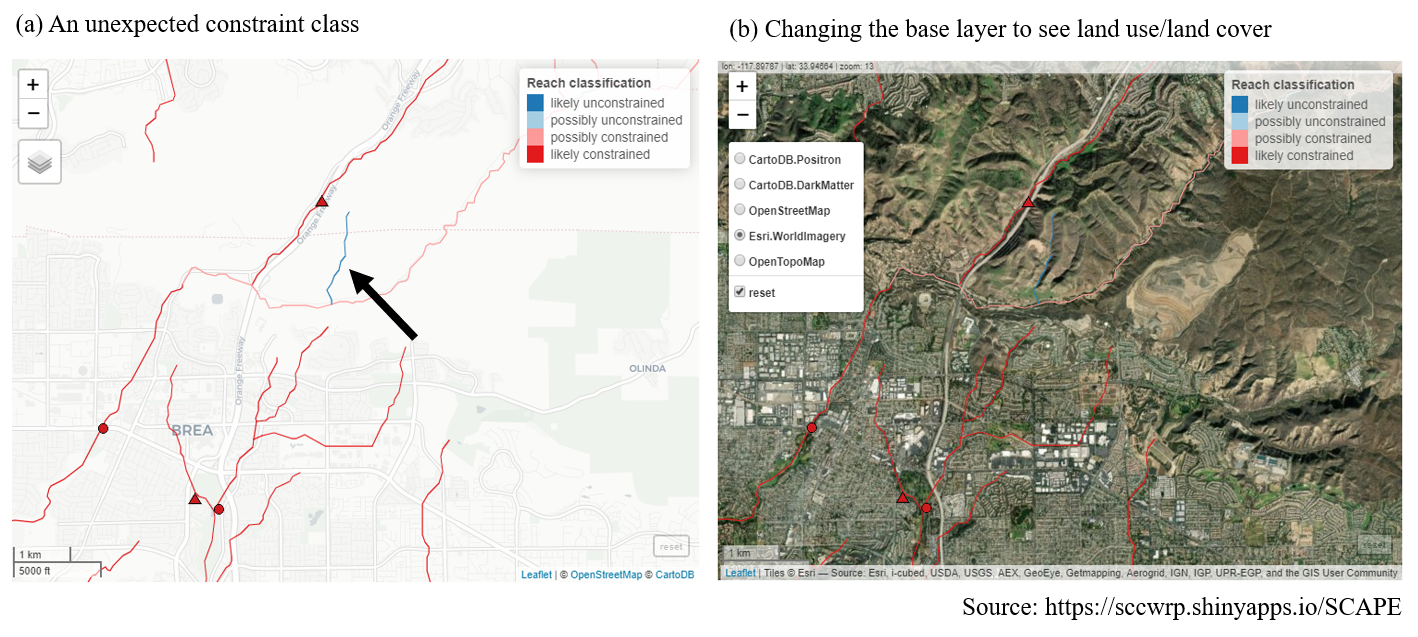


Figure 6: An unexpected stream class is validated by examining the land use, lane cover base layer.

For locations not in the San Gabriel watershed, a [shapefile](https://knb.ecoinformatics.org/view/urn:uuid:75411f50-32ed-42a5-bbfd-26833c7a441f) is available for all stream reaches in California where the landscape model was applied.

### Has land cover recently changed?

*Why is this a problem?*

The landscape model provides an expectation of biological condition 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 1). Because some of the primary products relate to a specific year, the associated constraint classes from the landscape model may not accurately reflect current constraints.

Table 1: 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). The measurement scales for each variable are at the riparian (100 m buffer), catchment, and/or watershed, scale relative to a stream segment.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | 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?*

There is no quantitative approach to verify if the constraint class accurately reflects the current landscape. However, 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. 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.

The Google Earth (<https://www.google.com/earth/>) time slider provides a useful tool to evaluate temporal changes in land use over time that could be affecting a site. 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 7 shows a screen shot from SCAPE for a current image, including some stream channels in the possibly unconstrained class. In the right image for current conditions, we see a parking lot adjacent to the stream channel. Using Google Earth, we can view an image from 2008 in the left image to see that this parking lot was not yet constructed. Because our land use data used to calibrate the model correspond roughly to this time period in the right image, this stream channel may be incorrectly classified. The channel may be more accurately described as constrained. A site visit could also confirm this information.

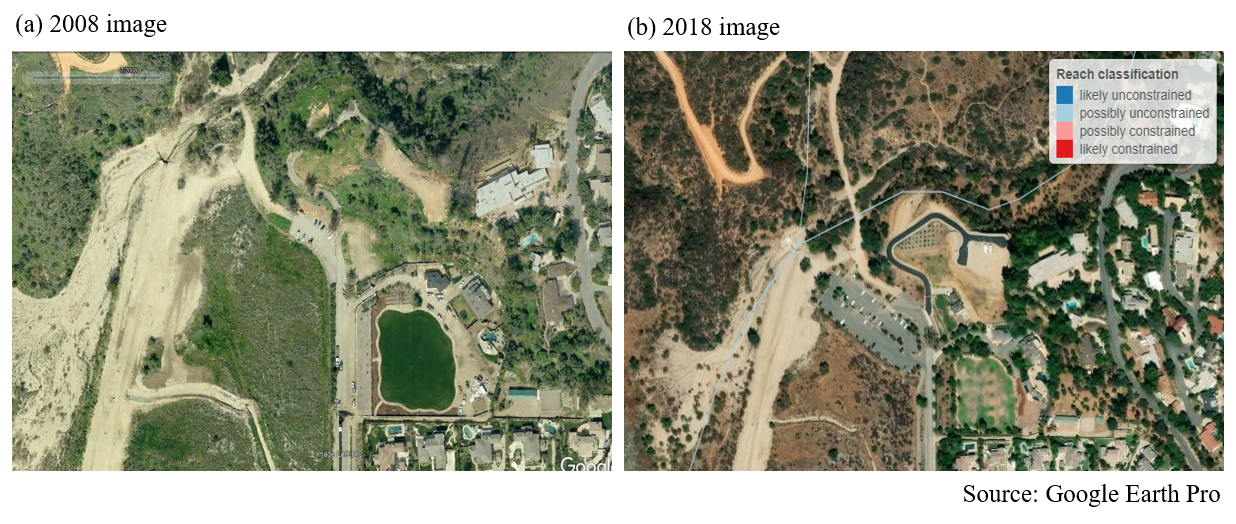


Figure 7: An example showing a (a) sampling location in 2008 and (b) unconstrained channels next to a new parking lot ten years later. Note the absence of the larger parking lot in the earlier photo.

### Are there constraints not captured by the landscape model that are affecting biology at the site?

*Why is this a problem?*

As noted in the previous question, the landscape model was calibrated using large-scale watershed characteristics (Table 1) that are unlikely to change dramatically over most management time scales (e.g., five to ten years). In some instances, there may be constraints affecting biology at a site that are not captured by the model. Many of these constraints may be short-lived or temporary disturbances that fall under the same category as those for question three for the CSCI questions (i.e., abnormal flow, fire, etc.). The same approach to addressing validity of the CSCI score also applies to the landscape model classification. Field notes or external datasets could be consulted to determine if the site classification is warranted.

*Where do you find an answer?*

A more general issue applying the landscape model is that other landscape stressors not captured by the model may have had a legacy or long-term impact on stream biology. For example, in the Sierras or northern California, timber harvesting, silviculture, or mining may have had historical impacts on biology which are not reflected in the current model. These legacy impacts might not be explicitly included in the landscape model and should be considered in the constraint classification if the current class does not accurately characterize biological expectations. A first step is to evaluate the variables in Table 1 and assess how well these variables capture stressor gradients at the site of interest. If they are not appropriate, consider that the current classification is invalid and alternative ways to assess constraints could be used (e.g., what is the range of scores for other sites in similar conditions for the region).

### Do the spatial data for the stream channel match the actual location on the landscape?

*Why is this a problem?*

The landscape model assigns a constraint classification to every NHD-Plus stream reach where data are available. Because of this, information from the landscape model is only as good as the spatial representation of stream locations in the hydrography dataset. Stream reaches may not be good representations of the actual channel location if 1) the stream is in a landscape with altered hydrology, or 2) the stream channel has migrated over time due to normal hydro-geomorphic processes.

*Where do you find an answer?*

Figure 8 shows an example of where landscape model results are available, whereas the stream lines represent the challenges of digitizing actual stream channels in altered landscapes. The actual stream channel that corresponds to the spatial data represented in the NHD-Plus should be compared to verify accuracy and if the results from the landscape model still apply.



Figure 8: Stream hydrography in altered lanscaped. Actual stream reaches may not be well represented by existing spatial layers.

### Are the results from the landscape model close to key decision points?

*Why is this a problem?*

As with many decision support tools, the constraint classifications derived from the landscape model are categories that are based on binning continuous data. This approach allows complex information to be distilled for communication in a more digestible format. This improvement in communication comes at the expense of specificity - binning results reduces the type of information that can be obtained. As an example, the landscape model would classify a stream reach as possibly constrained if the median score is 1/100 of a decimal place below the biological objective, whereas increasing that score by 0.02 points would cause the landscape model to classify the reach as possibly unconstrained. Because of this, an analyst should be aware of how the constraint classes are determined and if the results for a site are close to decision points.

Figure 9 provides a graphical description of how the landscape constraint classes are determined. Stream segments where the predicted 90th quantile score is below the biological objective threshold are considered likely constrained, whereas those where the predicted 10th percentile is above the threshold are considered likely unconstrained. Possibly constrained and possibly unconstrained are assigned if the score expectations overlap the biological objective and the median score is below or above the objective, respectively.

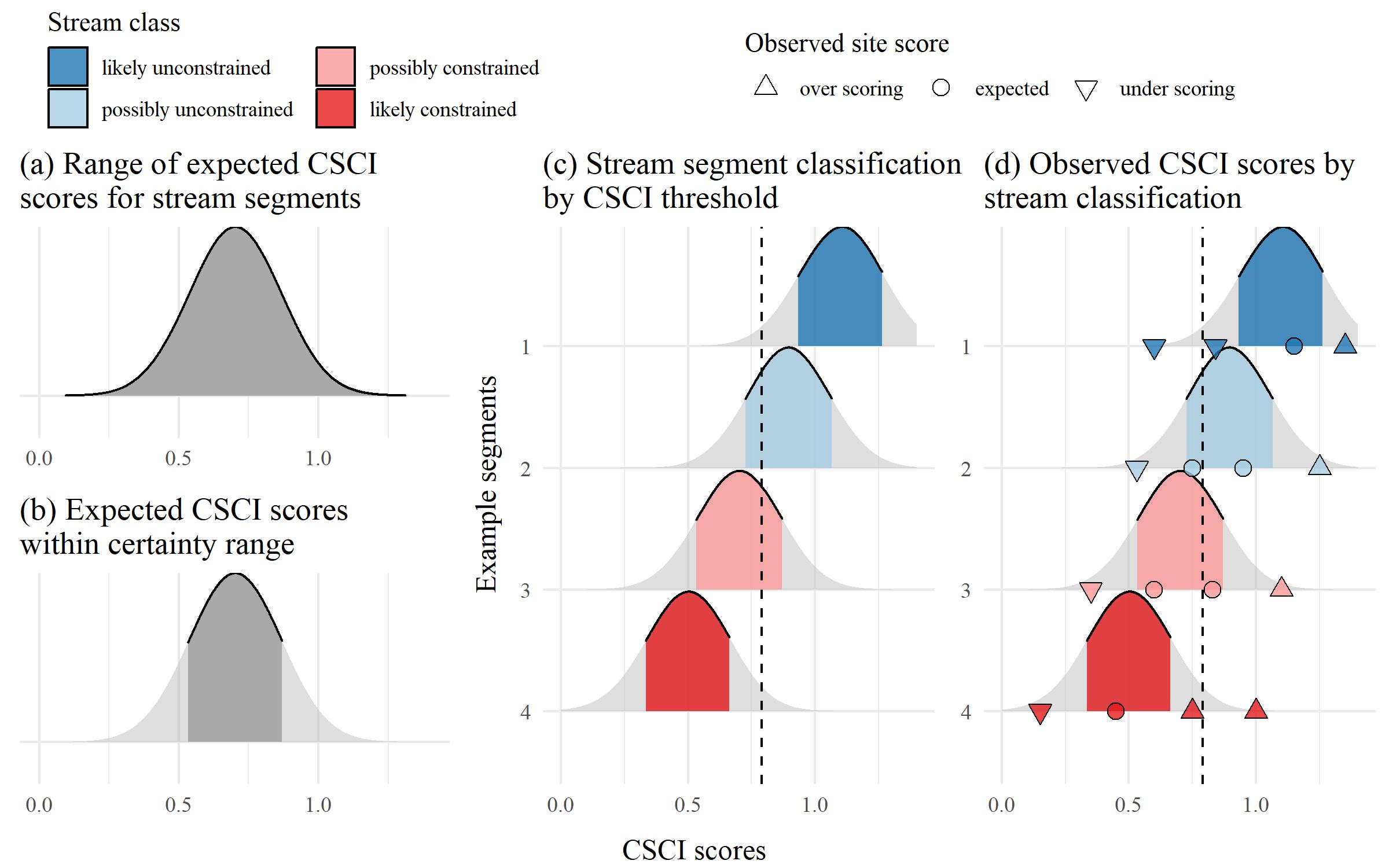


Figure 9: Decision points and methods for defining sream classifications from the landscape model.

*Where do you find an answer?*

Sites where the 10th, median, or 90th percentile of the score expectations are near the biological objective should be further evaluated. Does the site classification change if slightly different breakpoints are used, such as the 5th or 95th percentile of scores? Does the site classification change if the biological objective changes? The online SCAPE application provides options to explore the effect on stream classifications when changing the main decision points. If the stream class does not change with slightly different decision points, then the the stream class is robust. However, if small changes in decision points change the the stream class and different conclusions are made about biological expectation at a site, a judgment call can be made as to how much validity a stream class has for the particular reach.

## Data sources

The following is a list of resources that can provide information to address the validation questions.

* CSCI metadata- consult CSCI SOP in (Mazor et al. 2018) and package documentation on [GitHub](http://sccwrp.github.io/CSCI/)
* [SCAPE](https://sccwrp.shinyapps.io/scape/) website
* Reference site information - check the loadRefData() or loadRefBugData() in the CSCI package
* GIS data
  + [StreamCat](https://www.epa.gov/national-aquatic-resource-surveys/streamcat) watershed data
  + [NHD hydrography](https://www.usgs.gov/core-science-systems/ngp/national-hydrography) layers
  + Catchment/Watershed layers - consult SMC database or explore with [Streamstats](https://streamstats.usgs.gov/ss/)
  + Land use/land cover data - National Land Cover Database [2006](https://www.mrlc.gov/data/nlcd-2006-land-cover-conus), [2011](https://www.mrlc.gov/data/nlcd-2011-land-cover-conus-0), [2016](https://www.mrlc.gov/data/nlcd-2016-land-cover-conus)
  + GIS metrics for CSCI - consult SMC database
  + Google imagery and [time slider](https://support.google.com/earth/answer/148094?hl=en) - available through [Google Earth](https://www.google.com/earth/)
* Field data
  + [SWAMP](https://www.waterboards.ca.gov/water_issues/programs/swamp/), SMC Data Portal, [CEDEN](http://www.ceden.org/)
* Local knowledge
  + Field notes
  + Site photos
* Additional external datasets
  + Weather conditions ([noaa.gov/weather](https://www.noaa.gov/weather), [NOAA stations](https://www.ncdc.noaa.gov/cdo-web/datatools/findstation))
  + USGS [flow stations](https://waterdata.usgs.gov/nwis/rt)
  + [weekly drought map](https://www.climate.gov/maps-data/dataset/weekly-drought-map)
  + Fire perimeters at [Cal Fire Hub](https://hub.arcgis.com/datasets/653647b20bc74480b335e31d6d81a52f)
  + [Mining data](https://mrdata.usgs.gov/catalog/combine.php?term=3-685&with=1-fUS06)
  + Timber harvest/silviculture data - from [USDA Forestry Service](https://data.fs.usda.gov/geodata/edw/datasets.php)

# High priority sites in SGR watershed

* 405CE0280, SMC00480, SMC00144, SMC02972, SMC04524, SMC06496
* Why are these high priority?
* Validate CSCI/LSM results for each using available data to demonstrate the process
* What conclusions are made?

# Appendix

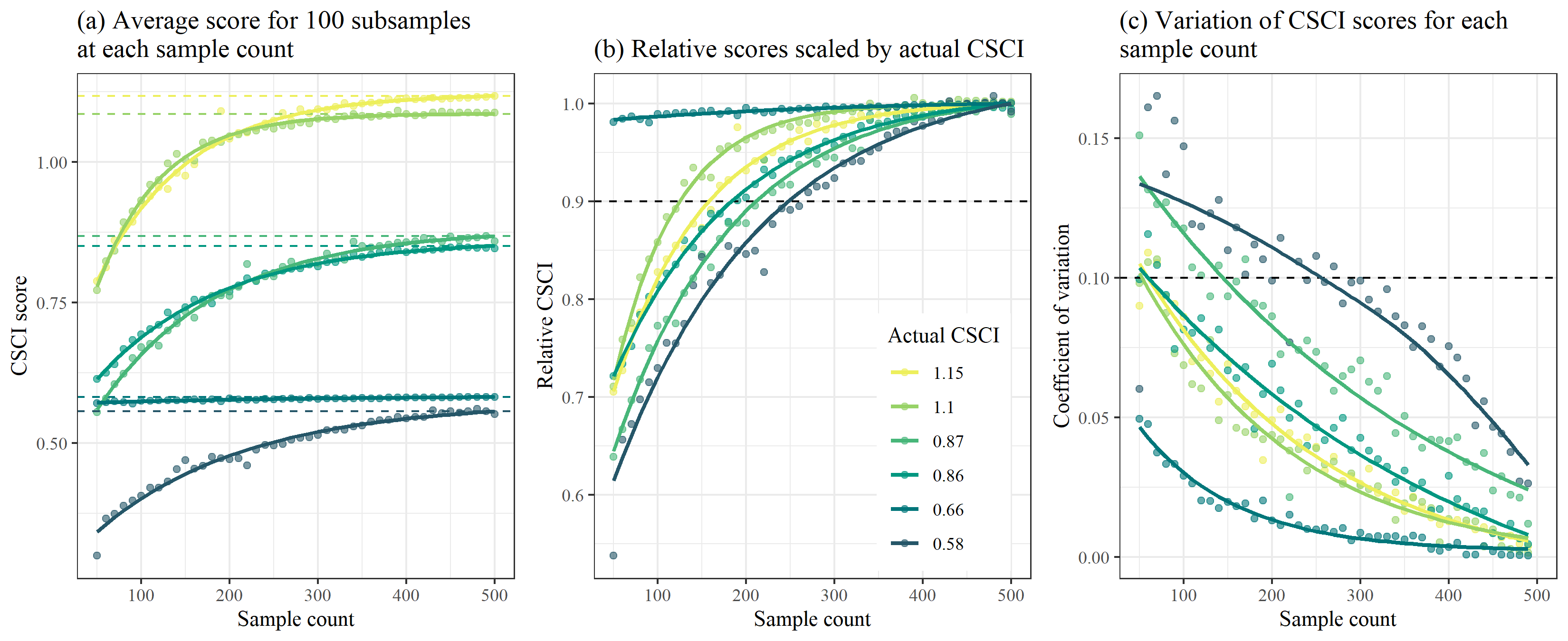


Figure 10: Effects of reducing sample size on CSCI scores.

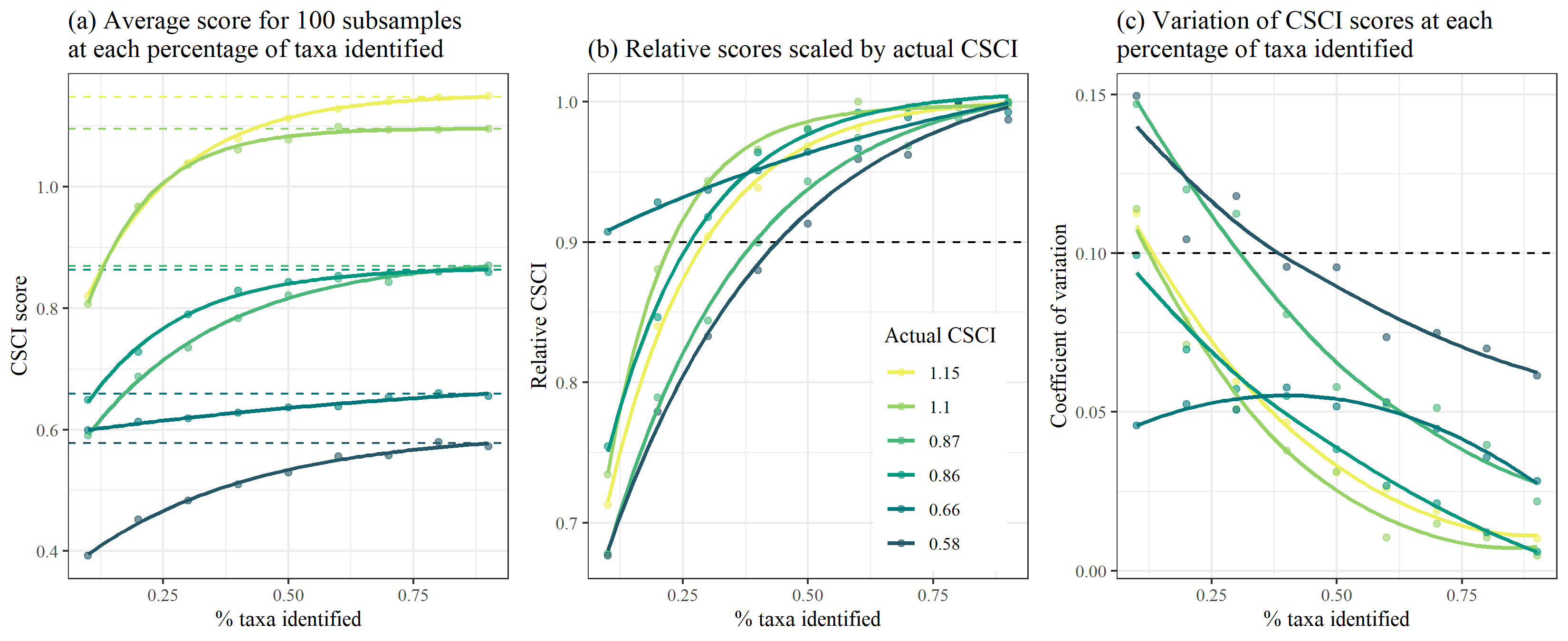


Figure 11: Effects of introducing ambiguous taxa on CSCI scores.

# References

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. n.d. “Prioritizing Management Goals for Stream Biological Integrity Within the Developed Landscape Context.” *Freshwater Science*.

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/>.

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>.

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>.

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/docs/combined_sop_2016.pdf>.

Oliver, Allison A, Michael T Bogan, David B Herbst, and Randy 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>.