Decision framework for SCAPE validation: SOP

Marcus Beck

Raphael Mazor

Scott Johnson

29 May, 2019

Table of Contents

# Background

* What are constraints and how are they determined with SCAPE and the landscape model (LSM)
* How is SCAPE used to prioritize/what are high priority sites in SGR
* We need to validate the information for high priority sites
  + Does the landscape model correctly reflect the landscape conditions affecting these sites?
  + Do observed CSCI scores correctly reflect local in-stream conditions?
  + If the above are true, what factors may explain the discrepancy between observed scores and predicted ranges?
* Who this document is for
  + Resource managers in SGR
  + Field crews and technicians
* What this document provides
  + Checklist of questions to evaluate for considering validity of CSCI and LSM scores
  + Organized in hierarchy from low to high effort, i.e., desktop exercise to collection/eval of external data
  + Decision is a judgment call based on available evidence
* What this document is not
  + Not a validation of the CSCI as an index - it is a validation of the sample
  + Not a validation of LSM as a model - it is a validation of the input data
  + No policy recommendations for considering a sample/score valid, this is part of normal QA/QC
  + Does not define what action is pursued once CSCI/LSM are validated

# Validation

* General framework from priority to follow-up action is below, validation is a component of this broader process

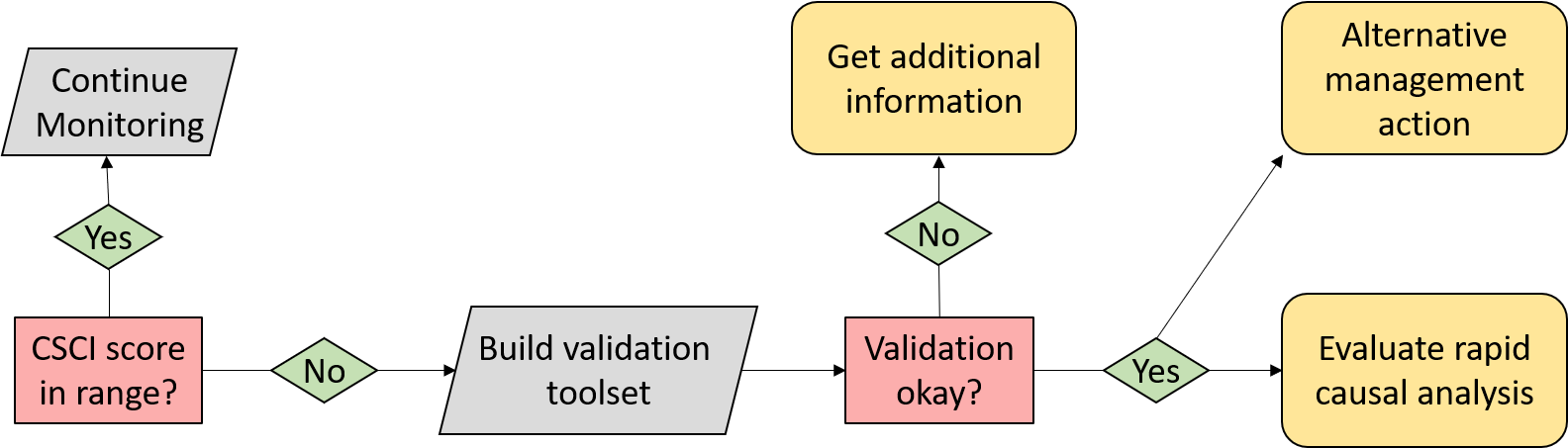


Figure 1 A simplified framework for validating CSCI and SCAPE information.

* The validation process evaluates metadata and external datasets
  + Identify sites/samples to validate
  + Build validation tool set
  + Verify against validation criteria
  + Make final decision

## Data sources

List of resources to assist with building the validation tool set - can go here or in appendix.

* Metadata QA/QC
  + CSCI metadata (consult CSCI SOP and package documentation)
  + SCAPE website
  + Reference site information
* GIS data
  + StreamCat
  + NHD hydrography
  + Catchment/Watershed layers
  + LU/LC data - NLCD 2006, 2011, NAIP aerial imagery
  + GIS metrics for CSCI
  + Google imagery + time slider
* Field data
  + SWAMP, SMC, CEDEN
* Local knowledge
  + Field notes
  + Site photos
* Additional external datasets
  + weather conditions (noaa.gov/weather)
  + Fire perimeters
  + Dredging (?)
  + Mining (?)
  + Timer harvest/silviculture (?)

## CSCI

### Questions to ask

The following is a list of several questions to consider when validating a CSCI score. Each question focuses on a specific issue that may influence a CSCI outside of the standard operating procedure for the index. For each question, a description of the issue is provided, how might the issue affect the score, and what data are needed to answer the question. The questions are also described as simple desktop evaluations (e.g., examination of the metadata that are included with standard CSCI output) or more challenging questions may require additional data (e.g., site visits) or analyses to fully evaluate. A CSCI score could be invalidated for one to any of the questions and it is up to the individual to determine when to stop considering additional questions.

#### Desktop validation

These questions can be addressed by consulting the SOP guidelines for the CSCI and/or the metadata that are included in the CSCI output.

1. Is the sample count sufficnet?

The CSCI provides a consistent measure of the degree of alteration of the macroinvertebrate community from reference conditions. This information is only valid if a sufficient sample has been collected in the field for calculating the index. Low sample counts may not provide a complete picture of the community that was present during sampling. The index output that is generated by the CSCI calculator 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:

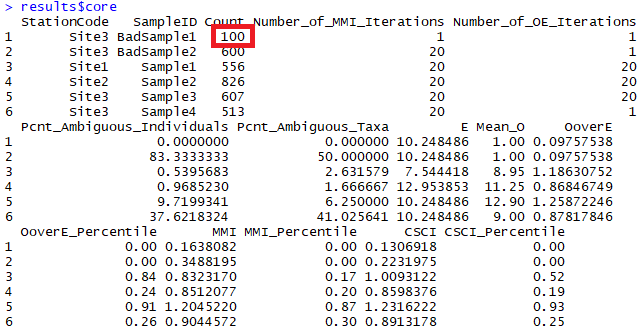


Figure 2 CSCI metadata that can be evaluated from the standard results. The first sample returns an invalid CSCI score because of a low sample count (in red).

In the first row, we see that the first sample was based on only 100 organisms. According to guidance for the CSCI, a minimum of 450 indviduals is needed for the pMMI and 360 for the O/E. Therefore, we can assume that CSCI scores from the first sample are invalid.

1. Are there many ambiguous individuals or taxa?

Ambigous taxa or individuals cannot be used for metric and O/E calculations in the CSCI. This might occur if, for example, a sample is dominated by midges all identified to family. In these cases, a lower sample count of unambiguous individuals is used that can cause depressed MMI or O/E scores.

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 observations in a sample that do 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, samples with high percentages may have invalid CSCI scores.

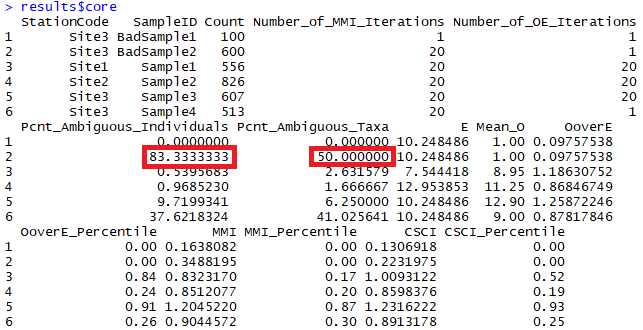


Figure 3 CSCI metadata that can be evaluated from the standard results. The second sample returns an invalid CSCI score because of many ambiguous indivdiuals and taxa (in red).

1. Was the sample outside of the typical index period?

The SOP guidelines for field sampling of macroinvertebraes [@Ode16] states the typical index period 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 soutern California, and later in this range for higher latitudes. Sampling that occurs outside of this range will likely not produce a representative sample for which the CSCI can be used. Sample dates can be verified from the raw data used to calculate the CSCI.

#### External validation

These questions require additional datasets or field visits for validation.

1. Was the sample affected by natural or temporary disturbances?

CSCI scores may be invalid if the sample was affected by natural or temporary disurbances, such as drought, scour from high flow events, or wildfire. The SWAMP sampling protocol indicates that sampling should occur during normal, baseflow conditions. Sampling outside of these conditions, even during the normal index period from May to September, may result in depressed CSCI scores. Field notes may indicate if abnormal conditions were present. Exernal datasets, such as flow records or time and location of fire events, may also provide clues of abnormal conditions.

1. Unusual sampling conditions (flow was too low/high for sample nets)?
2. Unusual settings where CSCI is known to give low scores?
3. Uncertainty in score with n = 1?
4. Bad watershed delineation?
5. High variability with repeat site visits?
6. Score is very close to decision points (e.g., 0.77 or 0.80)?

### Data to evaluate

* Weather data
* Fire perimeters
* QA reports, CSCI metadata
* Field notes
* Upstream/downstream samples or nearby
* ASCI, PHAB, CRAM, water quality observations
* Pictures
* Reference sites
* GIS data
* watershed data
* Degree of deviation from expectation

## SCAPE

### Questions to ask

For each question, provide a description of the issue, how might it affect the LSM category, and what data to evaluate. Rank the questions from easy to difficult, i.e., simple desktop validation vs compile external datasets.

* Close to landscape model breakpoints?
* Sampling reach is atypical of segment’s overall conditions (e.g., unconstrained surrounded by constrained)?
* Channel has migrated from nominal location (NHD issues)?
* Land cover had changed?
* Constraints not captured by model (e.g., fire impacts, dredging, mining)?

### Data to evaluate

* Satellite imagery
* Site photos
* Alternative land use/land cover data (2006, 2011 NLCD)
* PHAB data (metrics and field notes)
* CRAM
* Landscape stressors not characterized by StreamCat
* Google images
* Site location relative to NHD segment
* Catchment size
* When is lu/lc change important?
* Reference GIS data

## Conclusions

What decisions do you make once CSCI/LSM are validated?

* Validated - carry on
* Not validated - trust results anyway?
* Not validated - get more samples?
* Conduct RSCA (not covered here, but briefly describe)
* Others?

# 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?

# Colophon

The current Git commit details are:

#> Local: master L:/San Gabriel RSCA\_MB/Repositories/SGRrsca  
#> Remote: master @ origin (https://github.com/SCCWRP/SGRrsca.git)  
#> Head: [7e61be0] 2019-05-01: working on csci questions