The Algal Stream Condition Index (ASCI): instructions for calculating ASCI scores using R

Southern California Coastal Water Research Project, State Water Resources Control Board, Surface Water Ambient Monitoring Program (SWAMP)

Susanna Theroux

Raphael Mazor

Marcus Beck

Quynh-Thi Ho

Shuka Rastegarpour

30 July, 2019

Table of Contents

SWAMP SOP #XXXX

# Introduction

This document describes steps in calculating the California Algal Stream Condition Index (ASCI), a bioassessment index that measures stream health based on algae (diatoms and soft-bodied algae) taxonomy data. This document describes the data input requirements and the computational steps for calculating ASCI scores using the custom ASCI R package (<https://github.com/sccwrp/asci>). Additionally, this document contains advice on interpreting scores and explanation of the quality control flags for insufficient data. The development and interpretation of the index is described in (Theroux et al., in review).

# Background

The Algal Stream Condition Index (ASCI) was developed to provide an additional line of evidence for assessing stream biological health. The index was constructed after the California Stream Condition Index (CSCI, (Mazor et al. [2016](#ref-Mazor16))), leveraging a robust statewide dataset of algae taxonomy and environmental parameters. As in the CSCI development, the ASCI is calibrated using a subset of “reference” sites that are minimally-impacted by human activity. Multiple versions of the ASCI were developed and tested, including an observed versus expected (O/E) index and a multi-metric index (MMI) and a combination of both approaches. Each index was evaluated as both a traditional versus predictive approach, wherein site-specific reference-based expectations are individually selected for each test site. The strongest performing of all index variations was a traditional multi-metric index for diatom, soft-bodied algae, and hybrid assemblages. There are now three acceptable ASCIs (ASCI\_d, ASCI\_sba, ASCI\_hybrid) for each diatom, SBA, and hybrid assemblage, respectively.

# Software requirements

R version greater than or equal to v3.4.1 or RStudio current version

This document assumes that the user is familiar with basic operations in the R programming language, such as data import, export, and manipulation. Although not required, we recommend using graphic interface for R, such as R-studio, which can be downloaded at <http://www.rstudio.com>. New users are encouraged to pursue training opportunities, such as those hosted by local R user groups. A list of such groups may be found here: <http://blog.revolutionanalytics.com/local-r-groups.html>. R training material developed by SCCWRP can also be found online: <https://sccwrp.github.io/SCCWRP_R_training/>

# Data requirements

The ASCI calculator was designed to accept a SWAMP-format taxonomy data file in .csv format (please see example file provided with ASCI package). The following column headers are required in the input file:

Table 1: Required field names and descriptions of data needed to calculate ASCI scores.

|  |  |
| --- | --- |
| Field | Description |
| StationCode | Unique identifier of site |
| SampleDate | Sampling date |
| Replicate | Replicate number |
| SampleTypeCode | Indicates sample type. Valid values are: Integrated, Epiphyte, Macroalgae, Microalgae, and Qualitative. Integrated sample types are for diatoms, whereas other sample types are for soft-bodied algae. Qualitative sample types are not used in ASCI calculation and will be excluded from analysis. |
| BAResult | Diatom count data |
| Result | Soft-algae biovolume data |
| FinalID | Must match values in SWAMP organism lookup lists (<http://swamp.waterboards.ca.gov/SWAMP_Checker/XXX>). A few common misspellings are recognized. |

# Calculating the ASCI

The ASCI package automates all of the necessary steps to calculate ASCI scores from properly formatted input files. It uses the algae taxonomy data to calculate metrics and final ASCI values for the diatom, soft-bodied algae, and hybrid ASCIs. Additionally, it compares the endpoints to the expectations, relative to a reference distribution. We have automated all of these steps, with the goal of minimizing demands on the user.

The automated steps are as follows:

1. Verify all necessary input fields is present.
2. Compare algae species names to harmonized species name list. Report any unrecognized taxa.
3. Convert all taxonomy data to presence/absence data. Flag samples with low diatom valve counts. Calculate percentage of taxa present with trait attribute information.
4. Calculate biological metrics for each of the three ASCIs.
5. Calculate metric scores by comparing observed metric values to ranges observed at calibration reference and high-activity sites.
6. Average metric scores and standardize by the mean at reference sites to generate final ASCI values.
7. Calculate percentile of reference distribution for each ASCI score (see Table 2).

Table 2: Reference distributions of ASCI scores.

|  |  |  |  |
| --- | --- | --- | --- |
| ASCI | 10th percentile of reference scores | Mean | SD |
| Diatoms | 0.78 | 0.89 | 0.22 |
| SBA | 0.82 | 0.90 | 0.21 |
| Hybrid | 0.83 | 0.90 | 0.20 |

# Instructions

1. Prepare input datafile, ensuring all required column headers are present (Table 1). This typically involves manual editing of the raw data (e.g., the spreadsheet) to match the input requirements. Note that the R package is case-sensitive, meaning that upper and lower case letters should be exactly as shown in Table 1.
2. Open RStudio (or R). The ASCI package lives at <https://github.com/SCCWRP/ASCI>. To install packages from GitHub, you need to first install the devtools package.

* install.packages('devtools')

1. Once devtools is installed, you can load it with the library() function and install the ASCI package with install\_github().

* library(devtools)  
  install\_github('SCCWRP/ASCI')

1. Once the ASCI package is installed, it can be loaded with the library() function.

* library(ASCI)

1. A sample dataset comes with the ASCI package and can be accessed when the package is loaded. You can view this dataset to see the required format for the calculation functions.

* head(demo\_algae\_tax)  
  #> StationCode SampleDate Replicate CollectionMethodCode SampleTypeCode  
  #> 1 000CAT148 8/10/10 1 Algae\_SWAMP Integrated  
  #> 2 000CAT148 8/10/10 1 Algae\_SWAMP Integrated  
  #> 3 000CAT148 8/10/10 1 Algae\_SWAMP Integrated  
  #> 4 000CAT148 8/10/10 1 Algae\_SWAMP Integrated  
  #> 5 000CAT148 8/10/10 1 Algae\_SWAMP Integrated  
  #> 6 000CAT148 8/10/10 1 Algae\_SWAMP Integrated  
  #> BAResult Result FinalID  
  #> 1 6 NA Achnanthidium exiguum var heterovalvum  
  #> 2 225 NA Achnanthidium minutissimum  
  #> 3 2 NA Adlafia minuscula  
  #> 4 8 NA Adlafia suchlandtii  
  #> 5 1 NA Amphora copulata  
  #> 6 1 NA Amphora inariensis

1. ASCI scores can be calculated using the ASCI() function with the input data. Here we are showing how the function is run with the demo data. We assign the output to a new object called results.

* results <- ASCI(demo\_algae\_tax)

1. The ASCI results are now contained in the results object. We can get some information about what to do next from the results object.

* results  
  #> An object of class asci   
  #> Scores calculated for diatoms, sba, hybrid indices for 4 unique samples  
  #> Use these functions for access: scores, Supp1\_mmi

1. From the text output, we can see that our results object includes ASCI scores for diatoms, soft-bodied algae, and hybrid indices. We can also use the scores() and Supp1\_mmi() functions to retrieve the results.

# ASCI output

The ASCI output can be accessed using the scores() and Supp1\_mmi() functions after the ASCI() function is run (see above). The scores() function returns all scores for diatom, SBA, and hybrid ASCIs.

scores(results)  
#> # A tibble: 12 x 4  
#> taxa SampleID MMI MMI\_Percentile  
#> <chr> <chr> <dbl> <dbl>  
#> 1 diatoms 000CAT148\_8/10/10\_1 1.27 0.620   
#> 2 diatoms 102PS0139\_8/9/10\_1 0.920 0.0690  
#> 3 diatoms 105DLCDCC\_5/19/09\_1 1.31 0.696   
#> 4 diatoms 105DLCDCC\_6/23/09\_1 1.34 0.747   
#> 5 hybrid 000CAT148\_8/10/10\_1 1.19 0.836   
#> 6 hybrid 102PS0139\_8/9/10\_1 0.934 0.0843  
#> 7 hybrid 105DLCDCC\_5/19/09\_1 1.09 0.507   
#> 8 hybrid 105DLCDCC\_6/23/09\_1 1.13 0.648   
#> 9 sba 000CAT148\_8/10/10\_1 1.10 0.845   
#> 10 sba 102PS0139\_8/9/10\_1 0.966 0.298   
#> 11 sba 105DLCDCC\_5/19/09\_1 0.915 0.129   
#> 12 sba 105DLCDCC\_6/23/09\_1 1.07 0.740

The Supp1\_mmi() function returns all metric score values for each individual ASCI, in addition to data quality flags.

Supp1\_mmi(results)  
#> # A tibble: 168 x 4  
#> taxa SampleID Metric Value  
#> <chr> <chr> <chr> <dbl>  
#> 1 diatoms 000CAT148\_8/10/10\_1 cnt.spp.BCG3 10   
#> 2 diatoms 102PS0139\_8/9/10\_1 cnt.spp.BCG3 8   
#> 3 diatoms 105DLCDCC\_5/19/09\_1 cnt.spp.BCG3 7   
#> 4 diatoms 105DLCDCC\_6/23/09\_1 cnt.spp.BCG3 10   
#> 5 diatoms 000CAT148\_8/10/10\_1 cnt.spp.BCG3\_score 1.07   
#> 6 diatoms 102PS0139\_8/9/10\_1 cnt.spp.BCG3\_score 0.835  
#> 7 diatoms 105DLCDCC\_5/19/09\_1 cnt.spp.BCG3\_score 0.716  
#> 8 diatoms 105DLCDCC\_6/23/09\_1 cnt.spp.BCG3\_score 1.07   
#> 9 diatoms 000CAT148\_8/10/10\_1 prop.Cyclotella 0   
#> 10 diatoms 102PS0139\_8/9/10\_1 prop.Cyclotella 0   
#> # ... with 158 more rows

The components returned by the scores() and supp1\_mmi() functions are listed below (Tables 3, 4, 5).

Table 3: Core ASCI output

|  |  |
| --- | --- |
| Field | Description |
| Assemblage | Indicates index calculated, either diatoms, soft-bodied alage (sba), or hybrid |
| SampleID | Unique identifier of the sample from concatenated StationCode, SamplingDate, and Replicate |
| ASCI\_d, ASCI\_sba, ASCI\_hybrid | Diatom, SBA, or hybrid ASCI score |
| ASCI\_d\_Perc, ASCI\_sba\_Perc, ASCI\_hybrid\_Perc | The percentile of the ASCI score, relative to the reference distribution. A minimum threshold has not been established, but low values should be considered indicative of degradation. |

# Supplemental report (Supp1\_mmi)

Table 4: Supp1\_mmi output

|  |  |
| --- | --- |
| Field | Description |
| Assemblage | Indicates index calculated, either diatoms, soft-bodied algae (SBA), or hybrid |
| SampleID | Unique identifier of the sample from concatenated StationCode, SamplingDate, and Replicate |
| Metric | Indicates ASCI calculated |
| Value | ASCI metric score, scaled |

Table 5: Metric descriptions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metric | Description | Diatom | SBA | Hybrid |
| cnt.spp.BCG3 | Count species: BCG 3 taxa | x | x | x |
| cnt.spp.IndicatorClass\_Cu\_high | Count species: high copper indicators |  | x |  |
| cnt.spp.IndicatorClass\_DOC\_high | Count species: high DOC indicators |  | x |  |
| cnt.spp.IndicatorClass\_TP\_high | Count species: high total phosphorous indicators |  | x |  |
| prop.spp.most.tol | Proportion species: most tolerant taxa |  | x |  |
| prop.Cyclotella | Proportion species: Cyclotella taxa | x |  | x |
| prop.spp.Green | Proportion species: Green algae |  | x | x |
| prop.spp.IndicatorClass\_Cu\_high | Proportion species: high copper indicators |  |  | x |
| prop.spp.IndicatorClass\_DOC\_high | Proportion species: high DOC indicators |  |  | x |
| prop.spp.IndicatorClass\_TN\_low | Proportion species: low total nitrogen indicators |  |  |  |
| prop.spp.IndicatorClass\_TP\_low | Proportion species: low total phosphorous indicators | x |  | x |
| prop.spp.OrgN.NHHONF | Proportion species: NHHONF (N-heterotrophic-high organic N (facultative)) taxa | x |  |  |
| prop.spp.IndicatorClass\_NonRef | Proportion species: non-ref indicators |  | x | x |
| prop.Surirella | Proportion species: Suriella taxa | x |  | x |
| prop.spp.OxyReq.DO\_10 | Proportion species: taxa requiring at least 10% oxygen | x |  | x |

# Data flags

Table 6: Types of flags returned.

|  |  |
| --- | --- |
| Flag | Description |
| Unrecognized taxa | All submitted taxonomy names are compared to and harmonized with the California Standardized Taxonomic Effort (STE). This harmonization process converts ambiguous species names (e.g. Amphora sp. 2) to harmonized names (e.g. Amphora). Any species names not recognized by the STE are ignored for the calculations and reported as “unrecognized names”. |
| Low valve count | Samples are flagged that had diatom valve counts <200 |
| Low trait attribute value | Samples are flagged when less than 50% of taxa present have trait attribute data available (TBD). |

# Troubleshooting and FAQ

*Bad or missing field names*

All required field names must be present in input files. Please be sure to match the field names provided above. Although we have implemented scripts to make the inputs case-insensitive, we recommend conforming to the capitalizations shown above.

*Bad or missing sample type codes*

If your data are missing life stage codes or contain values that do not match acceptable values in SWAMP, the samples will not be scored. As qualitative fraction data is excluded in the ASCI analyses, a sample must contain additional field code variables to be scored.

*Bad FinalIDs*

Bad FinalIDs typically result from misspellings, but occasionally occur when taxonomists do not conform to the SWAMP Master Taxa Lists (LINK). All unrecognized taxa are provided as exported information (verify). If you believe a FinalID is erroneously missing from SWAMP’s lookup lists, please contact the SWAMP help desk ([OIMA-Helpdesk@waterboards.ca.gov](mailto:OIMA-Helpdesk@waterboards.ca.gov)). If you believe a valid FinalID is inappropriately rejected by the scripts, contact Susanna Theroux at [susannat@sccwrp.org](mailto:susannat@sccwrp.org).

*Missing taxonomy data*

It is possible that all FinalIDs submitted for a sample are considered unrecognized taxa. In these instances, the ASCI calculator will not calculate an algae score.

*Do I need GIS data?*

The ASCI calculator does not require site-specific GIS data.

# Cautions on Score Interpretation

*Unusual environmental settings*

Most wadeable streams can be accurately scored with the ASCI (including some nonperennial streams), as long as the sampling occurred during the index period and the sampling protocol was not modified (Ode et al. [2016](#ref-Ode16)). Preliminary analyses have revealed the ASCI responds to stressor gradients in both intermittent and channelized streams (Theroux et al., in review). However, the validity for sites from unusual environmental settings is unknown and should be reviewed on a case-by-case basis.

*Samples with low counts*

Samples with low algae counts may have erroneously depressed ASCI scores. We have not established a minimum count of diatoms or SBA for validating the ASCI, but as a rule of thumb, scores that are within 10% of the specified sample size (i.e., XXX) may be used for most applications of the ASCI. Smaller counts may be appropriate for certain applications. [can we do some rough analyses to try and constrain this?]

*Samples with few trait attribute data*

Not all California algae taxa have available trait attribute information, and the absence of this trait attribute information can erroneously bias ASCI scores if it plagues the majority of species in a sample. The ASCI calculator provides a flag if samples have a < 50% of taxa with available trait attribute information. These samples should be reviewed on a case-by-case basis to determine if the scores are reflective of the biological condition.

*Missing SBA samples? Or no SBA in quantitative sample?*

*Missing diatom samples? Or no diatoms in quantitative sample?*

# References

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.

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

Theroux, S., R. D. Mazor, M. W. Beck, P. Ode, E. D. Stein, and M. Sutula. In review. “A Non-Predictive Algal Index for Complex Environments.” *Ecological Indicators*.