**Family Level Index: Explanation, Usage, and Performance**

Installing and using the Family Level Index

Installing the Family Level Index User Interface (FLI-UI)

1. Use a web browser to go to <https://github.com/mengeln/FLI>
2. Find and click the button “Download ZIP” on the right hand side of the page.
3. Once the download has completed, unzip the file.
4. Navigate to the top level of the FLI folder.
5. Double click the “FLI.exe” program. This will normally boot up the FLI-UI; however, the MMI and OE models need to be rebuilt the first time the FLI-UI is launched. This may take a few minutes (> 5 minutes). Subsequent launches are much faster.

Using the FLI-UI

1. Double click the “FLI.exe” program. The application should boot up in your web browser (specifically, at localhost:5678).
2. First, input a CSV file containing your benthic macroinvertebrate (BMI) data. The download includes demo data (FLI-master/www/BMI\_example.csv). This file should contain five columns (column names are case sensitive):
   1. SampleID: Unique identifier of the sample
   2. StationCode: Unique identifier of the site
   3. FinalID: Taxon names. These must be exact matches with SWAMP FinalIDs (but not case-sensitive).
   4. BAResult: Counts of the taxon in that sample
   5. LifeStageCode: The life stage of the taxon. May be “L” (larvae), “P” (pupae), or “A” (adult) for insect taxa. All non-insects have a LifeStageCode of “X”.
3. Next, input a CSV file containing your station data. The download includes demo data (FLI-master/www/stations\_example.csv). This file should contain three columns (column names are case sensitive):
   1. StationCode: Station names that match with station names in the BMI data
   2. Lat: The latitude in decimal degrees
   3. Long: The longitude in decimal degrees. Use negative values to indicate degrees west.
4. Select the number of organisms per subsample your BMI data (specifically, 100, 200, 300, or 500).
5. Click the submit button. The number of samples submitted will increase computation time.
6. The core results should displayed on the screen. Additional reports may be downloaded as well by selecting the options from the drop-down menu. Pressing “Get Report” will cause the report to saved/opened as a CSV file. These reports are: pMMI results, O/E results, group probabilities, and GIS results. These reports are described below.

Reports

Core report provides scores for the FLI and its two components (the pMMI and O/E). Additional data quality information is also provided.

Core report

|  |  |
| --- | --- |
| Field name | Description |
| StationCode | Unique identifier of the site, provided by user |
| SampleID | Unique identifier of the sample, provided by user |
| Count | Number of specimens in the sample |
| Excluded | Number of specimens excluded from analysis (taxa not included in normal Training Academy workshops) |
| PctAmbiguousIndividuals | Percent of individuals with ambiguous identification (typically, order-level or higher) |
| PctAmbiguousTaxa | Percent of taxa with ambiguous identification (typically, order-level or higher) |
| E | Number of common taxa expected by the model |
| O | Number of common taxa observed in the sample |
| OoverE | Ratio of observed to expected common taxa. Expect 1 at reference condition. |
| MMI | Predictive multimetric index. Expect 1 at reference condition. |
| FLI | Family level index score. Expect 1 at reference condition. |

stationsGIS: Environmental parameters, derived from latitude and longitude of the site

|  |  |
| --- | --- |
| Field name | Description |
| StationCode | Unique identifier of the site, provided by user |
| Lat | Latitude in decimal degrees, provided by user |
| Long | Longitude in decimal degrees, provided by user |
| ppt | 10-y (2000-2009) average precipitation at the sample point, in hundredths of millimeters |
| temp | 10-y (2000-2009) average temperature at the sample point, in hundredths of a degree Celsius |
| elevation | Site elevation in meters |

[Other reports need revision to match CSCI reports, but here’s a description of what they currently look like:]

captureProbs: Probability of observing each taxon in the sample.

groupProbs: Probability of membership of a site in each reference group

metrics: Raw metric values for each sample. Will vary depending on subsample size selected.

scores: Scored metrics for each sample.

Troubleshooting

If the FLI UI run into a problem, it will return a “ThereWasAnError” message to the screen. Most errors are caused by formatting issues with the input data. The following checklist may help resolve problems:

* Do the BMI and stations CSV files have all of the required columns? The column names must exactly match the above specifications, and must match the case shown above. Whitespace at the end of the name (which can be difficult to see in MS Excel) will create errors.
* Are all of the coordinates are within the state of California? This tool contains GIS data only for stations within California. No predictor data will be generated for those that lie outside state boundaries, even if only by a small amount. For sites near the state boundary, minor adjustments of the coordinates may be helpful to ensure they plot within the state.
* Are all station codes found in both input files? Station codes in the two input files must match exactly (both case and whitespace sensitive). Also, make sure that each station is only listed once in the stations data.
* Are all FinalIDs are recognized names from SWAMP lookup lists? FinalIDs must match both in spelling and in case to the tool’s internal database of recognized taxa.

**Development and Performance Evaluation of the FLI**

Introduction

Although the California Stream Condition Index (CSCI) can be used to assess perennial streams throughout California, the data requirements of the index may be out of reach of citizen scientists. For example, the CSCI requires genus-level identifications of many benthic macroinvertebrate species, while many citizen scientists are trained to produce only family-level data. Other requirements of the CSCI (such as watershed delineations, or large sample counts) make it ill-suited for volunteer monitoring or other non-regulatory applications with limited resources. Although a few indices are available, few have statewide applicability, and none employ predictive modeling, which provides biological benchmarks that are appropriate for the unique environmental settings found at each site. Therefore, citizen scientist groups working in different parts of the state have no way of assessing their streams in a comparable way. Consequently, these groups have only a limited ability to work together and share their data.

We have developed a simple predictive index that has data requirements consistent with the coursework offered by the Training Academy, yet allows citizen-scientists to learn about more complex tools, such as the CSCI. Specifically, this tool is simpler than the CSCI with respect to three requirements:

1. Taxonomic requirements match the level provided by the training academy (that is, family level for most groups of benthic macroinvertebrates). Genus of species level identifications are not required.
2. The number of individuals required for a valid sample is 100, rather than 500.
3. Geographic data requirements are be limited to latitude and longitude. No GIS skills or resources are required.

Methods

*Index development*

The FLI includes two different types of indices, both based on predictive models: a predictive multi-metric index (pMMI), and an observed to expected richness model (OE). Although both models use the same data sets to assess site condition, they assess different aspect of the benthic community, which the FLI combines together by taking the mean score of the pMMI and OE components. Developing the index followed development of the CSCI as close as possible (Mazor et al., in review).

*Index calculation*

Before either the pMMI or OE model predictions are generated, several data pre-processing steps must be done, and these steps are handled automatically by the tool. First, the station coordinates (provided by the user) are used to estimate several environmental parameters: the site elevation, long-term (1971-2000) average annual precipitation, and long-term average annual water temperature from PRISM.

Second, the benthic macroinvertebrate (BMI) data are automatically subsampled based on user’s input. Available options are subsampling to 100, 200, 300, or 500 individuals. Subsampling is essential because many assessment metrics are sensitive to the number of organisms in a sample; therefore, subsampling to a consistent sample size removes a source of bias in assessment scores. Samples with fewer organisms than the number specified by the user are not subsampled.

The final step in pre-processing the data, the BMI data are aggregated to the correct taxonomic resolution. For insects, this is the family level, although many non-insects are aggregated to the ordinal level (e.g., Oligochaeta). When multiple FinalIDs belong to the same family, the counts associated with that family are summed. Additionally, for the data to be used in the OE component, some taxa are thrown out for two reasons: 1) they are not truly benthic (e.g., Culicidae); 2) the resolution is too high, and thus the taxon is ambiguous (e.g., Diptera). For data generated at bioassessment training workshops, taxonomic aggregation step is unnecessary; however, this step is essential for calculating FLI scores for datasets that were identified to the genus or species level (e.g., for comparing volunteer data to data generated by professional labs).

When you are done, close both the browser and the R console window.

*Performance evaluation*

To assess the FLI, we looked at three aspects of its performance: accuracy, precision, and responsiveness. We calculated several performance measures and compared them to the performance of the CSCI. For the purpose of assessing the FLI, the CSCI score will be considered to represent the “true” condition of a given site. Performance evaluations of the FLI followed Mazor et al. (in review) as close as possible.

Accuracy is the ability of an index to provide consistently high scores at reference sites, with no difference across environmental settings. To measure accuracy, we looked at regional biases in scores at reference sites. In addition, we looked for relationships between FLI scores at reference sites and natural gradients using random forest models and bivariate plots.

Precision is the variability of an index. To measure precision, we calculated the standard deviation of FLI scores at reference sites. Secondly, we looked at the pooled standard deviation of sites with multiple samples.

Responsiveness is the ability of an index to change in response to stress. To measure responsiveness, we calculated t-statistics on FLI scores, comparing reference and stressed sites. In addition, we looked at relationships between FLI scores and stressor gradients.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| MMI Model | FLI-100 | FLI-200 | FLI-300 | FLI-500 |
| Metrics | Diptera Taxa  Predator Percent Taxa  Intolerant Percent Taxa  Burrower Taxa  EPT Percent Taxa  Non-insect Percent Taxa  Shredder Taxa  Clinger Taxa | Intolerant Percent Taxa  Noninsect Percent Taxa  EPT Percent Taxa  Clinger Taxa  Ephemeroptera Taxa  Plecoptera Percent Taxa  Trichoptera Taxa  Shredder Percent Taxa | Intolerant Percent Taxa  EPT Taxa  Non-insect Percent Taxa  Plecoptera Taxa | Intolerant Percent Taxa  Clinger Taxa  Non-insect Percent Taxa  EPT Percent Taxa  Ephemeroptera Taxa  Trichoptera Taxa  Shredder Taxa  Plecoptera Taxa |

Table 1: Metrics used in different versions of the pMMI model. EPT refers to Ephemeroptera, Plecoptera, and Trichoptera. Metrics appended with “taxa” are counts of unique aggregated taxa occurring in that group, while those appended with “percent taxa” are the proportion of aggregated taxa out of the entire sample occurring within that group.

Results and Discussion

In many ways, the FLI exhibited good performance, but in general performance was slightly worse than CSCI, especially with smaller subsample sizes (Table 2). Accuracy of the FLI was good, as there was no indication of significant bias from natural gradients. Comparisons of reference condition sites between ecoregions show no significant effects, and a non-linear model based on natural gradient data (e.g., soil composition, long term climatic conditions, watershed descriptors), explained none of the variance in FLI scores. A small positive bias from stream gradient was observed (Fig. 1); this bias was also observed for the CSCI (Mazor et al. in review).

For precision, there was a noticeable difference between the performance of the CSCI and FLIs (Table 1). The pooled standard deviation among repeatedly sampled sites was greater than the CSCI, and tended to increase as the level of subsampling decreased. The better precision of the FLI may reflect the simplicity of the family-level data it uses, in contrast to the more complex genus- and species-level data used by the CSCI.

The FLI was less responsive than the CSCI, as reflected by its lower t-statistics, although it was still sufficient to detect differences between reference and stressed sites. Again, larger subsample sizes improved performance: at the highest subsample size evaluated (FLI-500), responsiveness to stress was very similar to that of the CSCI (Fig. 2).

In general, differences between the FLI and CSCI were more pronounced at sites in poor condition (Fig 3). That is, differences were largest where CSCI scores were poorest. The differences between the indices were largest for smaller subsample sizes. For example, when FLI-100 evaluates sites in very poor condition may give a score 10 points greater than the CSCI on average. On the other end, the FLI-500 shows almost no bias relative to the CSCI, and remains very close to the CSCI scores at all levels.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | | |  | Precision | Sensitivity | | |
| **Model** | **F (cal)** | **F (val)** | **VarExp** | **Cal Pooled SD** | **Val Pooled SD** | **t.cal** | **t.val** | **VarExp** |
|  |  |  |  |  |  |  |  |  |
| CSCI | 1.1 | 1.3 | -9 | 0.11 | 0.10 | 23 | 26 | 53 |
| FLI-500 | 1.3 | 1.1 | -16 | 0.09 | 0.09 | 19.4 | 21.0 | 58 |
| FLI-300 | 0.6 | 0.4 | -17 | 0.10 | 0.09 | 18.7 | 20.2 | 57 |
| FLI-200 | 1.4 | 0.8 | -13 | 0.09 | 0.10 | 19.3 | 20.6 | 54 |
| FLI-100 | 0.7 | 0.8 | -17 | 0.11 | 0.11 | 17.2 | 19.0 | 48 |
|  |  |  |  |  |  |  |  |  |

Table 2: Accuracy: F-ratio for O/E scores by PSA region; VarExp is from random forest model regressing reference scores against natural gradient variables (i.e., geology, watershed size, max elevation, etc.).

Precision: Pooled SD of repeatedly sampled sites

Sensitivity: T-value for comparison between reference and stressed sites; VarExp is from random forest model regressing scores against stressor gradient variables (i.e., % urban, % agriculture, road density, etc.)

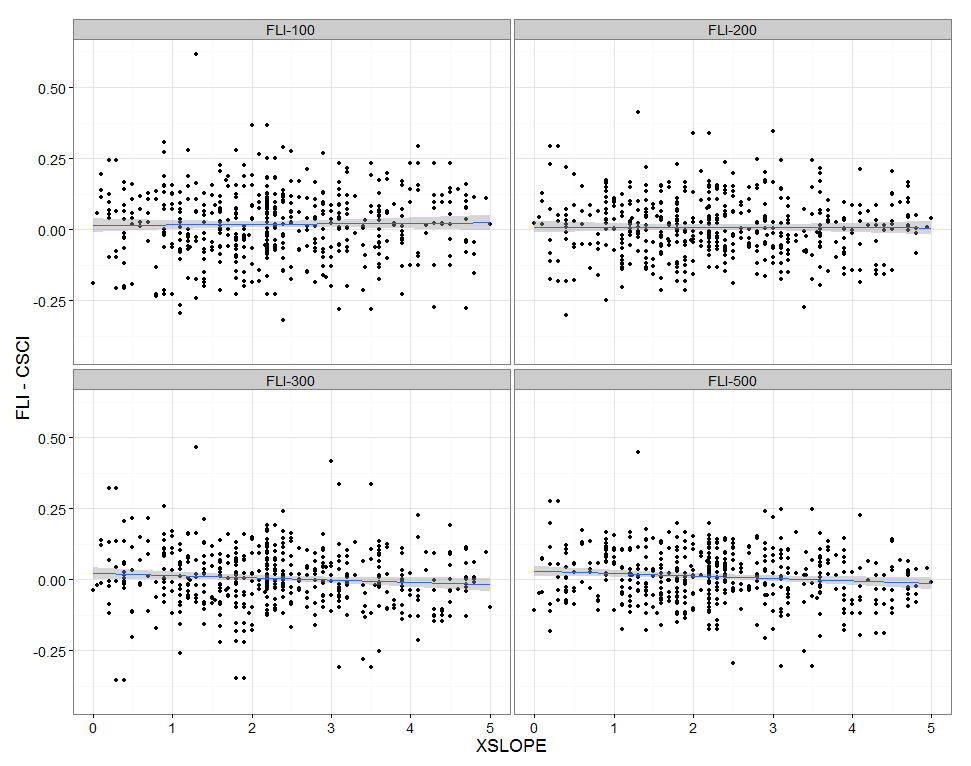


Figure 1: Residual FLI scores for reference condition sites against slope (as percent gradient).

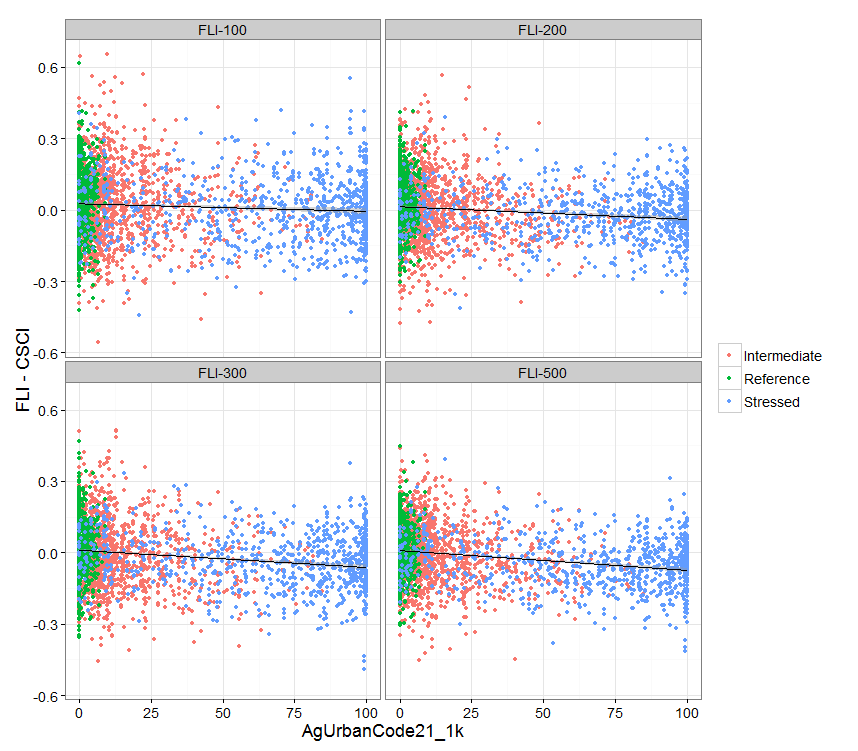


Figure 2: The residual FLI score against the level of human influence, as defined by the sum percentage of agricultural, urban, and managed land use within a 1-km radius of the site within its watershed.

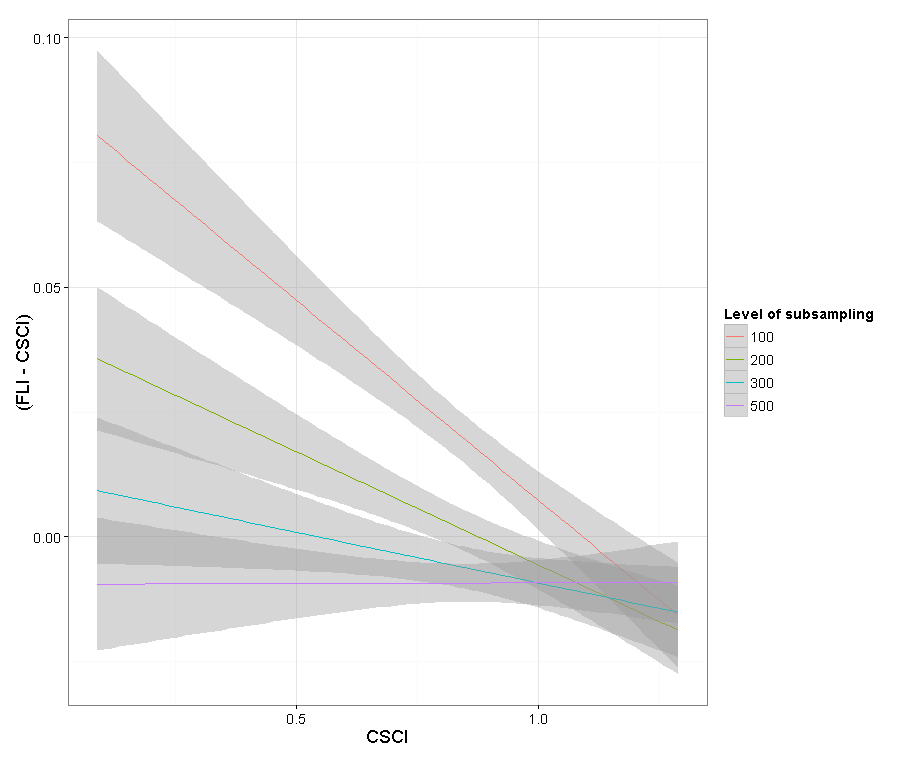
Figure

Figure 3: Residual of the FLI vs. CSCI scores.

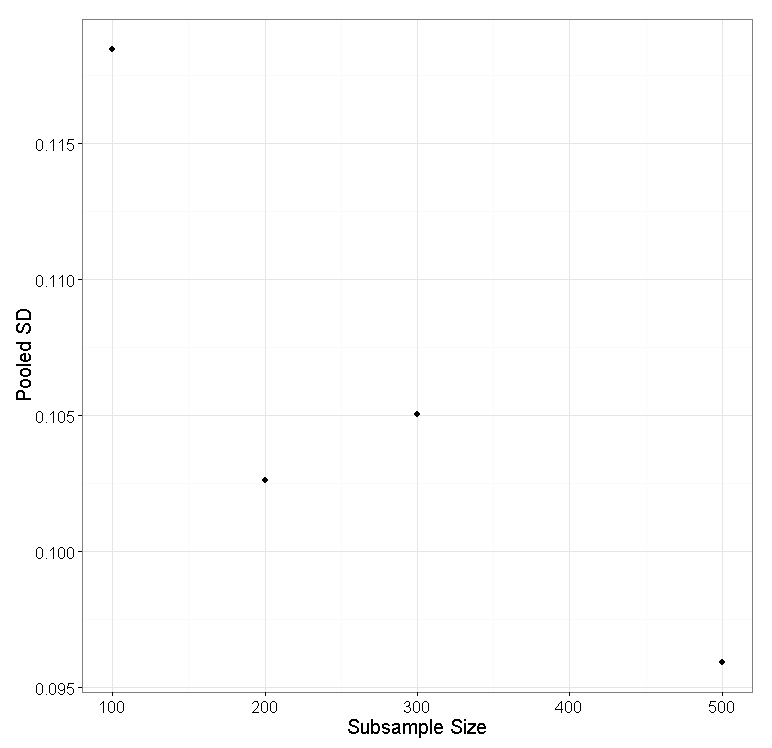


Figure 4: The pooled standard deviation of repeatedly sampled reference calibration sites as subsampling size increases.