Family Level Index: Explanation, Usage, and Performance

Installing the Family Level Index User Interface

1. 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 on first usage the MMI and OE models will be installed. This may take a few minutes.

Using the Family Level Index User Interface

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 macroinvertibrate (BMI) data. This CSV should contain five columns:
   1. SampleID: Names given to identify the sampling events
   2. StationCode: The stations where the sampling events occurred
   3. FinalID: Taxon names. These must be exact matches with SAFIT FinalIDs.
   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 may only have a LifeStageCode of “X”.
3. Next, input a CSV file containing your station data. This file should contain three columns:
   1. StationCode: Station names that match with station names in the BMI data
   2. Lat: The latitude in decimal notation
   3. Long: The longitude in decimal notation. Use negative values to indicate degrees west.
4. Select the level at which you want to subsample your BMI data
5. Click the submit button
6. Your core results should displayed on the screen. You can access other aspects of the results, such has the metric scores in the MMI component, by selecting the various reports from the drop-down menu. Pressing “Get Report” will cause the report to saved/opened as a CSV file.

Troubleshooting

If the FLI UI run into a problem, it will simply return “ThereWasAnError”. While not especially informative, it is likely the case that such errors are caused by formatting issues with the input data. Try checking the following to resolve the problem:

* The BMI and stations data inputs have all of the required columns. The column names must exactly match the above specifications, including having the same case. Erroneous extra whitespace at the end of the name (which can be difficult to notice in MS Excel) will count as a different name.
* All of the coordinates are within the State of California. This tool contains GIS data only for data points within California. No predictor data will be generated for those that lie even slightly outside of California.
* Your stations data has coordinates for all of the stations in your BMI data. Again, these names must match exactly (both case and whitespace sensitive). Also, make sure that each station is only listed once in the stations data.
* All FinalIDs are recognized SAFIT names. FinalIDs do not necessarily need to already be at the family level (the tool will automatically aggregate taxa at greater resolution to family level). However, whatever names are used, they must match both in spelling and in case to the tool’s internal database of recognized taxa.

Explanation and Interpretation of the FLI

The FLI is a hybrid of two different models: a multi-metric index (MMI), and an observed to expected richness model (OE). Though both models use the same data sets to predict site condition, they assess different aspect of the benthic community, which the FLI combinds together by taking the mean score of the MMI and OE components.

Before either the MMI or OE model predictions are generated, several data pre-processing steps must be done. First, using the station coordinates provided by the user, several environmental parameters are looked up against a reference GIS data for the State of California. These include the long-term (1971-2000) average annual precipitation, long-term average annual water temperature, and the site elevation.

Next, the benthic macroinvertebrate (BMI) data are subsampled based on user’s input. Available options are subsampling to 100, 200, 300, or 500 individuals. This subsampling will only affect those sample which have more individuals than the selected level of subsampling, e.g. a sample with only 95 individuals will always be left as-is.

To finish the pre-processing, the BMI data are aggregated to the correct taxonomic resolution. For insects, this is the family level, though 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 one of several reasons: they are excluded because, though they may be collected in freshwater streams, they are not truly benthic (e.g., Culicidae); the FinalID’s resolution is less than the resolution of the target taxonomic level, and thus are ambiguous (e.g., Brachycera); the taxon is both benthic and is not ambiguous, but never occurred in the reference sites used to calibrate the models (e.g., *Branchinecta campestris*).

The MMI begins by calculating a suite of metrics considering different aspects of the BMI data. A different suite of metrics is used based on which level of subsampling is used (see table 1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 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 MMI model. 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. Metrics listed in black are compared against a predicted value generated from a model; metrics listed in red are keep as their raw value before being scored.

Next, the station GIS data is used as predictor inputs into a non-linear model to generate expected values for the metrics assuming the site is in reference condition. These predictors are compared against the observed values to produce the residual values (for most metrics, see table 1 for metrics that are left as raw values). These are then scored by comparing ratio of the difference between the observed value and 5th percentile of the calibration data to the difference between the 95th and 5th percentile of the calibration data. This usually results in a score between zero and one, though for exceptional sites the score may be greater than one. The unweighted mean of the metric scores is the final MMI score.

The OE component uses the station GIS data to predict the probability of membership to a biotic groups. These biotic groups, which were previous delineated during the development of the California Streams Condition Index (CSCI), are used to account to account for the different communities and levels of richness that is expected in reference status sites across a range of natural conditions. These group membership probabilities are used in conjunction with a calibration data set, containing mean taxon counts for reference sites within a biotic group, in order to generate the probability of capturing a taxon at the site (assuming the site is in reference condition). Finally, the expected richness is calculated by counting the number of taxa with an expected capture probability greater than 0.5. The ratio of the actual richness observed at the site to this expected richness is the final OE score. Again, values tend to fall between zero and one, though it is possible to have a score exceeding one (that is, for the site to have a greater richness than expected given the natural conditions).

Performance

To assess the FLI, we looked at three aspects of its performance. First, the accuracy: does the index show any regional biases? Are reference condition sites in different ecoregions scored differently on average from each other? Can FLI scores be explained by a simple model using natural gradient data as an input?

Second, we are interested in the precision of the index. Specifically, when a site is sampled multiple times, there should be little variance in the score. Finally, we assess the sensitivity of the index; that is, how well can the index distinguish between sites in reference condition and those that have been impacted? Here we should hope to find that FLI scores show a strong negative response as known stressors increase.

Here we present the performance of the FLI on these measures in comparison to the California Streams Condition Index (CSCI). The CSCI uses significantly more detailed data inputs, such as a larger suite of GIS-based station data and greater taxonomic resolution, to generate predicted values for its MMI and OE components. For the purpose of assessing the FLI, the CSCI score will be considered to represent the “true” condition of a given site.

Overall, the FLI tended perform close to the level of the CSCI, though consistently worse, especially when using version of the FLI that implemented lower levels of subsampling (Table 2). Comparisons of reference condition sites between ecoregions show no significant effects, and a non-linear model which attempted to predict FLI scores based on natural gradient data (e.g., soil composition, long term climatic conditions, watershed descriptors), failed to explain the variance in either the CSCI or FLI scores. Additionally, the FLI responded to an increase in the slope of the stream (i.e., the gradient from low land to mountain streams) for reference conditions sites in the same way as the CSCI (Fig. 1).

For both precision, and sensitivity, there was a greater 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 FLI was able to produce average scores for the reference and impacted condition sites that differed significantly, though not as much as the CSCI. Here again, as the level of subsampling decreased, the power to distinguish between reference and non-reference weakened. Both the FLI and CSCI showed similar responses when human impacts of the area near the site increased, though difference between the FLI-500 and CSCI tended to be smaller than when using the FLI-100 (Fig. 2).

This difference is also seen clearly when residual FLI scores (i.e., the FLI score minus the CSCI score for the same sample) are compared against CSCI scores (Fig. 3). When the CSCI score is high (i.e., the site is in reference condition), the FLI tends to produce a similar score. However, as the CSCI score decreases, the FLI tends to deviate from the CSCI. This deviation has a positive bias, such that impacted sites tend to have a higher score with the FLI than with the CSCI. This effect is strongest in FLIs with a low level of subsampling. 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, and remains very close to the CSCI scores at all levels.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Accuracy** | | |  | **Precision** |  | **Sensitivity** | | |
| **Model** | **F (cal)** | **F (val)** | **VarExp** |  | **Within-site SD** |  | **t.cal** | **t.val** | **VarExp** |
|  |  |  |  |  |  |  |  |  |  |
| CSCI | 1.1 | 1.3 | -9 |  | 0.08 |  | 23 | 26 | 53 |
| FLI-500 | 1.3 | 1.1 | -16 |  | 0.09 |  | 19.4 | 21.0 | 58 |
| FLI-300 | 0.6 | 0.4 | -17 |  | 0.10 |  | 18.7 | 20.2 | 57 |
| FLI-200 | 1.4 | 0.8 | -13 |  | 0.10 |  | 19.3 | 20.6 | 54 |
| FLI-100 | 0.7 | 0.8 | -17 |  | 0.12 |  | 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.)

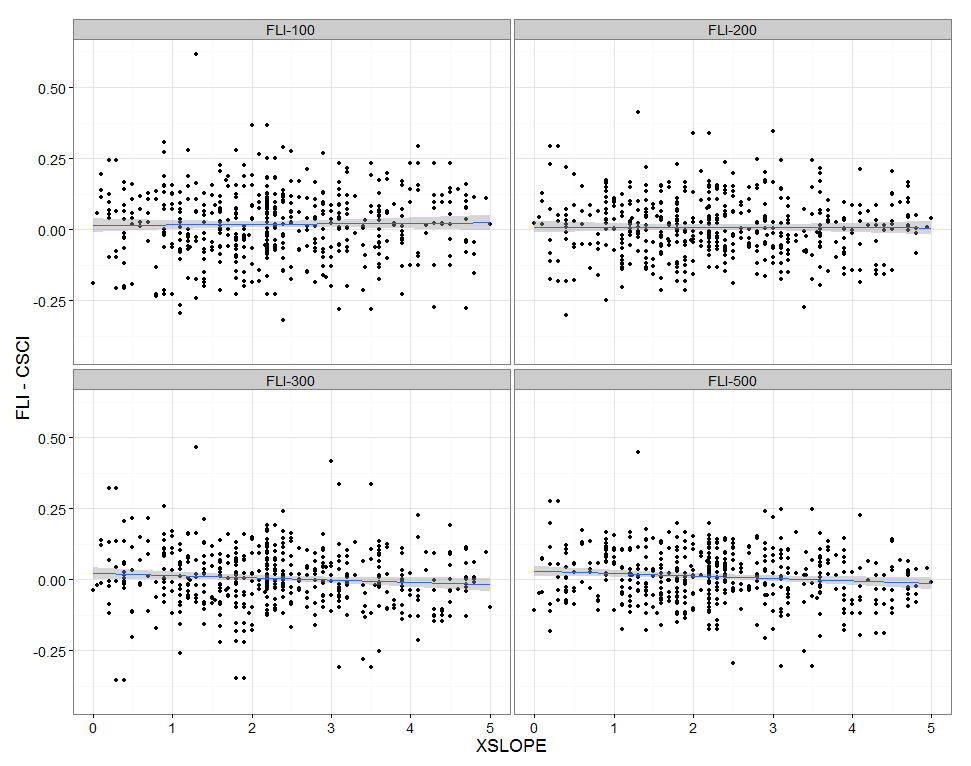


Figure 1: Residual FLI scores for reference condition sites against slope (as percent gradient).

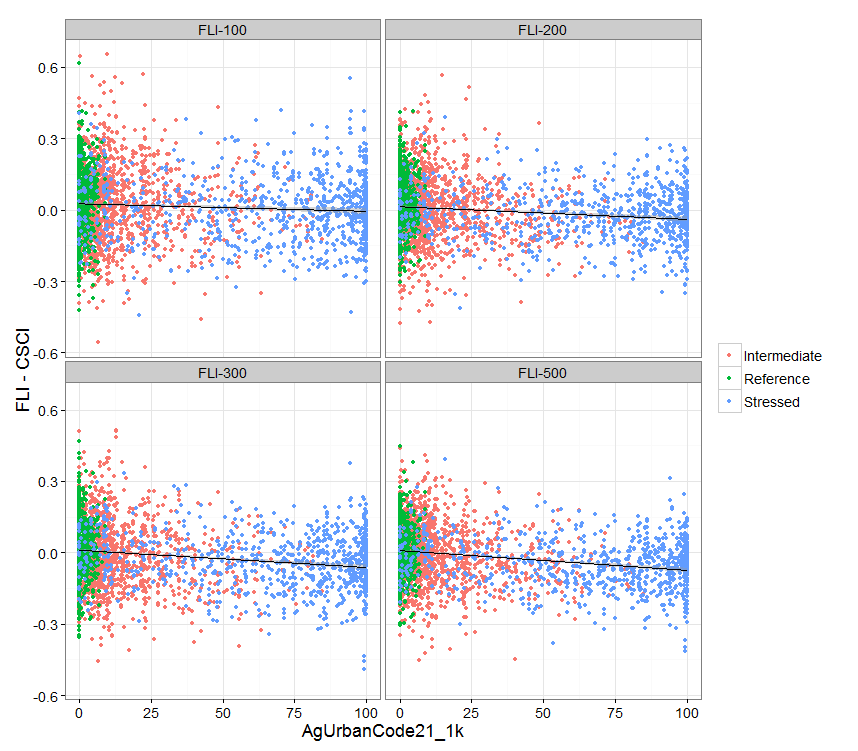


Figure 2: The residual FLI score against the level of human influence, as defined by the sum percentage of agricultural, urban, and managed land use within a 1 km radius of the site within its watershed.

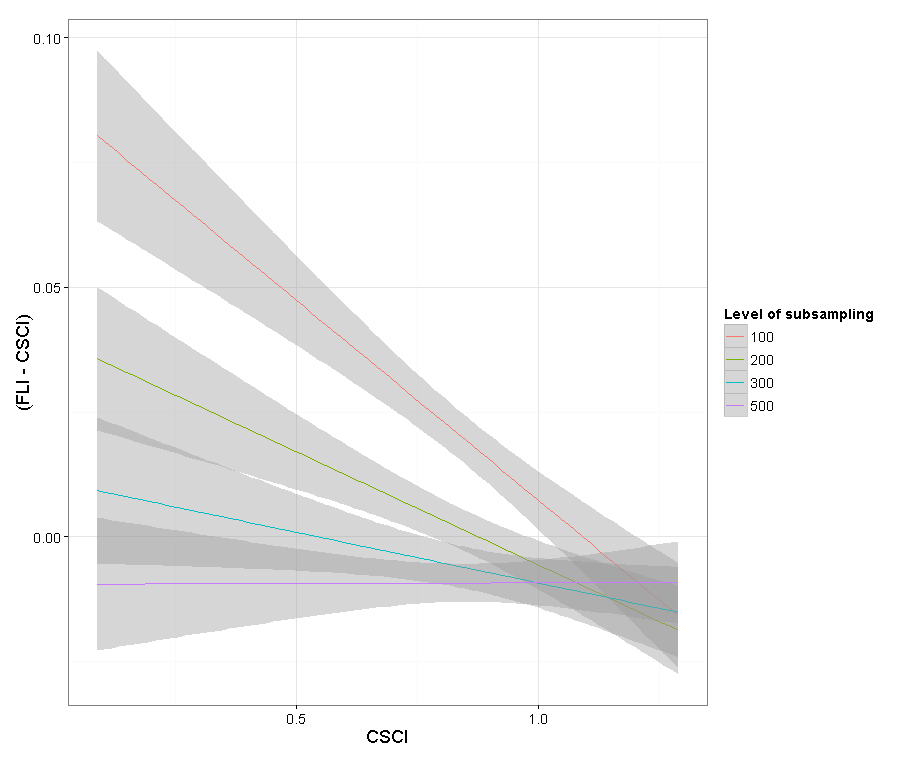
Figure

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