

MASS User Instructions

This document outlines the goals, outputs, and processes of MASS at its current state. It also discusses which user inputs and commands are involved in which processes.

Goals

The goal of MASS (Morphological Analysis of Shape and Size) is to provide a flexible and robust tool for analyzing leaf shape and morphology. Although we recognize that the utility of this tool may extend to other biological samples and systems, here we present instructions for using MASS to investigate leaves, for which the tool has been validated. To this end, a sample figure and sample data of leaves are provided at <https://github.com/gillianlynnryan/MASS> to facilitate user training on this software.

Installation

MASS source code is available as a collection of MATLAB '.m' source files that may be run within the MATLAB environment but is also provided as a standalone application (which requires MATLAB Runtime to already be installed), MASS.exe. Additionally an installer is provided, which includes installation of the necessary MATLAB Runtime. MASSInstaller.exe (not needed if you already have MATLAB 2017b or higher).

MASS Executable

Prerequisites for Deployment

Verify that version 9.4 (R2018a) of the MATLAB Runtime is installed. If not, you can run the MATLAB Runtime installer. To find its location, enter `>>mcrinstaller` at the MATLAB prompt.

NOTE: You will need administrator rights to run the MATLAB Runtime installer.

Alternatively, download and install the appropriate version of the MATLAB Runtime for R2018a from the following link on the MathWorks website: <http://www.mathworks.com/products/compiler/mcr/index.html>. For more information about the MATLAB Runtime and the MATLAB Runtime installer, see Package and Distribute in the MATLAB Compiler documentation in the MathWorks Documentation Center.

Files to Deploy and Package

Files to Package for Standalone

=====

-MASS.exe

-MASSInstaller.exe

Note: if end users are unable to download the MATLAB Runtime using the instructions in the previous section, include it when building your component by clicking the "Runtime included in package" link in the Deployment Tool.

-This readme file

Deployment within MATLAB

In addition to being run as a separate application, MASS can be run directly from MATLAB by executing the LeafAnalysis_GUI.mlapp file from the MATLAB IDE. If this option is preferred, users must ensure that the following folders are added to the path: “elliptic_fourier” and “RemoveWhiteSpace” (Right click->Add to Path->Selected Folder and Subfolders)

Sample Selection:

MASS is intended to be used with standard herbarium sheets produced by researchers and/or that can be found online in large databases. These sheets have a standard scale bar and color sample, allowing for calibration of digital images. To acquire readable outlines using MASS, however, the samples in these sheets cannot be folded, have damage around the edge, or tape over the leaves. Tape over the leaves causes issues with converting to binary and the software doesn't recognize the taped over portions as a part of the leaf, as indicated by the poor outline generated in Figure 1. A good sample has many leaves that are not damaged and do not overlap. Additionally, a good image has a high enough resolution to get accurate measurements. Low resolution images, or in which the outline is not clearly visible for individual leaves, will not produce quality analysis in MASS. Figures 2 and 3 are examples of good and poor samples, respectively.

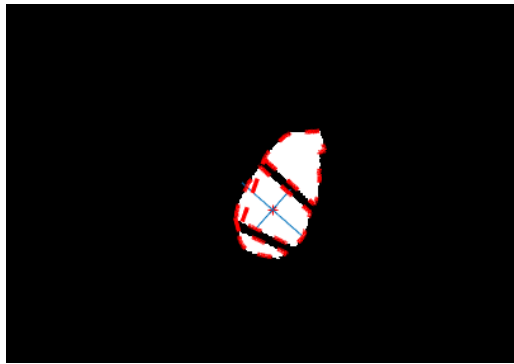


FIGURE 1: EXAMPLE OF AN OUTLINE GENERATED FROM A LEAF WITH TAPE OVER IT. ERRORS ARE PRODUCED IN THE ANALYSIS, SO THAT SUCH SAMPLES ARE NOT SUITABLE FOR ANALYSIS WITH MASS



FIGURE 2: AN EXAMPLE OF A GOOD SAMPLE HERBARIUM SHEET. SOME LEAVES ON THIS SHEET ARE CLEARLY VISIBLE, WITHOUT OBSTRUCTION AND WITHOUT TEARS, AND WOULD BE EASILY ANALYZED USING MASS.



FIGURE 3: AN EXAMPLE OF A POOR SAMPLE HERBARIUM SHEET. MOST OF THE LEAVES WITHIN THIS SAMPLE OVERLAP OR ARE OTHERWISE NOT CLEARLY DELINEATED FROM THEIR NEIGHBORS, MAKING THEM ILL-SUITED FOR ANALYSIS USING MASS.

ANALYSIS INSTRUCTIONS

1. **Start** the program by running the MASS.EXE file
2. **Configure** the analysis settings. Note that sample settings for the maple leaves test case is shown in Figure 4. *Basic analysis* is for single leaf analysis (although multiple leaves may be processed at once). *Standalone tools* are for aggregate analyses. Upon configuration, select 'Run Analysis.'

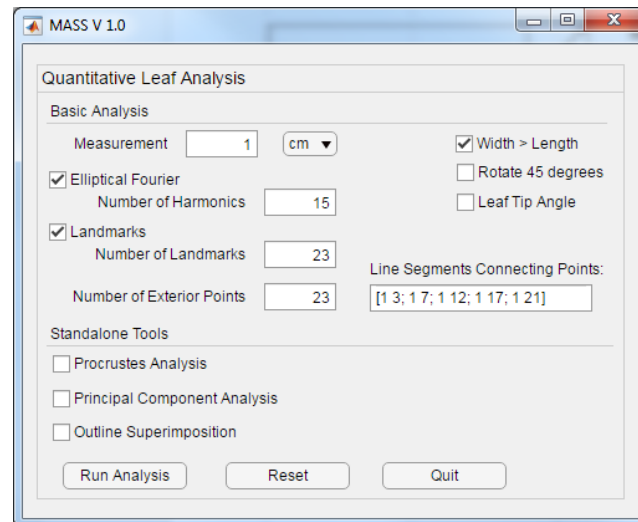


FIGURE 4: THE MASS GRAPHICAL USER INTERFACE. USERS MAY MANUALLY ADJUST ANALYSIS PARAMETERS AS DESCRIBED IN THE TEXT.

Configuration options include:

- a. **Measurement:** The line segment to be selected on the ruler for scale. Default is 1cm, but the measurement can also be in inches.
- b. **Elliptical Fourier:** The number of harmonics in the series can be selected for elliptical fourier descriptors.
- c. **Width > Length:** Some leaves are wider than they are long. If your leaves fall into this category, check this box to ensure that the rotation and dimensions will be accurate.
- d. **Leaf Tip Angle:** Check this box to measure the leaf tip angle of the leaves.
- e. **Landmarks:** Select the number of landmarks present in the leaves.
 - i. The number of landmarks are the total number of all landmarks present in the shape.
 - ii. The number of exterior landmarks are those present on the outline of the leaves. You would omit the landmarks that in the interior of the shape.
 - iii. The line segments connecting points will determine how the visualization of the landmarks will be displayed. In the example, to visualize a maple landmarks, not only do the exterior landmarks have to be connected, but there should exists line segments connecting landmarks 1 to 3, 1 to 7, 1 to 12, 1 to 17, and 1 to 21.
- f. **Procrustes Analysis:** Performs Procrustes analysis on a landmark data file to be selected by the user. If this analysis includes samples from more than one image/herbarium sheet the user must compile the necessary file before initiating analysis.
- g. **Principal Component Analysis:** Performs principal component analysis on a data file to be selected by the user. If this analysis includes samples from more than one image/herbarium sheet the user must compile the necessary file before initiating analysis. Files can either be:
 - i. Harmonics files if in the case of elliptic fourier, or
 - ii. Delta x/Delta y or displacement files if in the case of Procrustes analysis

- h. **Outline superimposition:** Performs outline superimposition a data file to be selected by the user. If this analysis includes samples from more than one image/herbarium sheet the user must compile the necessary file before initiating analysis.
3. **Sizing.** Upon selecting 'Run Analysis' a file explorer window will open and prompt you to select an image for analysis. After selecting an image file, the image will open in a new window, and you will be prompted to draw a box around the scale bar to magnify it, as shown in Figure 5A. Use the mouse to drag a box around a portion of the scale bar you wish to make your selection. Once a selection is made, a dialogue box will ask you to confirm the selection, as shown in Figure 5B. A new figure will pop-up with the magnified portion of the scale bar. Then, left-click on one of the tick marks on the ruler, and stretch the line to the other tick mark on the ruler, as shown in Figure 5C. Right click to mark your line, and when prompted confirm the line segment selection.

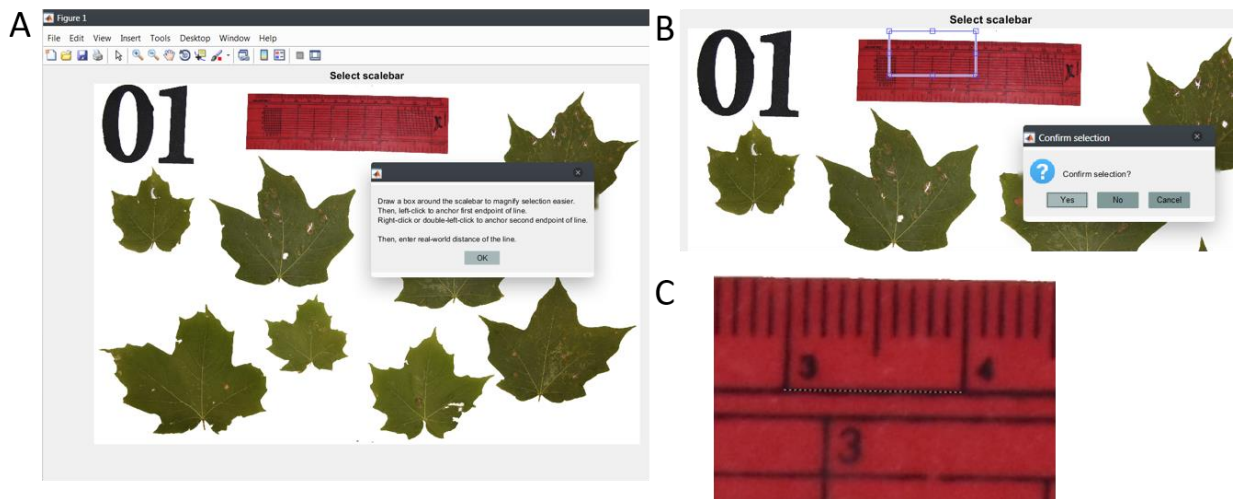


FIGURE 5: EXAMPLE OF USER-MEDIATED SIZE-SCALING PROCESS IN MASS. A) EXAMPLE OF FIGURE OPENING AND PROMPTING USER TO INITIATE SCALING SELECTION. B) USER SELECTION OF SCALE BAR WITHIN THE IMAGE C) MANUAL SELECTION OF RELEVANT SIZE-SCALE BY USER.

4. **Leaf Selection.** Upon successful input of the scale bar information, previous windows will close and a new window will open showing a grayed image, as shown in Figure 6. To select an individual leaf, the user should use their mouse to draw a polygon around the selected leaf by left clicking to drop connected points. Once the points are connected, the window will close. Leaf Selection should include the entire leaf, and should exclude all other objects for optimal analysis.

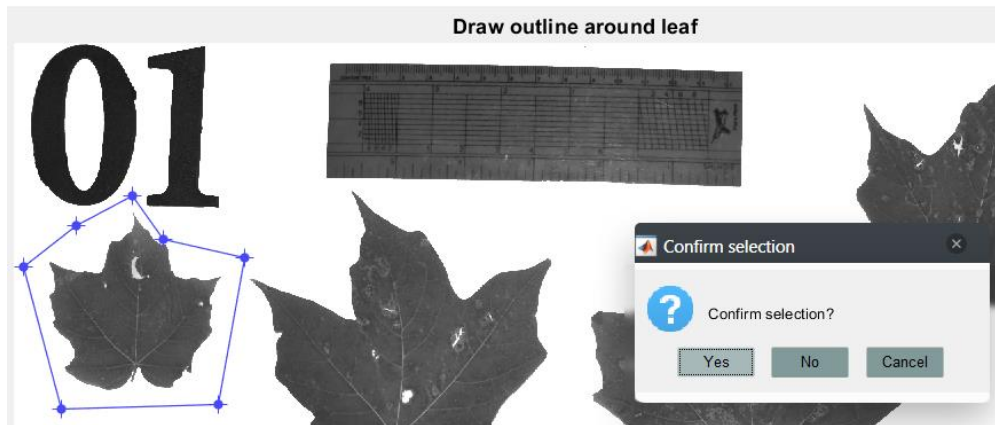


FIGURE 6: EXAMPLE OF USER'S SELECTION OF A SINGLE LEAF FROM THE CHOSEN IMAGE. THE USER-INPUT POLYGON IS SHOWN IN BLUE. LEAF SELECTION SHOULD EXCLUDE ALL OTHER OBJECTS (E.G. OTHER LEAVES) AS THEY WILL DISTORT OUTLINE ANALYSES.

Upon confirming the leaf selection, the leaf will be isolated from the background, smoothed, rotated about its major axis, and then have the outline traced. The length, width, area, Length/width ratio, fluctuation asymmetry (FA), and centroid of the image will then be displayed in the command window in Matlab (if open) and will be saved to the "measurements_YYYY_MM_DD_HHMMSS.csv" file.

5. **Leaf Tip Angle (Optional).** If selected, this analysis will prompt the user to drag a box around a portion of the leaf where the user wants to measure an angle, such as a leaf tip, as shown in Figure 7A. The confirmation prompt will pop up for this selection as well.

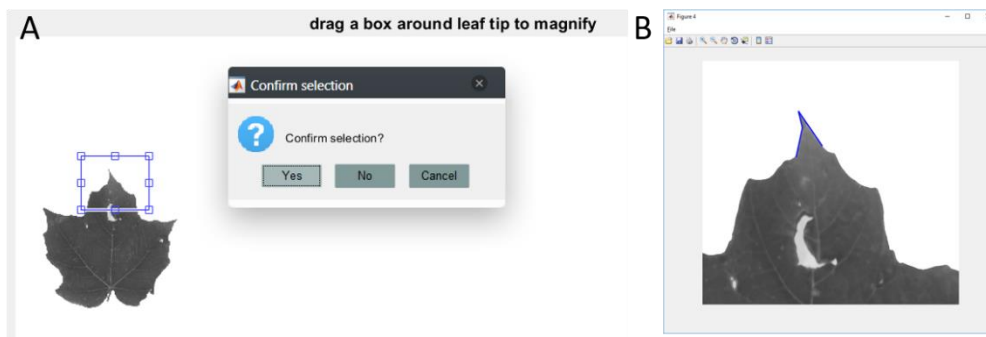


FIGURE 7: EXAMPLE OF USER'S SELECTION OF A LEAF TIP FOR ANALYSIS. A) USER SELECTION OF THE REGION OF THE LEAF TO BE EXAMINED B) SAMPLE USER SELECTION OF AN ANGLE.

Upon selection of the relevant region of the leaf, a second window will open, as shown in Figure 7B. The user then clicks on the points along the leaf tip, and MATLAB measures the angle between the two lines. Users should start these points from either side of the leaf, the 'tip' should be the center

point. The angle is displayed in the Command Window (if MATLAB is open), and is also written to the measurements file described in step 4.

6. **Elliptical Fourier Analysis(Optional).** If the option is selected, the Elliptical Fourier curve will be generated using the user-input number of harmonics. Generally, 15 harmonics is sufficient to recover overall leaf shapes described here, but more harmonics may be required to capture finer details.

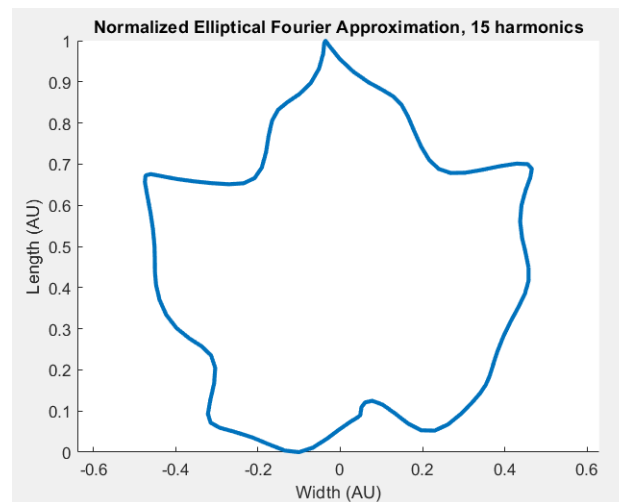


FIGURE 8: EXAMPLE OF AN ELLIPTICAL FOURIER CURVE GENERATED FOR THE PREVIOUSLY SELECTED LEAF, USING 15 HARMONICS.

By default, the outlines will be exported to the file with the naming format of outlines_YYYY_MM_DD_HHMMSS.csv. Outlines will be formatted using the following format: [x1, y1; x2, y2; ...; xn ,yn], and can be graphed or analyzed further with post-processing. All of the coefficients for the Fourier series ($4 \times$ number of harmonics) will be exported to the file with naming format of all_coefficients_YYYY_MM_DD_HHMMSS.csv. These may be useful for later principal component analysis of the elliptical Fourier fitting.

7. **Landmark Analysis (Optional).** Landmark analysis is a part of the individual leaf analysis workflow. If selected, the analysis will begin with the display of rotated, grayscale, and cropped image of the leaf will be displayed, as in Figure 9A.

The user should select the landmarks of interest in each, in order, by clicking on the image with their mouse. An example of completed landmark selection is shown in 9B. Order matters during landmark selection, as MASS compares landmarks between samples based on the order they were selected, so users should plan their landmark selection plan ahead of time. Landmarks may be unselected by pressing backspace, and the user indicates they are finished with selection by double click on the last landmark.

Upon selection, a visualization of the leaf's landmarks will be shown in a new plot, with lines connecting individual landmarks based on the user input from the initial MASS GUI. If the user selects interior landmarks, they will also be recorded, but for now landmark visualizations only support exterior features.

Landmarks will be stored for later use in the following format: x1, y1; x2, y2; x3,y3;...xn,yn. By default, the landmarks will be exported to the file with the naming format of landmarks_YYYY_MM_DD_HHMMSS.csv.

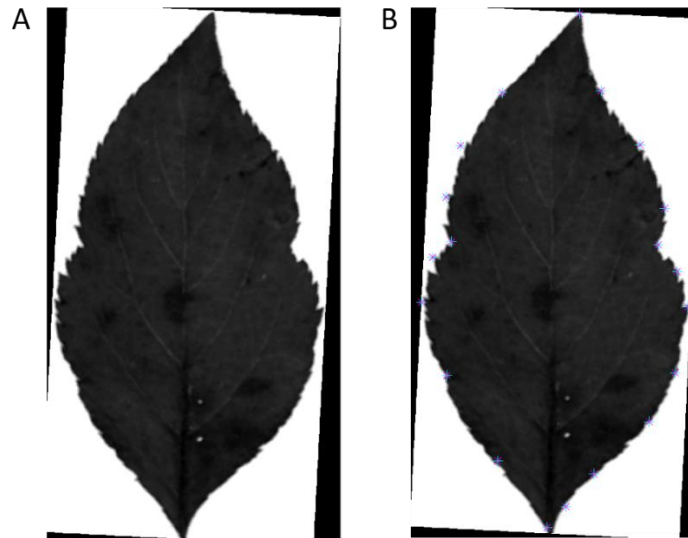


FIGURE 9: EXAMPLE OF LANDMARK SELECTION. A) THE SCALED, GRAY SCALE IMAGE THAT A USER STARTED THE ANALYSIS WITH. B) THE GRAY SCALE IMAGE DISPLAYING THE USER-SELECTED LANDMARKS, INDICATED AS BLUE '+' SYMBOLS.

Upon completion of the selected basic analyses, a dialogue box near the MASS GUI will ask the user if they want to process any more leaves, as shown in Figure 10. If the user wishes to process more leaves from the same sheet, without repeating the scale bar calibration they should select 'yes.'

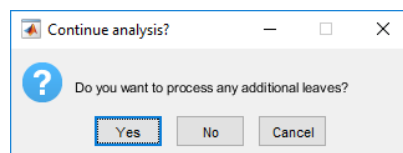


FIGURE 10: DIALOGUE BOX ASKING IF USER WOULD LIKE TO PROCESS ADDITIONAL LEAVES.

8. **Procrustes Analysis (Optional and standalone)** Procrustes analysis is an aggregate analysis that requires information from more than a single leaf. This analysis describes how one shape is fitted into another based on landmarks. In MASS the data of multiple leaves are fit to the mean, where the mean is taken across all of the samples provided for analysis. This analysis is generally run as a standalone process, as users may need to compile a csv file containing data from many samples for this analysis.

Upon initiating Procrustes Analysis a file dialogue will open, prompting the user to select a CSV file containing landmarks for analysis. (The upper left corner of the dialogue will identify this request as for Procrustes Analysis if more than one standalone analysis has been initiated.) The file should contain lists of landmarks formatted as described in the Landmarks analysis instructions. MASS will output the dissimilarity measure, a measure from 0 to 1, with 1 least similar and 0 most similar, and the size

similarity, 100 minus the factor which you would have to shrink the second. MASS will also display a plot comparing the individual samples to the mean, or standard, landmarks. An example of such a plot, for the sheet of maple leaves displayed in Figure 5, is shown in Figure 11.

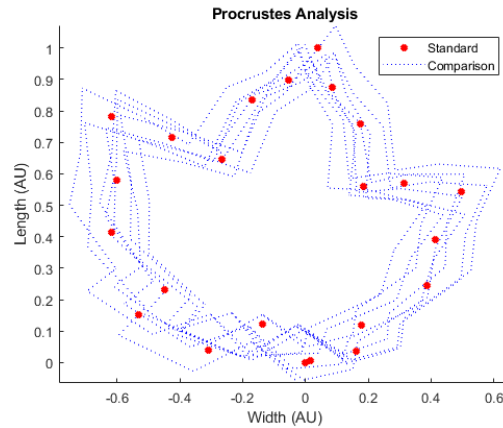


FIGURE 11: EXAMPLE PROCRUSTES ANALYSIS PLOT, DISPLAYING COMPARISONS FROM INDIVIDUAL SAMPLES IN BLUE, AND THE MEAN, OR STANDARD, LANDMARKS IN RED.

Data resulting from Procrustes analysis will be exported into three files. Each file also includes the dissimilarity and size similarity measure of each leaf. The files are as follows:

- Delta x/Delta y of all landmarks required to shift to the mean, saved as a CSV file (Procrustes_delta_x_y_YYYY_MM_DD_HHMMSS.csv)
- Net displacement required to shift each landmark to the mean, saved as a CSV file (Procrustes_distance_shifted_YYYY_MM_DD_HHMMSS.csv)
- The location of the mean landmarks will be saved as a CSV file (mean_landmarks_YYYY_MM_DD_HHMMSS.csv)

- Principal Component Analysis (Optional and standalone).** PCA analysis requires a CSV file containing the data the user wishes to analyze. Variable name headers should be included in the CSV columns as they will be the basis for PCA analysis in MASS. Users need not have generated this file using MASS, but each of the output files MASS generates includes necessary headers to allow them to be used for PCA.

Upon initiating PCA analysis, the user will be prompted to select a CSV file, and MASS will generate plots displaying the resultant data. The total possible number of graphs is generated from how many columns of data your CSV file contains, however MASS will display only the PC2 vs. PC2 graph and the PC coefficients biplot to minimize the number of windows popping up. An example of PCA scores for EFD coefficients of the maple leaves from Figure 5 are shown in Figure 12.

All PCA data (even those not displayed in the sample plots) are saved to file in CSV files with the naming formats `pca_coefficients_YYYY_MM_DD_HHMMSS.csv`, `pca_scores_YYYY_MM_DD_HHMMSS.csv` and `pc_variance_YYYY_MM_DD_HHMMSS.csv`

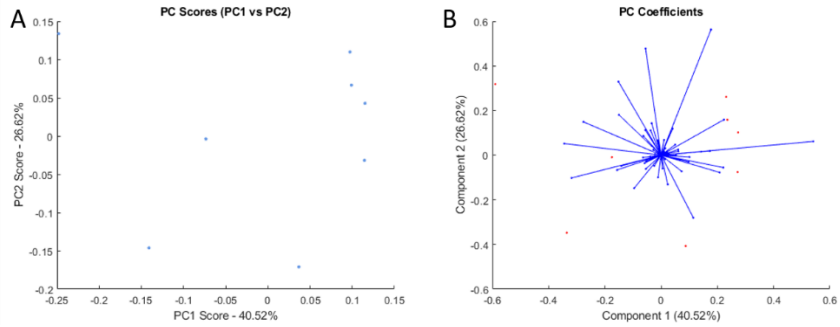


FIGURE 12: EXAMPLE PCA PLOTS GENERATED BY MASS A) PC SCORES FOR PC2 VS PC1. B) PC COEFFICIENTS BIPLLOT.

10. **Outline Superimposition (Optional and standalone)** Users may want to view outlines superimposed aggregately. This function is useful to aid such visualization. Upon initiation of this tool, users will be prompted to select a CSV file containing EFD outlines (as formatted from the MASS analysis tool). MASS will output a plot of all the individual outlines, as well as the mean outline, as shown in Figure 13. The coordinates of the mean outline will also be saved to file in a file with format mean_outlines_YYYY_MM_DD_HHMMSS.csv.

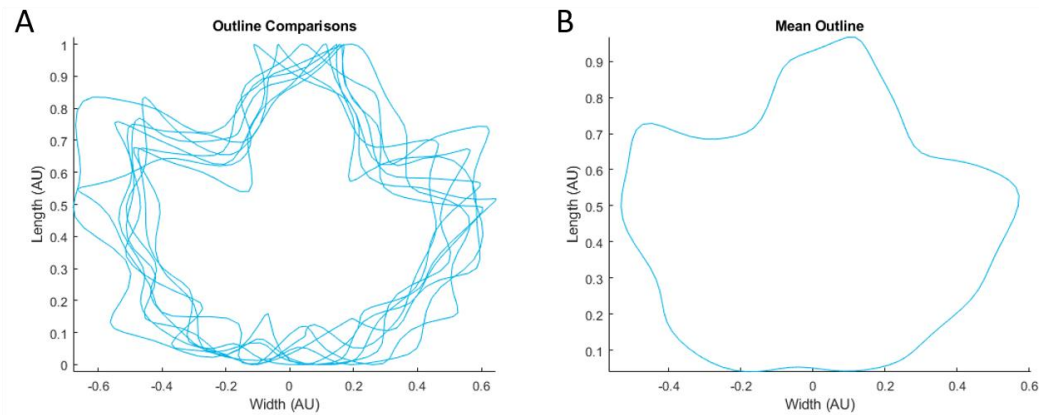


FIGURE 13: EXAMPLE OUTLINE SUPERIMPOSITION OUTPUT PLOTS BY MASS A) OUTLINE COMPARISONS FOR ALL INPUT SAMPLES B) MEAN OUTLINE, CALCULATED FROM INPUT OUTLINES.

Sample Dataset

The image shown in Figure 5, maple_tree_sample_01.jpg, is provided as a sample set to aid in user training.