YouTube video link: https://www.youtube.com/watch?v=_oOI6bqL5a4

Intro Page of app

Hello, and welcome! This video is an introduction to EMMALAB, a standalone app created using MATLAB app designer that is used to perform an end-member mixing analysis (or EMMA) on a mixture dataset. This video will not explain what an end-member mixing analysis is, but will rather introduce the workflow for performing this analysis as if the user understands these concepts.

Excel template for data

Before using the app, your data must be set up appropriately so that the data may be read into the app properly. Here is the included template, with columns representing different variables that are used within the app. The required format for the data will be explained later. The following identifier columns **must** be filled with data in order for the application to work: Sample Location, Data Type, and Sample Date. In addition, the columns starting from column 9 onward must be filled with the users chemical data, and there must be at least two columns of chemical data. Columns 4 – 11 are left for the user if they desire to use those variables for plotting time series data, but may be left blank. Please note that the user should not change the names or position of the first three columns.

Of the required identifier variables, Sample Location, Data Type, and Sample Date, the format of the data is simple. In the first column, the name of the location where the sample was taken should be placed. Another column may be used to place unique identifiers to the data, but should not be used for this column. For example, I would like to represent a body of water that we'll call Grand River, which I collected 100 samples from a single location and would like to perform a mixing analysis. Each sample may have its own unique name, such as "Grand River Sample 42", but I should simply label the sample in the "Sample Location" column as "Grand River".

The next required column of data is the Data Type. For this column, all of the locations that I would wish to perform EMMA on would be labeled as "Mixture" with a capital "M", and any endmember I had sampled for the analysis would be simply labeled as "Endmember".

The last required column of identifier data is the Sample Date. If the user collected multiple samples in the same location on a single date, only one sample from the sample date should be used. Otherwise, these samples should be classified differently within the Sample Location column.

With any of the elemental data that I had collected, that would be placed starting in the ninth column, meaning column number 9 would hold the first variable of chemical data that the user would wish to use in the analysis. Please note that all chemical data should be included in

however many columns are required. For example, if I measured 40 different chemical variables, I would place all of the data on this spreadsheet, regardless of whether the variable was thought to behave conservatively or not. This will be decided later on the app, and elements may be included or excluded in the analysis within the app itself. As a note: If the user includes chemical data with symbols (such as > or < sign) or just '0', the app will automatically replace those values with the minimum value in the column divided by two. If this is not desired by the user, all symbols or zero values should be adjusted by the user before submitting the data into the app.

After filling out the spreadsheet with your own data, it should look something like this.

Basic example of filled-out excel template.

You may want a more detailed spreadsheet if you would like to plot certain variables on a time series plot. If all of the required columns are filled out, save the file, then you are ready to use the app! Save the file.

App Window

When you are ready to start the analysis, launch the app. This introduction window will appear, and to start, the user should select the "Data" tab.

Data Tab

Click the "Load Dataset" button and a dialog box will appear, listing the files in the current folder. Navigate to the data organized following the template. Select the file and press "open". Your data will automatically populate within the table to your right, and your defined locations and endmembers along with the range of sample dates will appear on your left. Use the date dropdowns to select the date range that you would like to include in the PCA and EMMA. I will just use the default, which is the full extent of the data. Additionally, you may select here which locations or potential endmembers you would like to include in the analysis. Theoretically, you might begin this process with a handful of endmembers that you believe may explain the mixture data but are not yet sure. Including any number of endmember will not affect the results of the PCA so a good rule of thumb is to first include all potential endmembers. They may be removed before the mixing calculation, which will be shown later. However, at least two endmembers must be selected in order for the mixing calculation to work. After the desired date range, locations and endmembers are selected, click "Set Preferences" and the indicator light will change to green to show that the preferences have been set.

Plot Options

Now we will set plot options for our location or locations, and our endmembers. The location or locations in our dataset will appear in this “Location” box to the left. In our case, there are three locations that we would like to include in the analysis. In many cases, the user may want to only view a single location. Either cases will work within the analysis, but we will use the three locations in this example.

There are several plots that are used to display the data within the app, and each location and endmember should be classified differently. Here, the user may define the appearance of the location and endmembers by selecting a symbol and a color. I will use a single symbol, the circle, to represent my locations and differentiate between the locations using color.

Set the three locations symbols and colors

Now I’ll do the same thing, but for the endmembers. To assign a symbol and color to each endmember, I will select the “Endmember” button under the Data Type box, which will automatically refresh the above box with the names of my endmembers. I will choose both different symbols and colors for each of my endmembers.

Set endmembers symbols and colors

At this point, if I wish to save the categories I’ve defined by the locations, endmembers, and date range that I’ve selected, I would click on this button “Save Preferences”, which will open a dialog box which allows me to name the file and save it as a .mat file. This allows the user to skip through the first steps if the application is opened again. To upload the save file in a different session, simply click the “Load Preferences” button, navigate to the previously saved .mat file, and click “open”. Note that from uploading this file, the preferences are set.

Now that the different locations and endmembers are categorized by symbols and colors, I may begin to decide which of the chemical variables in my dataset may be used as tracers.

Correlation Matrix tab

Click on the Correlation Matrix tab, which show a bivariate scatter plot of our mixture data points within our selected time frame. The default with the app is to only show the first two variables in the dataset, and the user may iteratively include or exclude the variables they believe may behave as tracers. There is no strict guideline to what relationship on the scatter plots indicates a truly “conservative” relationship, but generally, if the trend of the data is linear, then the variable may be considered conservative and used in the analysis. After the tracer set is chosen, we may continue to the PCA tab.

PCA tab

This plot shows the initial spread of the data in 2-dimensional PC-space or U-space using the first and second principal components, with the locations and endmembers as defined in the "Plot Options" tab. Here, the user may continue to alter the tracer data to influence their decision with the mixing analysis. The PCA plot and percent explained table will be automatically updated by the removal or addition of a single tracer. The endmembers may also be averaged together to view their position on the PCA plot by selecting the "Averaged Endmembers" option within the "Endmember View" panel. From here, the user may decide which endmember sufficiently circumscribe their data to perform the mixing analysis.

3D Tab

The user may also view the data in 3D to observe the spread of the data and how well the endmembers circumscribe the data cloud, and click on the rotate view option to move view the data at different angles.

Residuals tab

The residual tab contains the ability to create residual plots for each of the selected tracers and a scree plot. The user should select the number of principal components they wish to include in the analysis, selected the tracer they would like to create a plot in the dropdown, and then click "Calculate". The residual plot will appear on the plot below. Many studies use these plots to determine the number of principal components to include within their mixing analysis, where if there is structure in the residuals, the structure may diminish by iteratively increasing the number of principal components used to represent the data. This feature is mainly important for those who wish to perform their calculation in U-space, but we suggest a separate method that will be explained later. To create a scree plot from the current PCA, simply press the "Scree Plot" option in the "Plotting Options" panel and the plot will update.

PC scores tab

Additionally, the PC scores for the locations are shown on this tab with their sample dates for easy viewing. If the user wishes to use this data outside the app, they may click on the "Save PC Data" button, which will prompt the user to insert a name for both the coefficient table and the PC scores tables. Both are saved in separate .txt files to the current folder.

Mixing calculation tab.

Once the user decides which endmembers circumscribe their data, they are ready to perform the mixing calculation. If you included more endmembers in the initial view than you would like

to include in the final calculation, simply navigate to the “Data” page, deselect the endmembers you would like to exclude, and then press set preferences. Then navigate to the Plot Options tab and click on the removed endmember to remove them from the analysis. Back on the mixing calculation page, you can see in the mixing summary under “Endmembers in analysis” which endmembers are being used in the calculation. You may also review the locations and tracers that are being used in the analysis to ensure that these are the variables you would like to use in the mixing. Next, you may decide the method in which you perform the mixing calculation. We would recommend performing the mixing calculation in tracer space, that is, using the standardized or z-scored original data for the mixture and endmembers in a system of equations. This minimizes error in the calculation by including more information than is retained in the principal component score data. However, if the user would like to perform the calculation using the principal component scores or in U-space, the option is available to calculate using either the first two or three principal components. When you have determined which calculation to use, click on the “Calculate Percent Contributions” button. Notice the “calculation status” light turns red, and will turn green once the calculation is finished. We used MATLAB’s function “fmincon” to calculate the mixing, the specifics of which can be found on the supplementary file or under the script “AnalyzeEMMA.m”. The table to the right will update with the mixing calculations, with the first column showing the location, the second showing the date, and the rest of the columns showing the name of the endmember and the fraction contributed to the mixture for the given sampling day. Each endmember in a row adds up to 1.

Mixing Plot Viewer tab

The user may choose to plot these fractional contributions on a stacked bar graph under the mixing plot viewer tab. Select the location of interest from the dropdown, and select the “Plot % contribution over time” button. The plot to the right will be updated with time on the x-axis and fractional contributions on the y-axis.

Predicted v. Observed tab

A common method to check for how well the mixing model predicted the fractional contributions accurately is by taking the mixing contributions and calculating what the concentrations of the various tracers would be for an individual mixture sample given only the fractional contributions of each endmember. Select from the dropdown the tracer that you would like to see whether it was predicted well by the model. Select “Plot Prediction” and wait for the plot on the right to update. The table below the “Plot Prediction” button will update and show the R^2, slope – 1, and y-intercept for the observed values plotted against the predicted values. To view other tracers, click the clear plot button, select a new tracer from the dropdown and press the “Plot Prediction” button again. The plot and table will update with the new tracer predictions.

Error Calculation

To check the error within the endmember sample sets themselves, we used the method of jackknifing the mean of the endmembers datasets. This method can take a long time to complete, especially if the users endmember datasets are large. The process that occurs here is that each endmember dataset has one observation iteratively removed, replaced by the average of the remaining samples, and is continued with every sample. Each new jackknifed value is then treated is an averaged endmember used in the mixing analysis. For example, if I had 10 precipitation samples, I would have averaged all of those values together for use in the mixing analysis, as is done in the app. With this technique, the jackknifing creates 10 new averaged together values, which we then treat as 10 different averaged endmember datasets to use in the mixing calculation. If I had three endmembers each with 10 different samples, then we would take each sample from each endmember and create every possible combination between the jackknifed averaged endmember values and calculate the mixing result. For our example of three endmembers and 10 samples in each endmember, that would be performing the mixing calculation 1000 times. Every possible combination of the three endmembers is calculated for a single sampling day, averaged together, and the standard deviation taken. How much these results deviate from each other is the result used in the error analysis. Thus, if someone had three endmembers but with 50 samples each, the mixing calculation would be performed 125,000. Depending on the number of cores the computer has, this calculation could take quite long to complete. The user should take this in mind when considering this error option. In our example, we'll use a significantly reduced dataset of three endmembers, each with only three samples. To begin the error calculation, select the "Calculate Percent Error". The status button will change to red and will change back to green once the calculation is complete. Once the calculation is finished, the time elapsed box will update to the amount of time (in minutes) the app took to complete the calculation. With the river dataset used in the paper, there were 10,000 different iterations and the calculation usually took ~6 hours. To plot the error, select the location you would like to view under the dropdown and select the "Plot error" button. The average input of the endmember along with the error bars for each endmember is plot over time. Additionally, if the user would like to save the averaged fractional contribution values and standard deviations from the jackknifing, the user may use the "Save Jackknife Data" button and assign a name for both tables, which will be saved as .txt's to the current working folder.

At this point, the steps within the mixing calculation are complete, but if the user would like to observe general time series trends in their data, you may use the "Plot Viewer" tab.

Time series tab

This page will allow you to view any of the variables on your original, uploaded spreadsheet and view their relationships over time. The default x-axis value is the sample date, and the y-axis may be changed to any variable. By selecting a variable, the plot will automatically update with

whichever location the user initially chose to include in the mixing. Additionally, the x-axis, y-axis, and title may be manually adjusted by the user.

Saving Graphs

To save any graph within the app, hover your mouse over this icon and a dropdown will appear with options to save the graph as a .png file, to copy the image to the clipboard, or to copy the image as a vectorized graphic. We will save our image as a .png by clicking on the first icon, inserting a name for the file, and clicking save. The image will be saved to the current workspace. This method may be used to save any graph throughout the app.