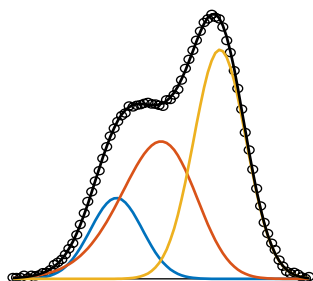


AnalySize
Software for grain size unmixing and analysis
v0.9.0r

August 24, 2015



Contents

1	Introduction	2
2	Version history	3
3	System Requirements	4
4	Using AnalySize	5
4.1	Installation	5
4.2	The main window	5
4.3	Loading data	7
4.3.1	Censoring and removing data	9
4.4	Unmixing analyses	10
4.5	Saving and exporting plots and data	13
4.6	Saving and loading an AnalySize session	13
4.7	Other plots and analyses	14
4.7.1	Descriptive statistics	14
4.7.2	End member statistics	14
4.7.3	Multi-Specimen Spectra plot	15
4.7.4	Ternary plots	16
4.7.5	CM plot	17
4.8	Customizing plots	17

1 Introduction

Advances in instruments to measure particle size distribution, particularly laser diffraction particle size analyzers, has led to a increase in the number of data available and the resolution of grain size measurements. With this advance, new tools are required to processes these large data sets quickly and in a fashion that fits with the physical constraints of the data. To that end we have developed AnalySize, which is a **MATLAB** based software tool to aid Earth scientists make sense of their grain size data. AnalySize incorporates a range of powerful unmixing techniques alongside several routine methods for quantifying grain size data. AnalySize presents these tool in a clear, easy to use graphical interface that is compatible with a wide range of software platforms. AnalySize is a mathematical tool to help understand data set. Like all such tool and analyses, these mathematical procedures require physical ground-truthing with data and information outside of AnalySize. Users should strive to use physical constraints to find a balance between physical plausibility and numerical optimization.

If users encounter bugs or problems using AnalySize, or have any suggestions for improving the software, we would really like to hear from as all input is welcome. If you use AnalySize in your work, please make it clear what version was used and the what method of unmixing was used. We would also be grateful if you cited the publication that introduced AnalySize: Paterson, G. A., and D. Heslop (2015), New approaches and software for unmixing sediment grain size data, *Journal*, **vol.**, **pages**, doi: xxxxx/xxxxxx.

Greig A. Paterson (greig.paterson@mail.iggcas.ac.cn)

Institute of Geology & Geophysics, Chinese Academy of Science, Beijing, China.

David Heslop

Australian National University, Canberra, Australia.

August 24, 2015

2 Version history

Pre-v0.9.0r

Draft versions during development.

v0.9.0r – August 24, 2015

The version of AnalySize released during the peer review process.

3 System Requirements

AnalySize is written in **MATLAB** 2014a, and requires **MATLAB** v7, or higher, to run. AnalySize is self-contained and doesn't not require any additional **MATLAB** Toolboxes. AnalySize will run on Windows, OS X, and Linux platforms capable of running **MATLAB**. Processing time of the unmixing analyses are CPU dependent and will vary from system to system, but the basic requirements to operate AnalySize are the same as those required for **MATLAB**. The graphical elements of AnalySize require a minimum screen resolution of 1024×720. Users that experience any difficulties should contact the authors for assistance.

4 Using AnalySize

4.1 Installation

All of the files required for AnalySize are available from <https://github.com/greigpaterson/AnalySize>. The field can be downloaded individually or as a single zip file (<https://github.com/greigpaterson/AnalySize/archive/master.zip>). Once downloaded, the files should be put in an AnalySize folder and move to your desired destination (e.g., your MATLAB user directory).

From the “Home” tab in the main MATLAB window, select “Set Path”. This will bring up a window that lists all of the directories that are currently set in the MATLAB path. Click “Add Folder...” and navigate to the location of the AnalySize folder. Click on the folder name, then click “Open”. The AnalySize folder should now appear at the top of the directory list. Now click “Save” and then “Close”.

The AnalySize program can now be called from any directory by typing “AnalySize” in the MATLAB command window.

4.2 The main window

When AnalySize is launched the main window is opened (Figure 1). From here, all of the AnalySize functions can be accessed.

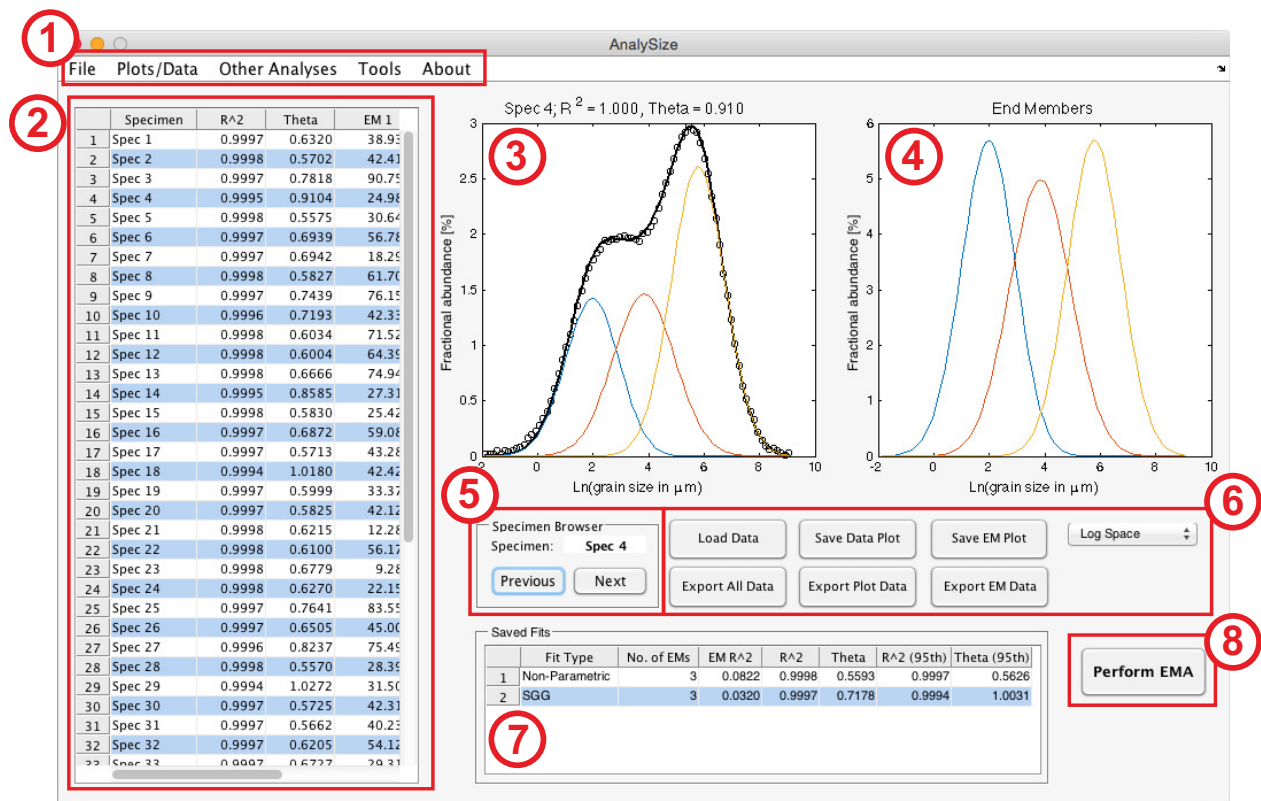


Figure 1: The main AnalySize window.

1. Menu bar

The main menu bar contains access to all the main function in AnalySize.

<i>File:</i>	Contains functions for loading data into AnalySize. It also includes functions for saving and loading an AnalySize session.
<i>Plots/Data:</i>	Contains functions for saving the main window plots and exporting data and end member fits.
<i>Other Analyses:</i>	Contains functions for performing other grain size analyses. This includes descriptive statistics, ternary diagrams, and CM diagrams.
<i>Tools:</i>	Contains functions for setting up and customizing AnalySize. This also contains function to censor grain size bins and remove specimens from the loaded data.

2. Specimen table

This table displays a list of specimen names from the loaded data. When end member fits have been made the table also includes the squared linear correlation (R^2) and angular distance in degrees (Theta) between the measured specimen and the fitted end member reconstruction. In addition, the percentage abundance of each end member is given. By clicking on individual specimen rows the user can jump between different specimens.

3. Data plot

This plots the individual specimen data (black circle) and end member fits that are made. The total end member fit is shown by the solid black line and the color order of the end members can be accessed from the “Tools” menu.

4. End member plot

This plots the unscaled end members of the currently selected fit (i.e., each end member sums to 100%). The color order of the end members can be accessed from the “Tools” menu.

5. Specimen browser

Controls to browse through specimens one-by-one.

6. Main control buttons

These buttons provide quick access to the main functions for loading and exporting data used to generate the plots as well as exporting the end member data. The pulldown menu allows the scale of the plots to be adjusted and can be set to a log space scale, log linear scale, or ϕ scale.

7. Saved fits

The saved fit table presents basic information about the end member fittings that have been performed on the currently selected data set. By selecting different rows of this table user can navigate between different fits that have been previous made.

<i>Fit Type:</i>	The type of analysis performed. Either "Non-parametric" EMA or parametric EMA (given as the distribution type).
<i>No. of EMS:</i>	The number of end members fitted.
<i>EM R^2:</i>	The maximum of the squared linear correlation between all fitted end members. This is a measure of the linear independence of the end members. If this value is too high then the data are likely over fitted and the number of end members should be reduced.
<i>R^2:</i>	The squared linear correlation between the measured data set and the data set constructed from the fitted end members.
<i>Theta:</i>	Angular distance (in degrees) between the measured data set and the data set constructed from the fitted end members.
<i>R^2 (95th):</i>	The lower 95 th percentile of the squared linear correlation between an individual specimen and it's reconstruction using the fitted end members.
<i>Theta (95th):</i>	The upper 95 th percentile of the angular distance (in degrees) between an individual specimen and it's reconstruction using the fitted end members.

8. Main analyses

The main button to call the End Member Analysis (EMA) routines.

4.3 Loading data

AnalySize supports a wide range of data formats, which includes both original laser particle size analyzer data files and post-processed text files and Excel spreadsheets. Data can be loaded by clicking the "Load Data" button on the main window, or by selecting "Load Data" from the "File" pulldown menu. Both of these will open up a file browser window, similar to that shown below (Figure 2), which allows the user to select their data files. Multiple files can be selected.

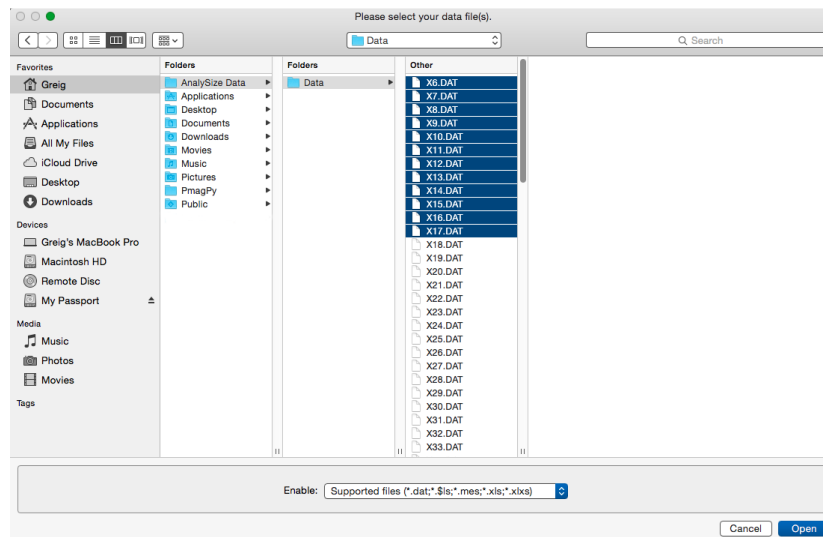


Figure 2: File browser window for loading data.

Useful Note...

OS X users should note that Beckman Coulter particle size analyzer files have the extension “*.sls”. This is not recognized by OS X as a valid extension and should be loaded by selecting the “All Files (*.*)” option from the file browser window.

Once the user has selected the file(s) to load one of two windows will open. For Excel files (*.xls, or *.xlsx) the window in Figure 3 will open. AnalySize supports different spreadsheet layouts, which can be specified from this window. Specimen data can be given as either rows or columns of the spreadsheets and key columns/rows can be specified to identify the data. This allows AnalySize to process spreadsheets that may contain additional meta-data not directly required for analysis (e.g., depths or ages). An example Excel file with both layout (using the default row/column values) is provided in the “*Example_Data*” directory of AnalySize.

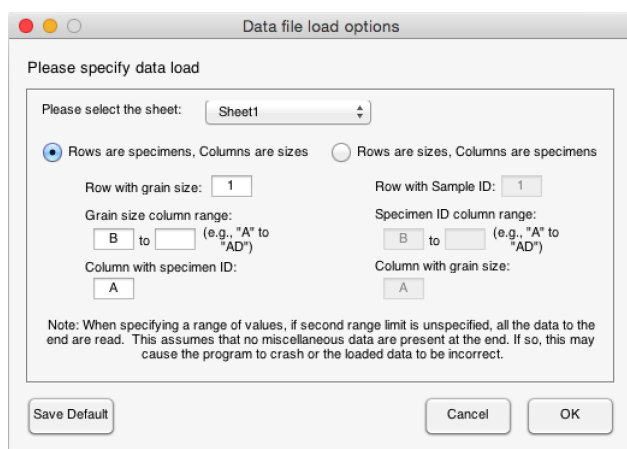


Figure 3: Load options window for Excel files.

For all other data files, AnalySize assumes that they are text files that are either delimited or direct outputs from particle size analyzer software. In this case the window in Figure 4 will open, which allows users to select the type of file they wish to load.

<i>Beckman Coulter</i>	Beckman Coulter *.sls data file.
<i>SALD-3001</i>	SALD-3001 data files.
<i>MicroTrac CSV</i>	MicroTrac comma separated variable data file.
<i>Cilas</i>	Cilas 940 data file.
<i>Delimited text file</i>	Delimited text file with columns of grain size bins and data and with a single header line. The delimiter can be selected from the pulldown menu (tab, comma, or space). AnalySize assumes that the first column contains the grain size bins and that proceeding columns are specimen frequency data. Data files can contain single specimen data or multiple specimens per file (“Multiple specimens” should be checked). For single specimens, the data file name without the extension is taken to be the specimen name. For multiple specimens, AnalySize assumes that the column headers are the specimen names. An example file is provided in the “ <i>Example_Data</i> ” directory of AnalySize.

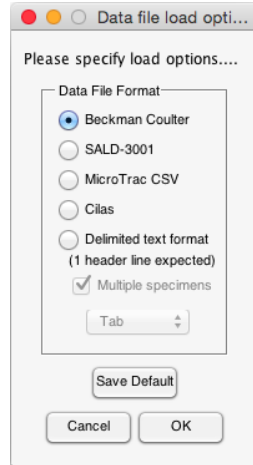


Figure 4: Load options window for data files.

Useful Note...

If a data file that you commonly are in a format that is not currently supported, please contact the authors with an example of your data format and it will be added to AnalySize.

For both of these load options windows, you can click “Save Default”, which will save the data file format choices to a user configuration file. When AnalySize is relaunched these default settings will be remembered.

Once the data are loaded, the specimen table and data plot will become populated (numbers 2 and 3, respectively in Figure 1). The user can browse through the data by selecting the corresponding specimen row in the specimen table, or by using the specimen browser panel (number 5 in Figure 1).

4.3.1 Censoring and removing data

From within the “Tools” pulldown menu, AnalySize contains functions to censor grain size bins and to remove individual specimens from the loaded data set. Choosing the “Censor Data” option will open the window shown in Figure 5. This window allows user to specify a grain size range to exclude from the data set. This function can be used to censor certain size bins that may be invalid (e.g., due to known measurement artifacts). Censoring the data will exclude these grain sizes from all loaded specimens. If multiple grain size ranges need to be censored the process can be repeated.

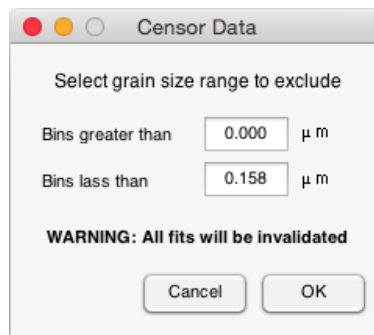


Figure 5: Censor grain size bins.

Choosing the “Remove Specimen” option will open a dialogue box confirming if the user wishes to remove the currently selected specimen. Only one specimen at a time can be removed, but the process can be repeated. When browsing through the loaded data the currently selected specimen is given in the specimen browser panel (number 5 in Figure 1).

Censoring data will invalidate all fits and all fits will be cleared from the AnalySize session. Removing a specimen will also effect EMA fits, but since the effects are often small, after removing a specimen, no fits will be cleared from the session. Before final interpretation, however, it is recommended that any fits are re-run after specimens have been removed.

4.4 Unmixing analyses

After data have been loaded into AnalySize EMA can be performed by selecting the “Perform EMA” button highlighted in panel 8 of Figure 1.

Choosing “Perform EMA” open up the window shown in Figure 6 and allows the user to undertake the EMA analyses as outlined by [Paterson and Heslop \(2015\)](#). The user has three end member options for EMA:

<i>Non-Parametric</i>	The form of the end members is determined from the data set. This routine uses the HALS-NMF algorithm of Chen and Guillaume (2012) .
<i>Parametric</i>	The form of the end members is based on parametric distributions (e.g., log-normal or Weibull). This routine uses a search algorithm to find the best fit distribution parameters. For each tested set of parameters the simplex projection algorithm of Heylen et al. (2011) to determine end member abundances.
<i>Defined</i>	The form of the end members is taken from a used defined input file containing the end members to unmix. This uses the fully constrained least squares algorithm of Heinz and Chang (2001) to determine end member abundances.

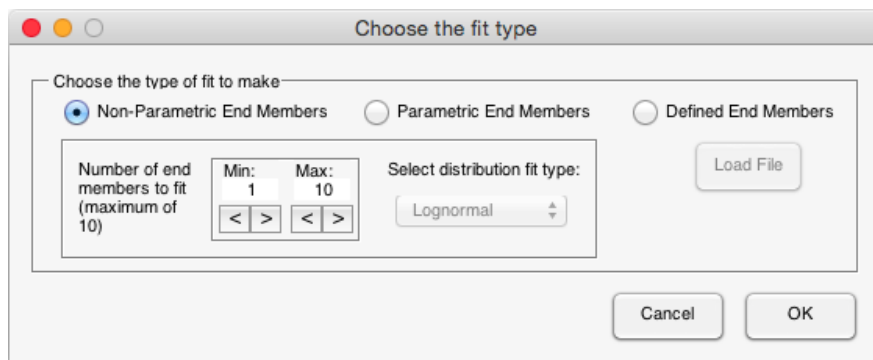


Figure 6: EMA fitting options.

For non-parametric and parametric end members, the user can set the minimum and maximum number of end members to test. The maximum number of end members support by AnalySize is 10, which more than is usually required to model a typical data set. For the parametric end members, users can choose from 4 parametric distributions given in Table 4.4.

Table 1: Parametric distributions supported in AnalySize.

Type	Parameters to fit	Comment
<i>Lognormal</i>	2	Parameters control the location and scale of the distribution. Lognormal distributions are fitted in linear space.
<i>Weibull</i>	2	Parameters control the shape and scale of the distribution. Weibull distributions are fitted in bin number space, where each size bin is number consecutively.
<i>Gen. Weibull</i>	3	General Weibull distribution. Includes an additional location parameter to the Weibull distribution. General Weibull distributions are fitted in bin number space, where each size bin is number consecutively.
<i>SGG</i>	3	Skewed Generalized Gaussian (SGG) distribution (Egli, 2003). The SGG is a 4 parameter distribution, but a maximum entropy approximation is used to reduce the number of free parameters to fit to 3 (Egli, 2004). When the parameters are saved or exported, all 4 parameter are output. Parameters control the location, scale, and skewness/kurtosis of the distribution. SGG distributions are fitted in log space.

When using defined end members, the user must first load an end member file by clicking on “Load File”. This will open a file browser where the user can select the desire end member file. The end member file should be a tab delimited text file with one header line. The first column must be the grain size bins in microns. The proceeding columns can be in any order and can contain additional data not needed for the unmixing (e.g., grain size bins in different units). The end member columns must be headed with “EMx”, where x is an integer number denoting the end member number.

Once the end members options have been setup, simply click on “OK” to perform the fitting. For defined end members, AnalySize will determine the unmixing proportions and return to main window where the results will be presented. The goodness-of-fit statistics will be shown in the fit table (part 7 in Figure 1). If the R^2 value is low and theta is high, this is an indication of a poor fit. This may be because the given end members do not fit the data well, or that additional end members are needed to model the data.

For non-parametric and parametric end members, AnalySize will determine the best end members and unmixing proportions for the range of end members given by the maximum and minimum number of end members to fit (Figure 6). The number of end members that best represents the data set must then be selected. To facilitate this, the window shown in Figure 7 is displayed after all the fits have been determined.

This window presents the fitting statistics for the different number of end members tested. The left hand plot displays various squared linear correlations as a function of the number of end members. The right hand plot displays angular differences (in degrees) between the reconstructed and observed data sets as a function of the number of end members. The lower panel displays some quantitative statistics for the selected number of end members, which can be controlled from the left hand side of this panel. These two plots can be saved to an EPS format file by clicking “Save Plot”. When the number of end members has been selected AnalySize will return to main window where the results will be presented.

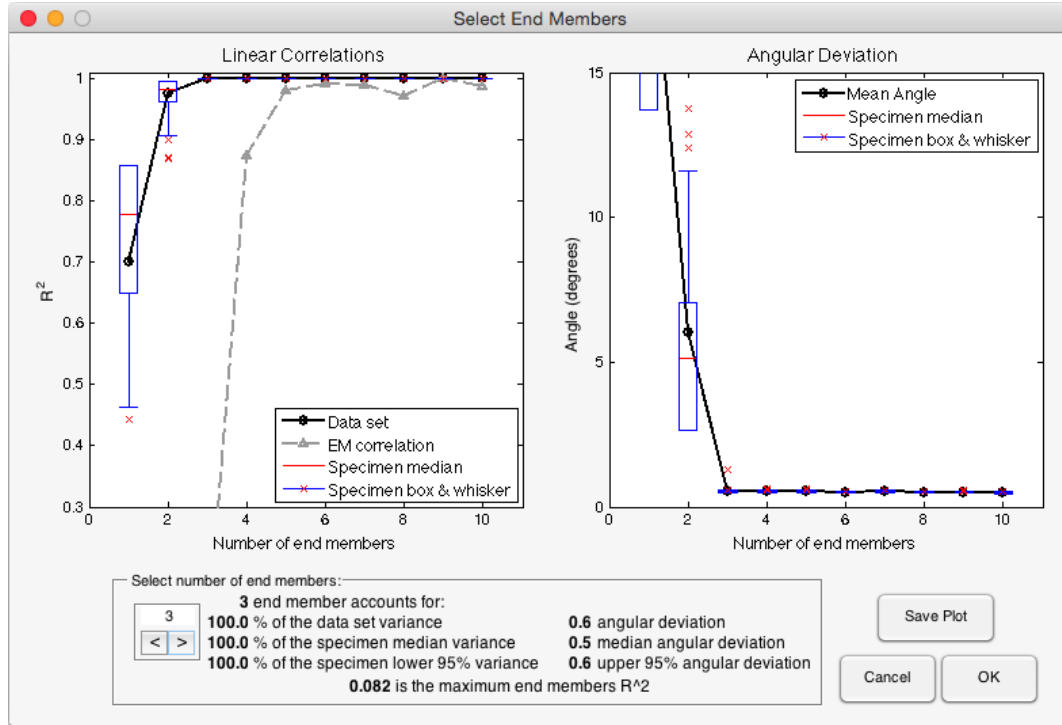


Figure 7: The end member selection window.

In both plots, the solid black lines and circles are determined for the whole data set, the blue box and whiskers represent the values determined from individual specimens. The red bars are the median values, the blue boxes are the interquartile range, and the blue whiskers mark out the one-sided 95th percentiles (i.e., the 95% coverage interval). For R^2 this is the lower 95th percentile and for angular deviation it is the upper 95th percentile - these represent a measure of the lowest quality fits. The red crosses represent outlying specimens (i.e., specimens that lie outside of the 95% coverage interval). The grey dashed line and triangles on the left had plot is the maximum squared linear correlation between the different fitted end members. This is a measure of the linear independence of the end members.

Useful Note...

A low linear correlation between end members is desirable, but there are physically reasonable situations where some correlation will occur (i.e., cases where end members contain a common source component). High end members correlations (i.e., $R^2 \gtrsim 0.7 - -0.8$), however, are a strong indication of overfitting where one end member is a near duplicate of another. As with all goodness-of-fit statistics, physical considerations of the data at hand are needed to make an informed decision.

If parametric unmixing is performed, AnalySize remembers the previous fits that were tested to generate Figure 7. If the parametric unmixing with the same distribution is repeated, AnalySize will ask if the existing tested fits should be reused (Figure 8a). The user can choose to use the existing ones or generate new ones. After the analysis, the user will then be asked if they wish to keep the existing fits, overwrite the existing fits with the new ones, or to merge the two sets (Figure 8b). This allows users to perform piecewise analyses, which can be useful for large data sets where parametric fitting can be slow.

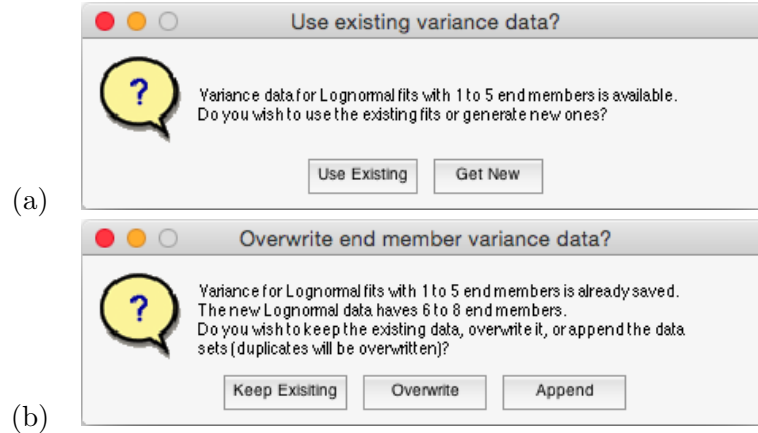


Figure 8: Prompt widows allow the user to reuse and save the parametric EMA fit data for different numbers of fitted end members.

After EMA fitting has been performed the fitted end members are plotted in the main window (panels 3 and 4 in Figure 1) and the fit statistics shown in the “Saved Fits” table (panel 7 in Figure 1). For non-parametric EMA, the end members are sorted by their median grain size. For parametric EMA, end members are sorted by their modal grain size.

4.5 Saving and exporting plots and data

All of the data and plots in the main AnalySize window (i.e., the measured grain size data and unmixed end members; Figure 1) can be exported through either the “Plots/Data” pulldown menu or the relevant button in the main control button panel (number 6 in Figure 1).

<i>Save Data Plot</i>	Save the currently selected specimen plot to a publication ready EPS file (number 3 in Figure 1).
<i>Save EM Plot</i>	Save the currently selected end member fit plot to a publication ready EPS file (number 4 in Figure 1).
<i>Export Plot Data</i>	Exports the data associated with the currently selected specimen (number 3 in Figure 1) to a tab delimited text file. This includes the measurement data, individual end members, and the total end member fit.
<i>Export EM Data</i>	Exports the data associated with the currently selected end member fit (number 4 in Figure 1) to tab delimited text files. This exports two files; a file containing the end member densities and a file containing the end member abundances and the goodness of fit statistics for all specimens.
<i>Export All Data</i>	Exports all of the measurement data into a tab delimited text file readable by AnalySize. The first column is the grain size in microns, which is followed by frequency data (sum to 100%) with each column representing a specimen.

4.6 Saving and loading an AnalySize session

For large data sets, some analyses can be time consuming and users may need to stop a fitting session or wish to save their fitting session for future work. To facilitate this, AnalySize includes a “Save Session” function (accessed from the “File” pulldown menu), which saves the current fitting

session to a MATLAB data file. Users can then load this data file at a later time (using the “Load Session” function in the “File” menu) and continue their analyses where they left off. This also allows user transfer their results and data to others user or computers and maintain the integrity of their analyses without having to repeat the time consuming steps.

4.7 Other plots and analyses

AnalySize contains other basic functions for processing and quantifying grain size data. These can be accessed from “Other Analyses” pulldown menu in the main AnalySize window.

4.7.1 Descriptive statistics

The Descriptive Statistics option generates tables of descriptive statistics and percentiles for the loaded data set, which can be readily exported to tab delimited text files. The available statistics include mean, standard deviation, skewness, and kurtosis determined by geometric and logarithmic methods of moments and geometric and logarithmic graphic analysis, as well as commonly used percentiles (*Krumbein and Pettijohn, 1938; Folk and Ward, 1957; Blott and Pye, 2001*).

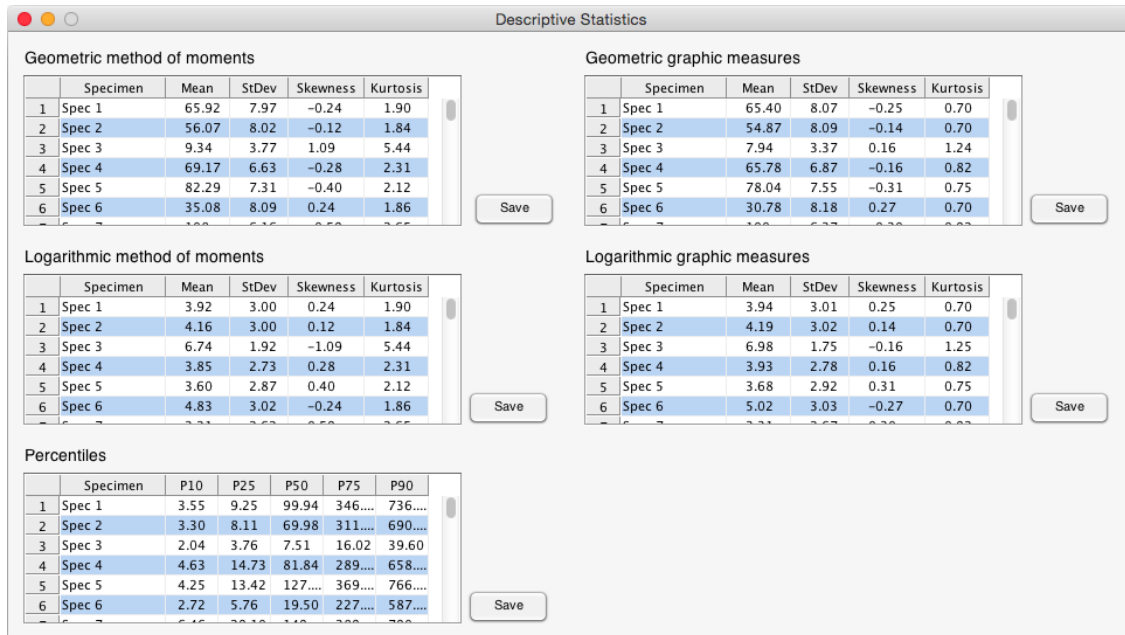


Figure 9: Descriptive statistics for the measured data.

4.7.2 End member statistics

The “End Member Stats” function generates tables of the descriptive statistics for the fitted end members, as described above. An example of the tables for EMA end members is given below in Figure 10.

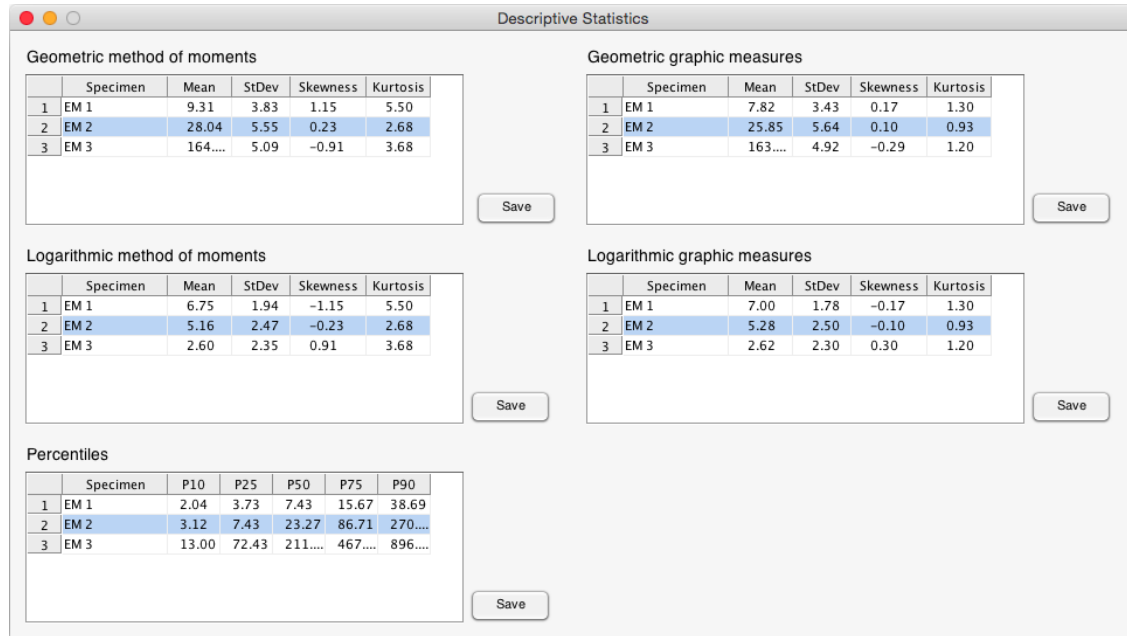


Figure 10: Descriptive statistics for EMA end members.

4.7.3 Multi-Specimen Spectra plot

The “Multi-Specimen Plot” option generates line plots for the grain size spectra for all loaded specimens. Users can select which specimens to plot using the checkboxes in the table provided and change the x-axis scale using the pulldown menu (Figure 11). The final plot can be exported to a publication ready EPS file by selecting “Save Plot”.

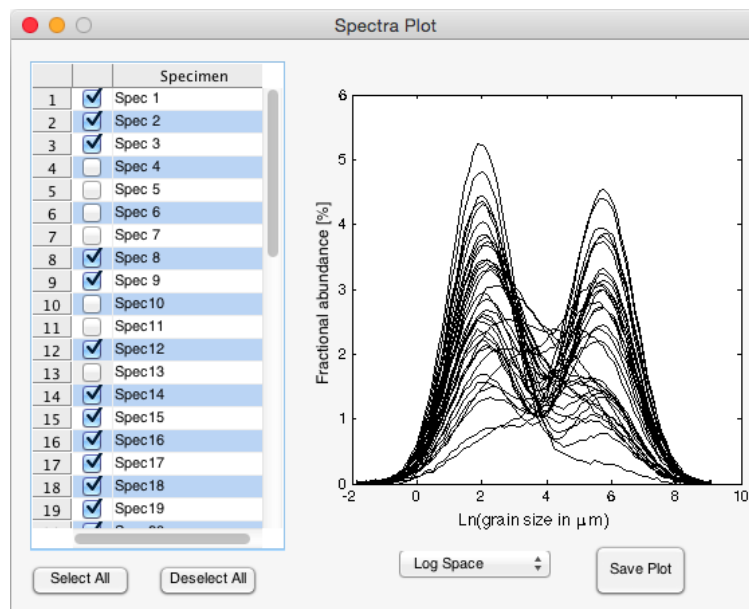


Figure 11: Multi-Specimen spectra plot.

4.7.4 Ternary plots

The “Ternary Plots” options plots the loaded data onto a range of ternary diagrams (Figure 12). AnalySize supports fine and coarse ternary plots of [Shepard \(1954\)](#), as modified by [Schlee \(1973\)](#), and [Folk \(1954\)](#). Clay, silt, sand, and gravel proportions follow the definitions of [Wentworth \(1922\)](#). On fine plots, AnalySize warns if significant gravel is present and excludes these specimens from the plot (e.g., lower left panel in Figure 12). Significant gravel is $> 0.01\%$ for Folk’s classification ([Folk, 1974](#)) and $> 10\%$ for Shepard’s ([Shepard, 1954](#)). All grain size proportions can be exported to a tab delimited text file by selecting “Export Data” and the diagrams can be saved to publication ready EPS files by selecting “Save Plot”. Plot symbols can be customized by selecting “Set Symbols” (described further in Section 4.8).

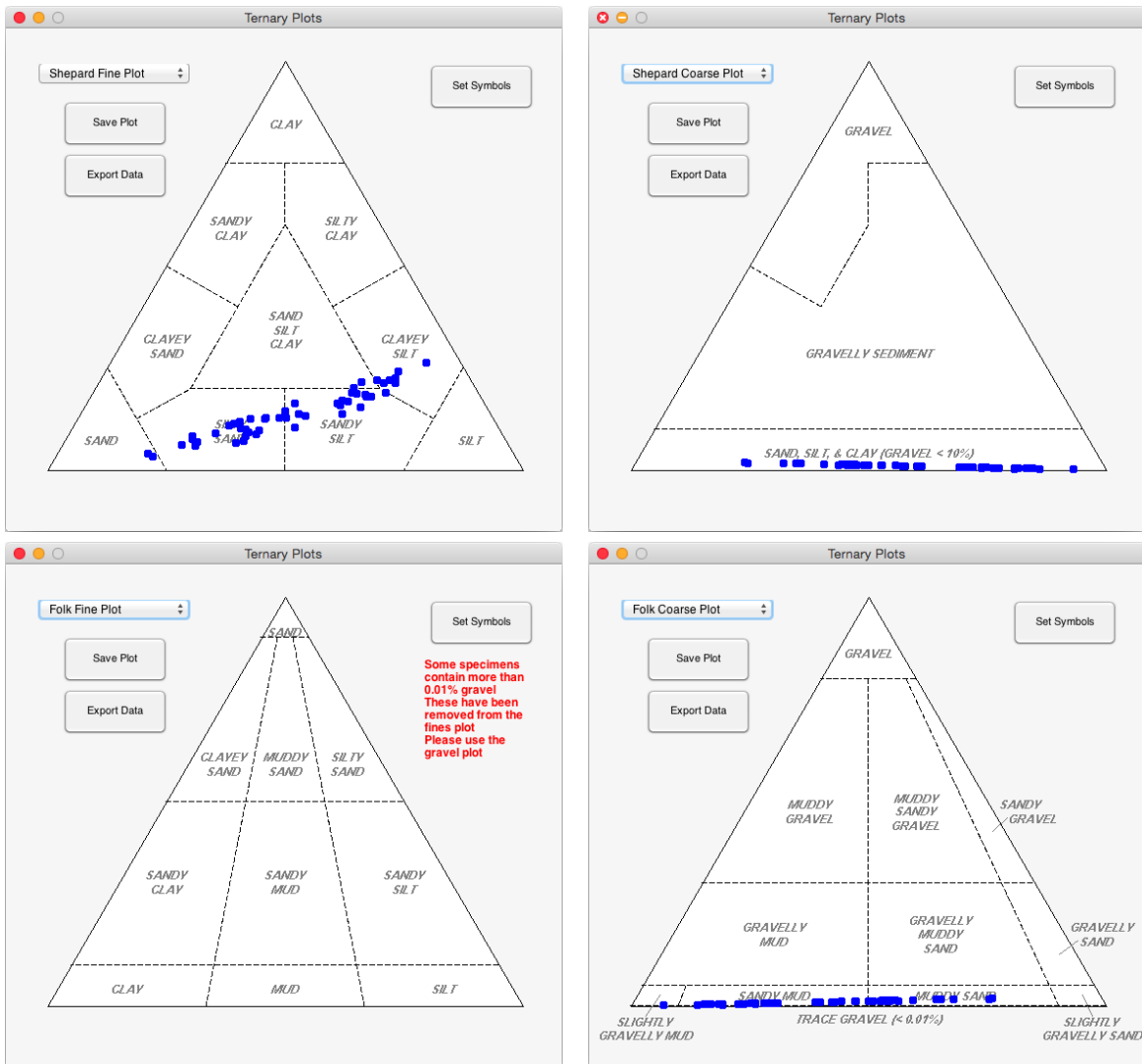


Figure 12: Ternary plots generated by AnalySize.

4.7.5 CM plot

The CM-plot of *Passega* (1964), which plots the first percentile (C) of grain size against the median grain size (M) can be accessed by selecting “CM Plot”. The data for this plot can be exported to text file and the plot can be saved to a publication ready EPS file. Plot symbols can be customized by selecting “Set Symbols” (described further in Section 4.8).

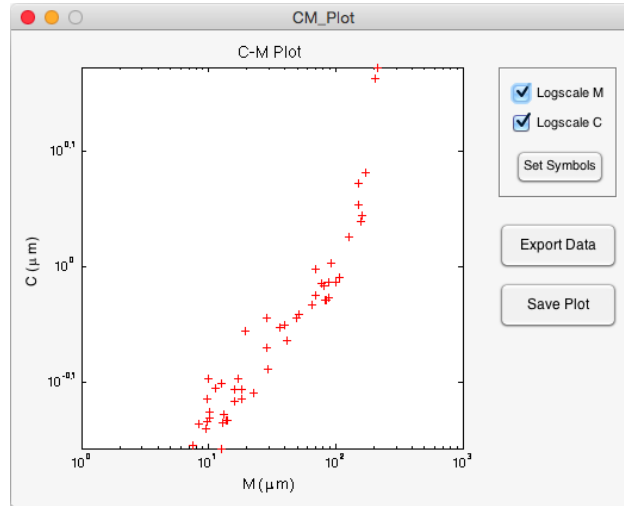


Figure 13: A CM plot from AnalySize.

4.8 Customizing plots

The “Tools” menu of the main AnalySize window (Figure 1) contains options to allow users to customize the plots generated by AnalySize. “Set End Member Colors” opens up the window shown in Figure 14, which allow the colors of the different end members to be modified. Clicking on each end member button opens MATLAB’s color palette tool, which provides a wide range of colors to choose from. Clicking “OK” will change the colors for the current AnalySize session, but these will revert to the standard default colors when AnalySize is closed. By selecting “Save Default”, AnalySize will save the choice of color to a user default configuration file, which allows AnalySize to recall user preferences each time it is run.

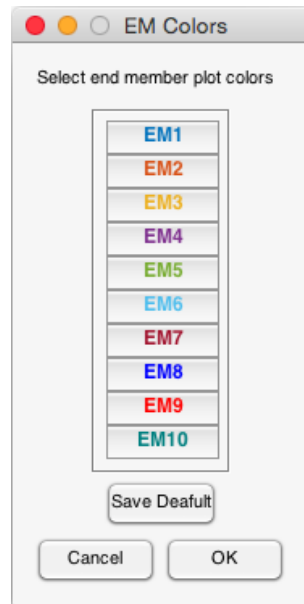


Figure 14: End member color options.

For ternary plots and CM plots the symbols are fully customizable and can be adjusted by selecting “Set Symbols” button in the relevant window (i.e., Figure 12 or Figure 13) or from the “Tools” menu in the main AnalySize window (Figure 1). The symbol style, size, and color can be adjusted. Clicking “OK” will change the symbols for the current AnalySize session only. “Save Default” will save the choice of symbol style and color for all future AnalySize sessions.

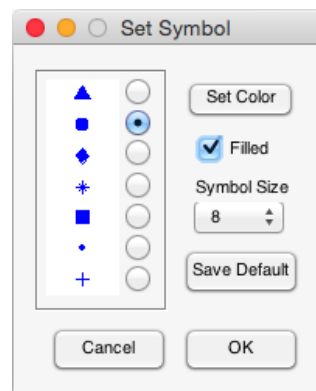


Figure 15: Ternary and CM plot symbol options.

References

- Blott, S. J., and K. Pye (2001), Gradistat: a grain size distribution and statistics package for the analysis of unconsolidated sediments, *Earth Surf. Proc. Land.*, *26*, 1237–1248, doi:10.1002/esp.261.
- Chen, W., and M. Guillaume (2012), Hals-based nmf with flexible constraints for hyperspectral unmixing, *EURASIP J. Adv. Sig. Pr.*, *2012*, 54, doi:10.1186/1687-6180-2012-54.
- Egli, R. (2003), Analysis of the field dependence of remanent magnetization curves, *J. Geophys. Res.*, *108*, 2081, doi:10.1029/2002JB002023.
- Egli, R. (2004), Characterization of individual rock magnetic components by analysis of remanence curves.: 2. fundamental properties of coercivity distributions, *Phys. Chem. Earth*, *29*, 851–867, doi:10.1016/j.pce.2004.04.001.
- Folk, R. L. (1954), The distinction between grain size and mineral composition in sedimentary-rock nomenclature, *J. Geol.*, *62*, 344–359, doi:10.2307/30065016.
- Folk, R. L. (1974), *Petrology of Sedimentary Rocks*, 182 pp., Hemphill Publishing Co., Austin, TX.
- Folk, R. L., and W. C. Ward (1957), Brazos river bar [texas]; a study in the significance of grain size parameters, *J. Sed. Res.*, *27*, 3–26.
- Heinz, D. C., and C.-I. Chang (2001), Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery, *IEEE Trans. Geosci. Remote Sensing*, *39*, 529–545, doi:10.1109/36.911111.
- Heylen, R., D. Burazerovic, and P. Scheunders (2011), Fully constrained least squares spectral unmixing by simplex projection, *IEEE Trans. Geosci. Remote Sensing*, *49*, 4112–4122, doi:10.1109/TGRS.2011.2155070.
- Krumbein, W. C., and F. J. Pettijohn (1938), *Manual of Sedimentary Petrography*, Appleton-Century-Crofts, New York.
- Passaga, R. (1964), Grain size representation by cm patterns as a geologic tool, *J. Sed. Res.*, *34*, 830–847.
- Paterson, G. A., and D. Heslop (2015), New approaches and software for unmixing sediment grain size data.
- Schlee, J. S. (1973), Atlantic continental shelf and slope of the united states - sediment texture of the northeastern part, *U.S. Geol. Surv. Prof. Pap.*, *529-L*, 64 pp.
- Shepard, F. P. (1954), Nomenclature based on sand-silt-clay ratios, *J. Sed. Res.*, *24*, 151–158, doi:10.1306/d4269774-2b26-11d7-8648000102c1865d.
- Wentworth, C. K. (1922), A scale of grade and class terms for clastic sediments, *J. Geol.*, *30*, 377–392, doi:10.2307/30063207.