EXCITATION-EMISSION-MATRIX'S CONVERSION, CORRECTION, COGNITION, COMPARISON, AND CALCULATION (EFC) V1.0

USER GUIDE

Dr. Wei He[¶,¶]

Prof. Fuliu Xu[¶]

[¶]School of Water Resources and Environment, China University of Geosciences (Beijing)

[¶]College of Urban and Environmental Science, Peking University

CONTENTS

1. Overview of the software

The efc is a software platform used to process three-dimensional fluorescence spectrum data (excitation-emission-matrix, EEM). The software can process absorption spectrum (abs.) as well because abs. is typically employed when correcting the EEM. The efc includes five functions: (1) data-conversion from the measured EEM or water Raman spectrum by instrument to required dataset which is needed in the subsequent analysis; (2) data-correction by considering inner filter effect, blank subtraction, and quinoline sulfate calibration; (3) component-cognition of EEM by PARAFAC analysis; (4) model-comparison with other PARAFAC models through quantifying their similarity and fitting the EEMs with previous PARAFAC models; and (5) indicator-calculation from EEM and abs. spectra and score-calculation of PARAFAC components.

The software is developed using MATLAB GUI base on the correction procedure of EEM, calculation of fluorescence indicators, and parallel factor (PARAFAC) theory. It is a very friendly toolbox for researchers who are not familiar with PARAFAC analysis with MATLAB. The application scope of this software is not limited to the study of EEM of natural organic matter. We can expand their application in EEM of other substances. Considering that the software's five important functions: conversion, correction, cognition, comparison, and calculation of EEMs mentioned above, we name this as EEMs' five c, shortly efc.

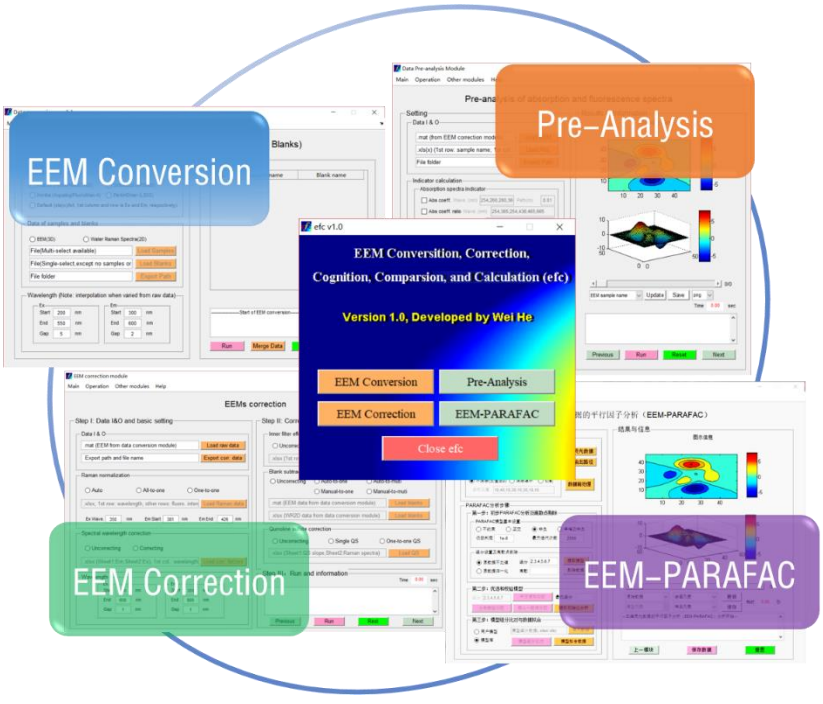


Figure 1 The main functions of the efc software

The main four functions of efc software are shown in Figure 1, including EEM conversion module, EEM correction module, pre-analysis of EEM and abs. module, and EEM-PARAFAC module. The former two modules are used to pretreat the data, the latter two modules are used to calculate the indicators and to conduct parallel factor (PARAFAC) analysis.

2. The software development environment

The software's interface is developed by MATLAB 2011b GUI. The main code is in MATLAB language. In addition to FDOMcorrect, n-way toolboxes, all other codes are written after learning FLtoolbox and DOMFluor toolbox. The software has 64-bit exe version and p-code version. There are some example data in the software for beginners' early learning.

3. The introduction of the main interfaces

3.1 Interface of main function list

The interface of main function list (Figure 2) includes four function buttons, which link to "Interface of data conversion module", "Interface of EEM correction module", "Interface of data pre-analysis module", and "Interface of EEM-PARAFAC module", respectively. The "interface of main function list" can be turned off by clicking "close efc" button or cross button.

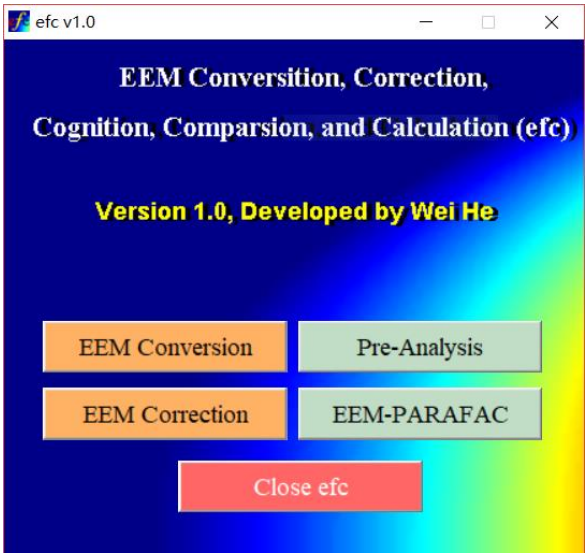


Figure 2 The listed modules in the efc main interface

The interface switch between the interface of main function list and other

interfaces is shown in Figure 3. The running sequence of those modules generally follow the red arrows, i.e., data conversion is ran firstly, and the second step is EEM correction, the following data pre-analysis and EEM-PARAFAC analysis can be carried out simultaneously.

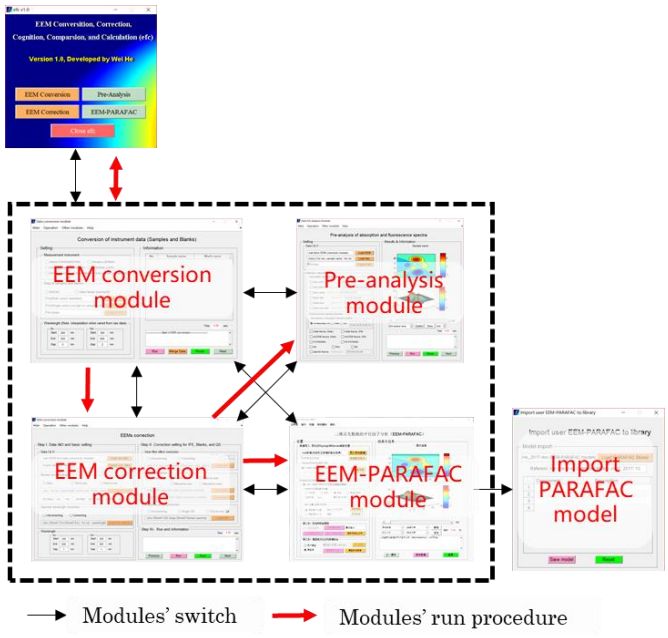


Figure 3 The relationship among each module in efc software

3.2 Interface of data conversion module

The interface of data conversion module includes two panels of "Settings" and "Information". The "Settings" panel can realize selecting the measurement instrument, importing the measured data (samples and blanks), selecting the export path, and setting up excitation and emission wavelengths' start, end, and the gap (interval).

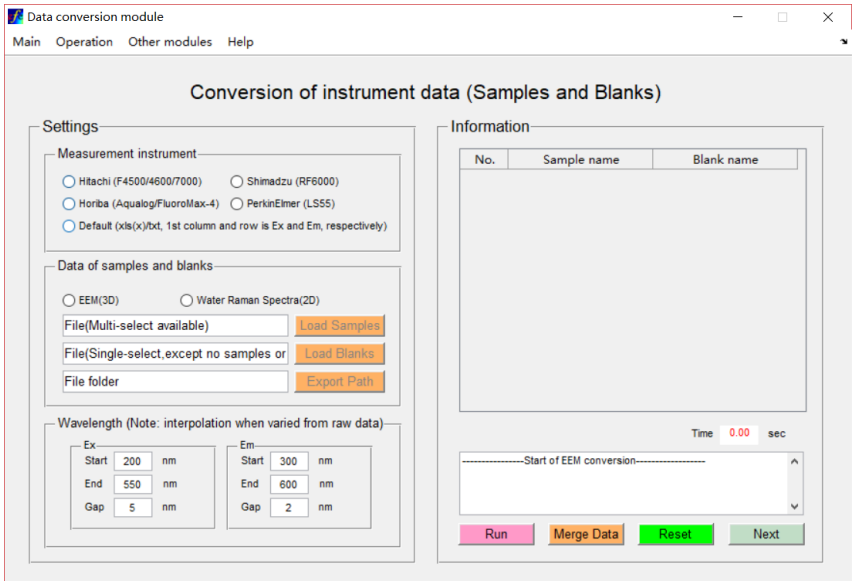
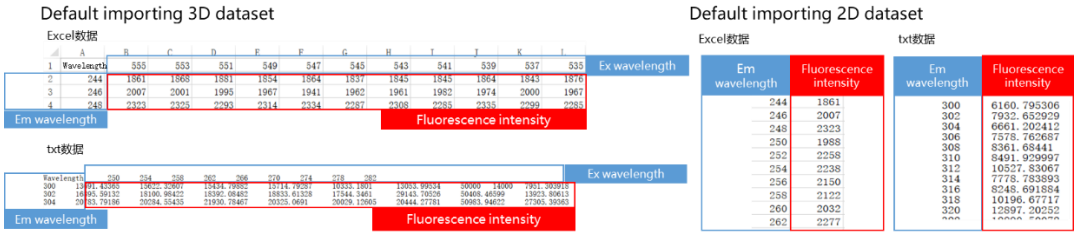


Figure 4 Interface of data conversion module

The previous literature investigation shows that the mainstream fluorescence spectrophotometers used by researchers include Hitachi F series, Shimadzu RF series, Horiba Aqualog/FluoroMax, Perkin-Elmer LS series, and Agilent/Varian Cary Eclipse. After practically investigating the usage of those instrument in Peking University, which is one of the institutes which had the most scientific instruments in China, we consider raw data produced by four manufacturers, Hitachi, Shimadzu, Horiba, and Perkin-Elmer. The measured EEM (3D) and Raman scan (2D) data, which is determined for the later Raman normalization, are the imported data in this module (Figure 4).

For data produced by other instruments, we also provide the default option as follows, which has similar data organization style with the data style produced by Horiba. The imported 3D data in default option is required that 1st row, 1st column, and other parts are excitation wavelengths, emission wavelength, and fluorescence intensity, respectively. The imported 2D data in default option is required that 1st row is emission wavelengths and other rows are fluorescence intensity. The default excitation wavelengths of Raman scan is 350 nm in efc software.



The software can read the names of imported data and summarized them in a separate files. That file will be used in latter analysis. Therefore, it is suggested that users should better give a clear name of each data.

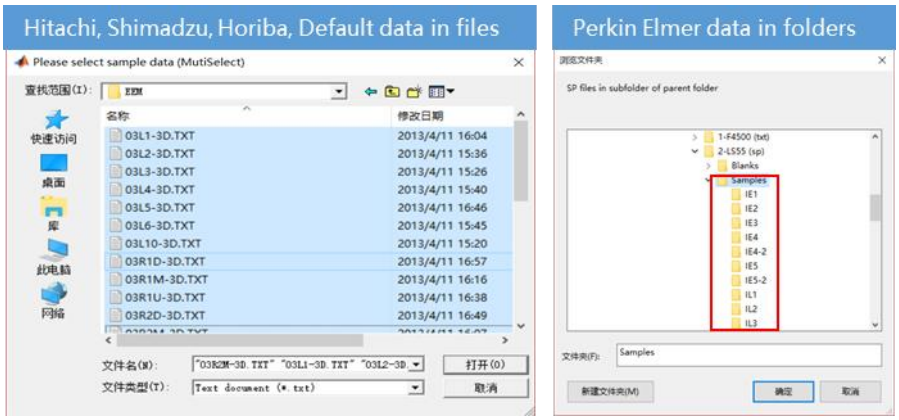


Figure 5 Importing of sample data

The data import button is not available before the type (EEM (3D) or Water Raman spectra (2D)) of imported data is selected. The imported data function support multiple-selection. The imported data file format support Excel (xlsx, xls, and csv

formats) and text document (txt format), which are common data format for measured data from instruments. For Perkin-Elmer LS series, the imported data is in file folders, not in specific files because the produced 3D data is separately saved in a certain number of 2D SP files (Figure 5). Thus, the folder, including sample folders, which contain SP files, is simply selected when importing Perkin-Elmer LS series data.

The water blank is generally measured accompanying with samples to eliminate the disturbance from the instrument' background and samples' matrix effects. Three ways of measurement based on instrument steady degree and experimental requirement are employed: (1) one sample corresponds to one blank; (2) some samples corresponds to one blank, others corresponds to another blank; (3) all samples corresponds to only one blank. When importing the blank data measured by first way, the names of imported blanks should contain the names of samples, i.e., Sample01blk or BLK_sample01. In other word, a blank prefix (BLK_) or suffix (_blk) should be added based on samples' names, otherwise, the blanks' sequences might be different from samples', which causes errors in latter analysis. When importing the blank data measured by third way, it is simple to just import the blanks using "Load Blanks" button. When importing the blank data measured by second way, it is suggested that parts of samples and their corresponding blank are imported by the third way and then several exported datasets are combined using the "Merge data" button in the "Information" panel. The software can convert the sample or blank data separately. When the sample or blank data is imported separately, the result dataset will only include the sample or blank data.

Setting requirements for wavelength ranges and intervals is that: (1) the start and end wavelengths needs to be within the wavelength ranges of imported data, and the software cannot achieve extrapolation; (2) the interval should not be too small, otherwise the interpolation data may be distorted, and the wavelength interval is generally not greater than 5 nm and not less than 0.1 nm, otherwise, the software cannot work properly. For EEM data, start, end, and interval of both Ex and Em need to be set; for Raman spectrum data, start, end, and interval of Em is only required to be set and Ex will be a single wavelength scalar. The interpolated data might be produced if the set interval is not in accordance with that of the original instrument data. It is suggest using the original wavelength interval for the measurement data, which has equal interval. For the measurement data without equal interval such as data from Horiba Aqualog whose Em wavelength interval is not equal, it is suggested setting the wavelength interval to a fixed interval, which will facilitate later data analysis. Reminder: appropriate setting of wavelength range and interval not only facilitate later indicators' calculation, but also increase the running efficiency of PARAFAC analysis.

The "Information" panel mainly provides the display of data and the operation

button of the module. The top part of the panel shows the number of the imported data, the sample name, and the corresponding blank name. If only samples or only blanks are imported, only samples' or blanks' names are shown. The runtime is also shown in this panel. Some important tips for operating this interface and reminders are shown in the prompt box at the bottom of the panel. The data conversion is processed by "Run" button. The EEM or Raman data of the sub-dataset are merged by "Merge data" button. All data I & O information on the interface will be reset by "Reset" button. The "Interface of EEM correction module" can be reached by clicking "Next" button.

The operation on this interface can also be implemented using the menu in the upper left corner of the interface, and each operation corresponds to the shortcut key. The exported result data files as shown in Figure 6. The name of converted EEM (3D) data has the prefix of EEM_dataset_ and the suffix of current time. The .mat format data includes 3D datasets (EEM_samples and EEM_blanks), number of samples (nSample), excitation wavelengths (EX), emission wavelengths (EM), sample name list (Data_name). In the folder, namely EEM_dataset_time, the data in mat format files is exported in the Excel (,xlsx) format files, which facilitate accessing those data for users who do not have MATLAB software. The name of converted Raman spcectra (2D) data has the prefix of WR2D_dataset_ and the suffix of current time. The .mat format data includes 2D datasets (WR2D_samples and WR2D_blanks), number of samples (nSample), specific excitation wavelength (Ex_set), emission wavelengths (EM), sample name list (Data_name). The data in mat format files is exported in the Excel file (.xlsx) WR2D_dataset_time.xlsx, which facilitate accessing those data for users who do not have MATLAB software. The mat format data will be used in later interface of EEM correction module.

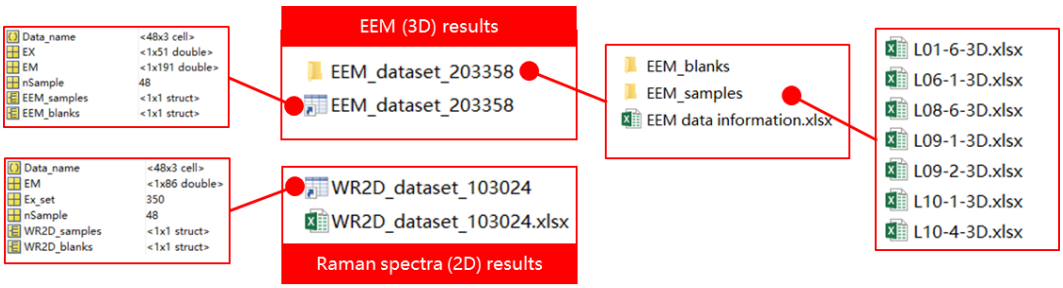


Figure 6 The results of instrument data conversion module

3.3 Interface of EEMs correction module

Three step-panels were included in the interface of EEM correction module as follows (Figure 7): (1) Step I Data I&O and basic settings, (2) Step II Correction settings for IFE, Blanks, and QS, and (3) Step III Run and information.

In Step I panel, the uncorrected data and path and name of exported file is required to be set first. Then the imported data and wavelengths should be set in

Raman normalization panel. Generally, "Auto" and "One-to-one" is chosen to use Raman spectra extracted from EEMs and Raman spectra measured by instrument, respectively. The "All-to-one" is chosen when the varieties among samples are very small and could share the same Raman spectrum. Then the spectral wavelength correction is require to be set: correcting or uncorrecting. The correction factors should be imported when "Correcting" is chosen. The settings of wavelengths is similar to that in interface of data conversion module. The interval should be proper and the ranges should be within the ranges of exported data. In practical, the software will show the wavelength ranges and intervals after EEMs is imported. The users do not need to change wavelengths unless they have their own purpose.

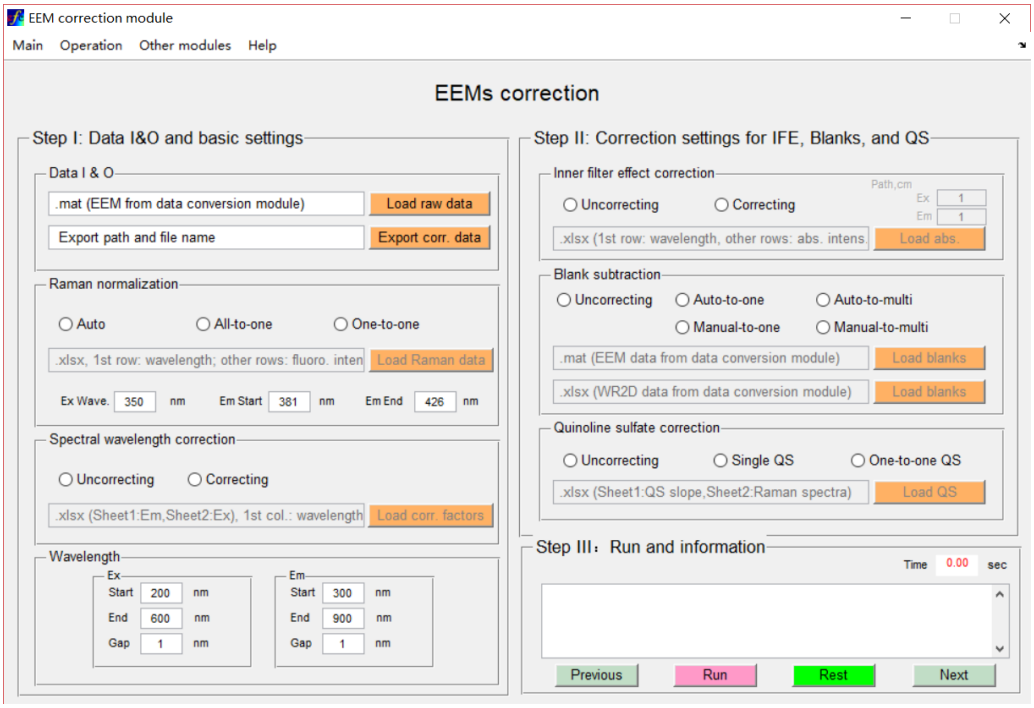


Figure 7 Interface of EEMs correction module

In the inner filter correction panel of Step II panel, the excitation and emission optical paths can be set when correcting. Generally, default values are assigned when the optical cuvette has a 1 cm * 1 cm cross section. In blank subtraction panel, "Uncorrecting", "Auto-to-one", "Manual-to-one", "Auto-to-multi", and "Manual-to-multi" separately mean that no blank subtraction data, one blank and automatically extracted Raman spectrum form that blank, one blank and one measured Raman spectrum form that blank, sample-corresponding blanks and automatically extracted Raman spectra form those blank, and sample-corresponding blanks and measured sample-corresponding Raman spectra are used to conduct blank subtraction, respectively. In the panel of quinoline sulfate (QS) correction, "Uncorrecting", "Single QS", and "One-to-one QS" mean that no QS slope, one QS slope, and sample-responding QS slopes are used to conduct QS correction.

In the Step III panel, the runtime is also shown in this panel. Some important tips for operating this interface and reminders are shown in the prompt box at the bottom of the panel. The EEM correction is processed by "Run" button. All the data I & O information on the interface will be reset by "Reset" button. The "Interface of data pre-analysis module" can be reached by clicking "Next" button. And the "Interface of data conversion module" can be reached by clicking "Previous" button.

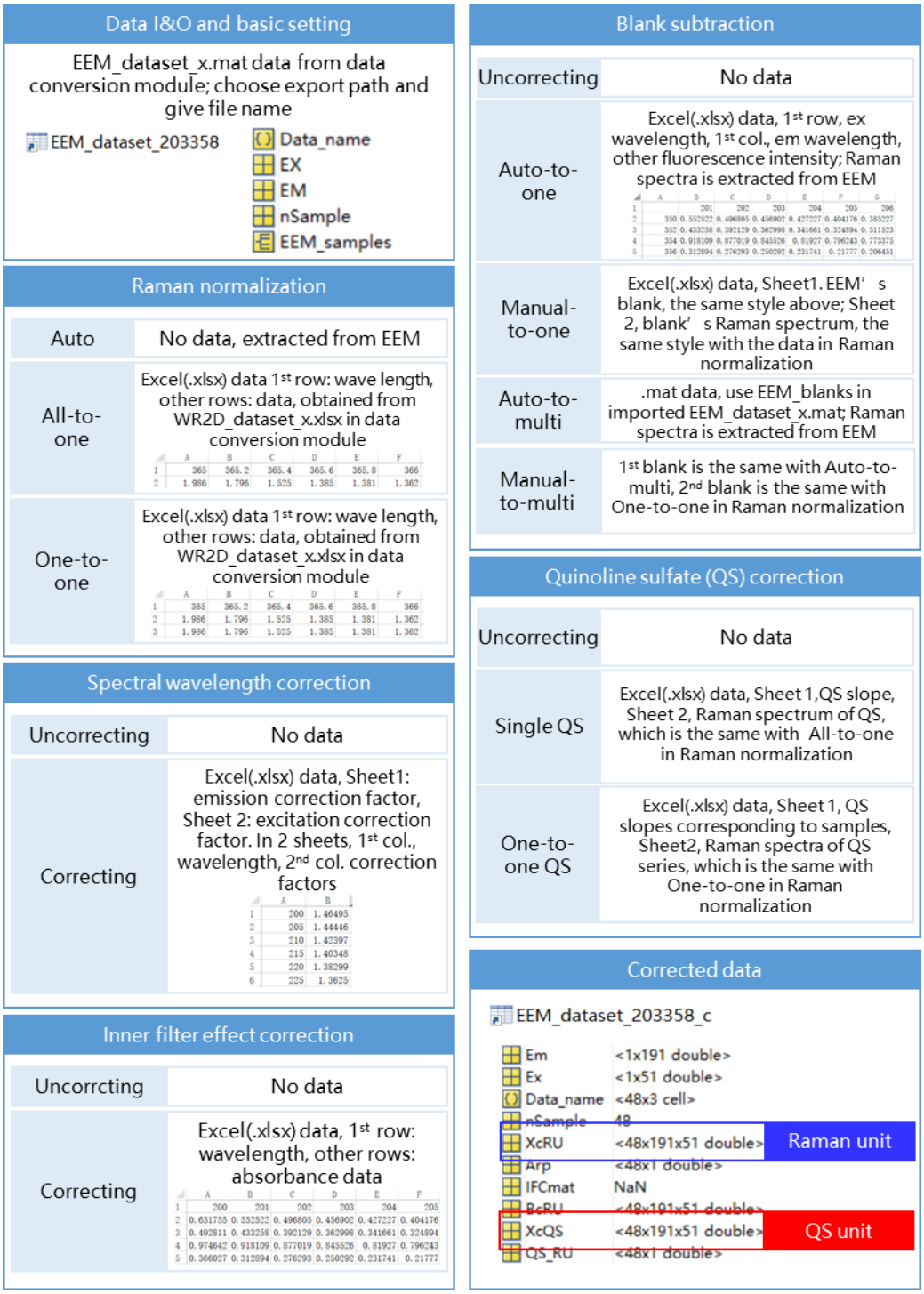


Figure 8 The formats and styles of imported data before EEM correction

Many selection items are included in this interface and different data formats and styles are required. Therefore, the Figure 8 is used to illustrate the required formats and styles of imported data when different options are chosen. In addition, the exported data from interface of data conversion will be used in the present interface. The sample (EEM_samples) and blank (EEM_blanks) EEMs are in the mat datset, namely EEM_dataset_time.mat. If there are no blanks or only one blank, we will conduct no blank subtraction or "Auto/Manual-to-one" blank subtraction, repectively. And Excel dataset will be imported. The "Auto" in Raman normalization means the Raman spectra are automatically extracted from sample EEMs. And the excitation wavelength setting and emission wavelength ranges should be within those in sample EEMs. The "Auto-to-one/multi" items in blank subtraction also means the Raman spectra are automatically extracted from blank EEMs. When conducting QS correction, some samples have one QS slope, others have another QS slope. In that situation, we still choose "One-to-one QS" item. It is suggested that the number and sequence of slope values should be in correspondence to samples' sequence and number. The samples which have the same QS slope will have the same slope value in that Excel dataset. The exported data, including excitation wavelengths (Ex), emission wavelengths (Em), sample names (Data_name), Raman normalized EEMs (XcRU), QS calibrated EEMs (XcQS), and other procedure parameters, are also in mat format, which will be used in last two interfaces. The user-defined mat dataset should have the same names and organized structure as shown in Figure 8, otherwise it cannot be used in last two interfaces.

The operation on this interface can also be implemented using the menu in the upper left corner of the interface, and each operation corresponds to the shortcut key.

3.4 Interface of data pre-analysis module

The following two interfaces are used for data analysis. The interface of EEM pre-analysis is mainly used in the calculation of absorption and fluorescence indicators. In addition, the contour and surface plots of fluorescent spectra can be displayed in interface and saved in image format.

This interface is divided into "Settings" and "Results and information". In the "Settings" panel, the import and export of the data and the selection of the calculation indicators should be conducted. The imported EEM data is mat format and from the previous interface. When importing data, it will be asked to choose "RU data" or "QS data" as shown in Figure 9. The input absorbance spectrum data is in Excel (xls(x)) format. The 1st row and column should be wavelengths and sample names, respectively. The data output path should also be set, otherwise the reminder dialog will appear.

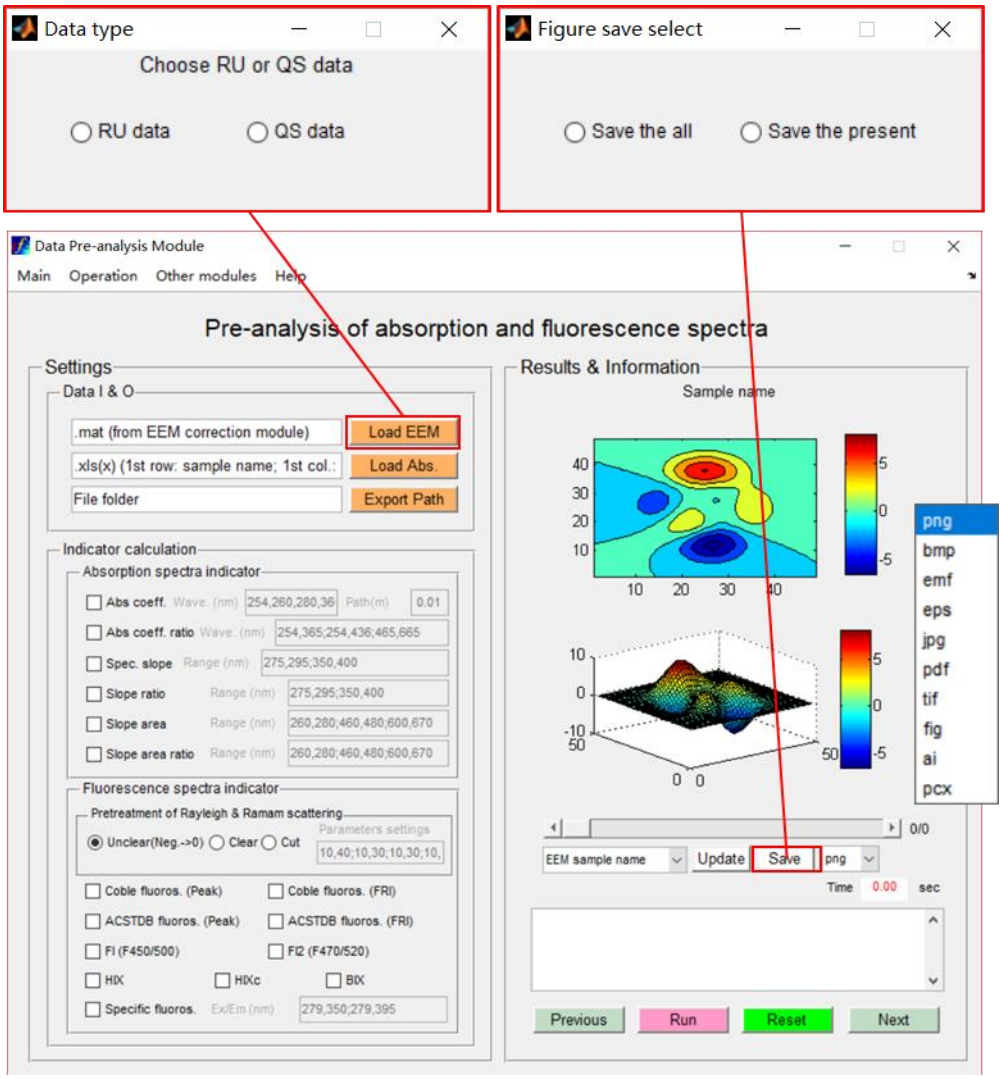


Figure 9 Interface of data pre-analysis module

In the absorption spectra indicator panel, the hidden parameter settings will show up when check box is selected and the input of wavelength should meet the requirements: each wavelength data is separated by commas and cannot have spaces and there is no commas or space behind the last data for calculation of absorption coefficient (absorptivity); for other indicators, a pair of wavelengths are separated by commas and cannot have spaces and two pairs are separated by semicolon and cannot have spaces, and there is no semicolon or space behind the last data. In addition, the spectral slope ratio and the spectral area ratio should have at least two pairs of data input, and the ratio results will be the ratio of any two pairs when the input data contain at least three pairs.

In the fluorescence spectra indicator panel, the Raman and Rayleigh scattering peaks can be eliminated or cut. When "Clear" item is chosen, the removed scattering peak data area will be interpolated to supplement the missing data. When "Cut" item is chosen, the removed scattering peak data area will be assigned to 0. The eliminating

parameters have a style of "x1,y1, x2,y2,x3,y3,x4,y4". The xi and yi forms a pair of parameters to control the elimination area of scattering peaks. The four pairs of parameters control the 1st Rayleigh scattering, 1st Raman scattering, 2nd Rayleigh scattering, and the 2nd Raman scattering, respectively. The parameters x and y control the upper and lower boundary of elimination, respectively. After adjusting the parameters, the best eliminated EEMs will be obtained, which facilitate the indicator calculating and EEM PARAFAC analysis. The input pair data have the same requirement and expressed in detail above.

In the upper part of interface, the EEMs' contour and surface figures can be shown. Figures of different samples can be shown by moving the slider or selecting name in the name list. The "Save" button is used to save all figures or just present figures of EEMs' in 10 image formats (Figure 9). In addition to the graphical data, other results data, including scattering peaks removed EEMs, fluorescence indicators, and absorption indicators, can also be exported as shown in Figure 10. Those data have prefix of EEM_RRR, EEM_Indicators, and UV_Indicators, respectively. And their suffix is the current time.

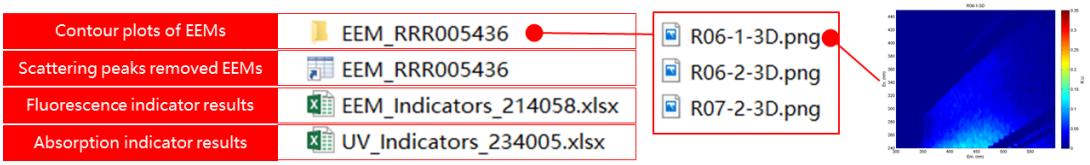


Figure 10 Result files and folders of data pre-analysis module

The runtime is also shown in this panel. Some important tips for operating this interface and reminders are shown in the prompt box at the bottom of the panel. The EEM correction is processed by "Run" button. All the data I & O information on the interface will be reset by "Reset" button. The "Interface of EEM-PARAFAC module" can be reached by clicking "Next" button. And the "Interface of EEM correction module" can be reached by clicking "Previous" button. The operation on this interface can also be implemented using the menu in the upper left corner of the interface, and each operation corresponds to the shortcut key.

3.5 Interface of EEM-PARAFAC module

The interface of EEM-PARAFAC module is the same with interface of EEM pre-analysis module and has two parts, "Settings" and "Result and information" (Figure 11).

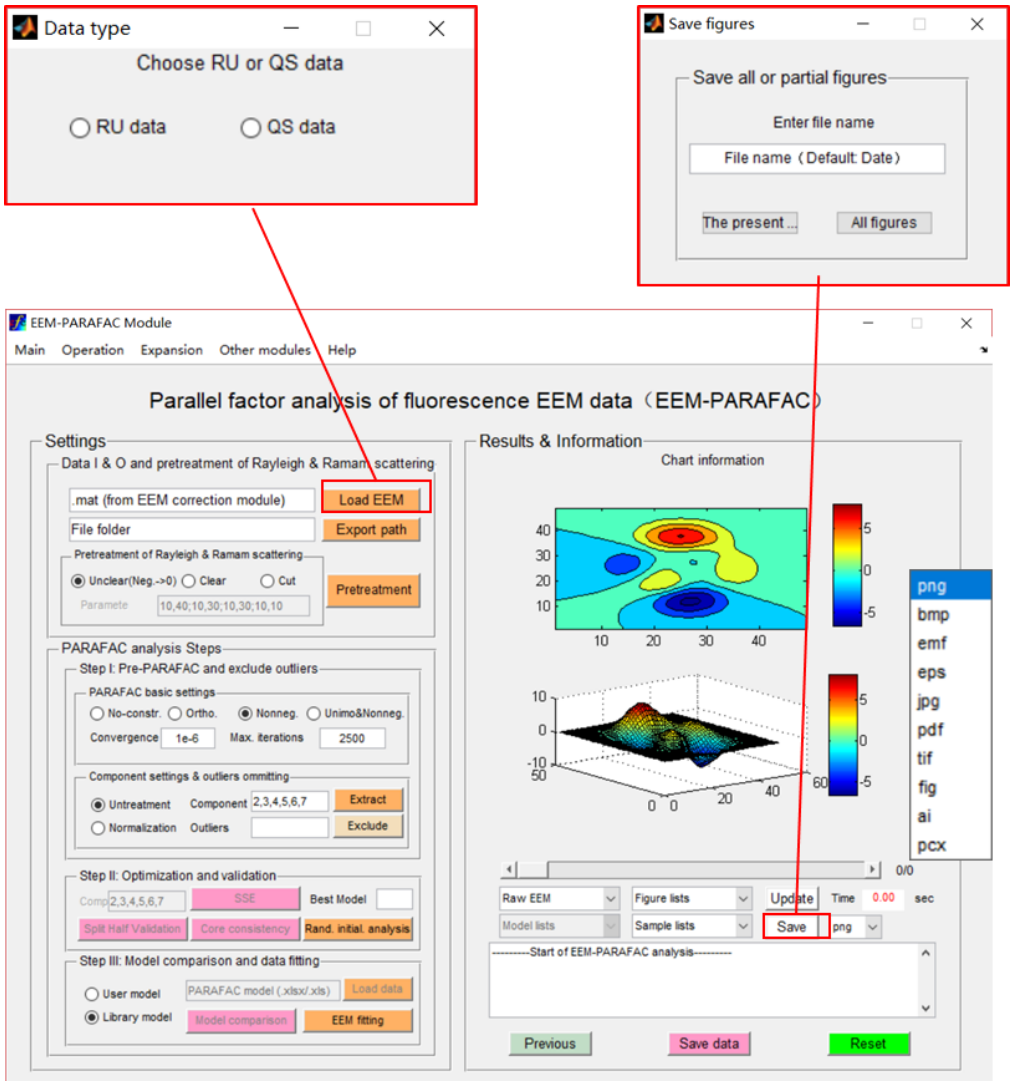


Figure 11 Interface of EEM-PARAFAC module

In the "Setting" panel, the top panel, including function of data import, data export, and elimination/cutting of scattering peaks, has the similar function to the previous interface. Then, the PARAFAC analysis is carried out in three steps:

(1) The preliminary PARAFAC analysis is first conducted, and then the second PARAFAC analysis is conducted after omitting the outliers judged by the leverage of samples and EEMs' contour figures. The panel of "PARAFAC basic settings" mainly set the constraint value, the convergence diagnostic value, and the maximum number of iterations. The panel of "Component settings & outliers omitting" can realize the normalization of EEMs and settings of models' components and series numbers of outliers. The models' components and series numbers of outliers should be separated by commas without spaces. The preliminary PARAFAC analysis will be conducted after clicking the button "Extract", the second PARAFAC analysis is conducted after omitting the outliers will be conducted after clicking the button "Exclude". When no outliers are input, the results of the "Exclude" is from the results of the "Extract".

(2) In the second step of "Optimization and validation", the best PARAFAC model is obtained through sum of squared errors (SSE), core consistency, split half validation, and finally random initialization analysis. The selected models in panel of "Optimization and validation" should be within the PARAFAC models constructed in the 1st step. The "SSE", "Core consistency", and "Split half validation" function will be disable unless the preliminary PARAFAC analysis is conducted in the 1st step. After input the optimized models' component number (a single value), the random initialization analysis can be sequently conducted.

(3) In the third step of "Model comparison and data fitting", the constructed model will be quantitively compared with user-imported model or library model and the imported EEM data will be fitted with user-imported model or library model. The software has a built-in database, which includes 37 PARAFAC models. By default, constructed PARAFAC models will be compared with the models in library and the models in library are also employed to fit the imported EEM dataset. When "User model" radio-button is chosen, the imported Excel (xls(x)) data should has a sheet named Ex Loadings and another sheet named Em Loadings. In both sheets, the first column is the wavelengths, and the other columns are loading values.

Considering that the PARAFAC models in library have wavelength ranges from 200 nm to 600 nm for both excitation and emission and the interval of 1 nm, the wavelength ranges of EEM data should not exceed that 200 – 600 nm and the intervals of EEM data should not be less than 1 nm, otherwise there may be some calculation error.

In the panel of "Results and information", the top part displays the results' figures and the central part contains sliders and list controls, which updating the results' figures. After clicking the "Save" button, the figures' folder should be given a name, otherwise a folder with name of that day will be made. Then we can choose to save the present figures or all figures. All calculated results will be exported into a folder with name of "Resutls_date_time" after clicking the "Save data" button. The relevant result files are shown in Figure 12. The runtime is also shown in this panel. Some important tips for operating this interface and reminders are shown in the prompt box at the bottom of the panel. All the data I & O information on the interface will be reset by "Reset" button. And the "Interface of data pre-analysis module" can be reached by clicking "Previous" button. The operation on this interface can also be implemented using the menu in the upper left corner of the interface, and each operation corresponds to the shortcut key. In the menu part, the "Expansion" includes functions of "Load EEM results" and "Import user model". The former function will import the previously exported PARAFAC result data, namely EEM_parafac_results_time.mat and then the interface can show figures, save figure, and perform PARAFAC analysis again. The latter function can lead to the interface of

"Import user model into library", which will be introduced later.

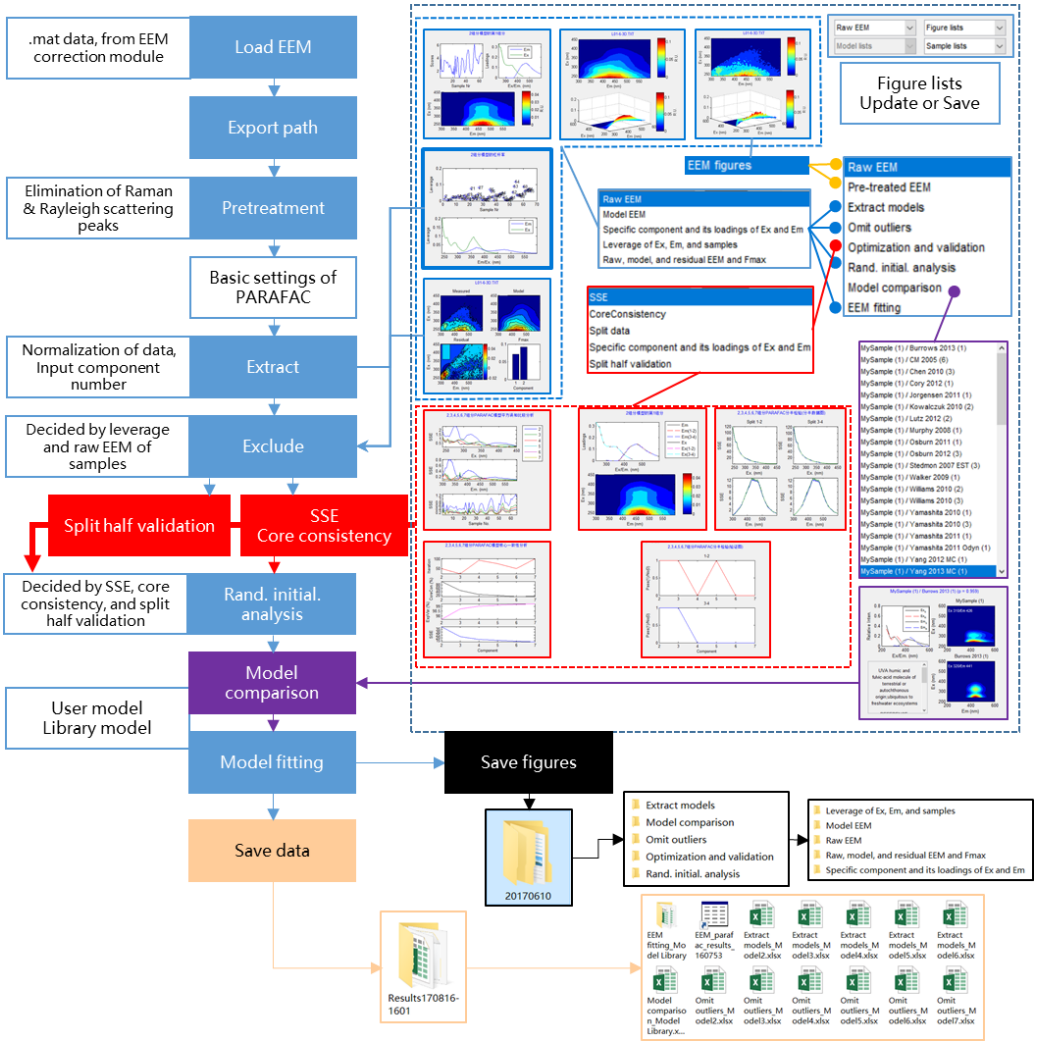


Figure 12 The connection among each operation of EEM-PARAFAC module and its exported figures and data files/folder

Figure 12 shows the sequence of each operation button, the main graphical results, and exported data files. Through the flow chart a complete PARAFAC analysis can be achieved from the data preprocessing (elimination of scattering peaks and omitting outliers), to models' optimization and validation, to model comparison and data fitting, and to other expanded applications.

3.6 Interface of import user EEM-PARAFAC to library

The interface of import user EEM-PARAFAC to library can be reached from the "Expansion" in the menu bar of EEM-PARAFC module (Figure 13). The imported Excel (xls(x)) data should has a sheet named Ex Loadings and another sheet named Em Loadings. In both sheets, the first column is the wavelengths, and the other columns are loading values. Those data can be found in the results files exported by the EEM-PARAFAC modules and directly be imported. After importing the PARAFC

model, the reference or relevant information of the model constructor should be entered. The meanings of all the components should be entered in the tables behind. Finally, the library will be expanded by clicking the "Save model" button.

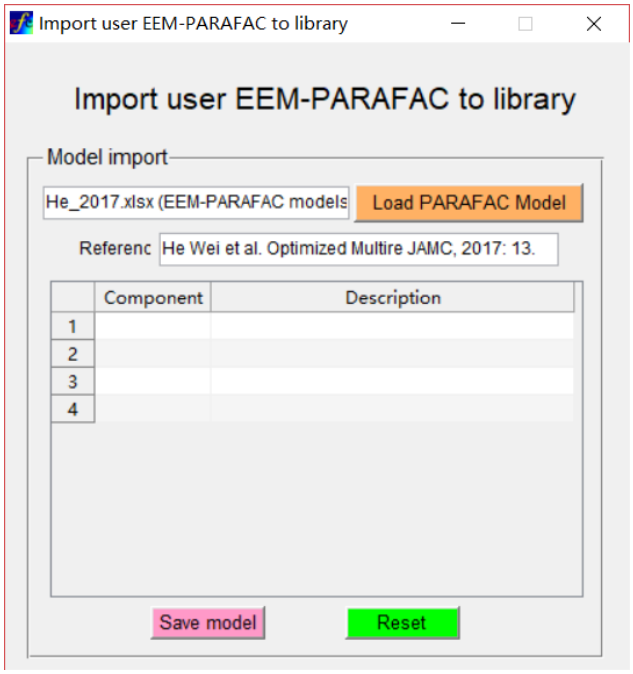


Figure 13 Interface of import user EEM-PARAFAC to library

3.7 Working path, log record, and example data

In the main working path of efc executable software, there are folders, namely "Examples", "Results", and "LOG" (Figure 14). The "Example" folder contains all data used in the "Operation method" part below. The "Results" folder store the exported results. In practice, the user can made their own result folders. The "LOG" folder contains LogReport-date.txt files, which records information in prompt box of each interface, the time of starting, proceeding, and ending of each run, some error messages.

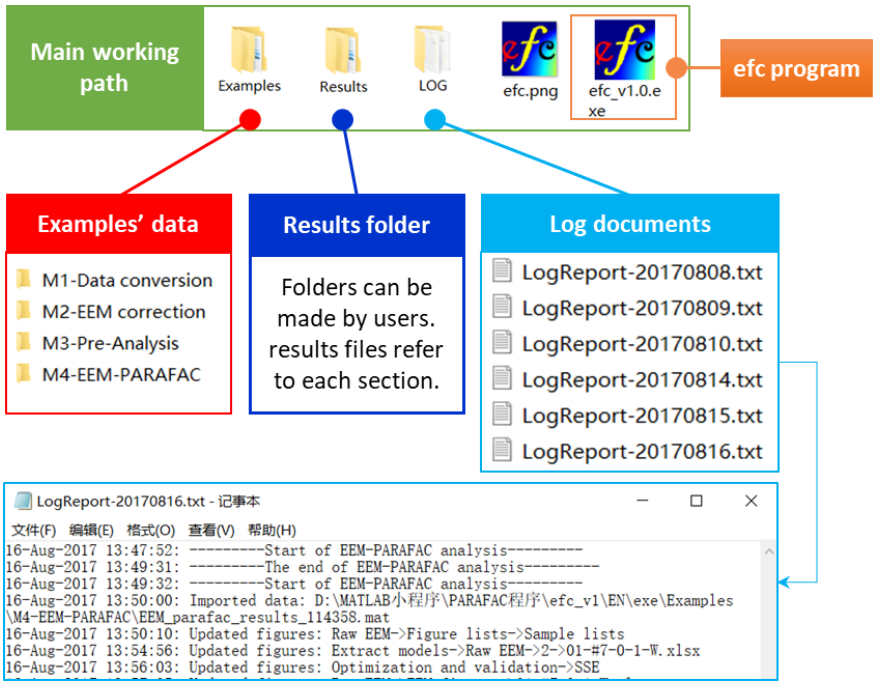


Figure 14 Working path, log record, and example data

4. Operation methods

4.1 Installation and authorization

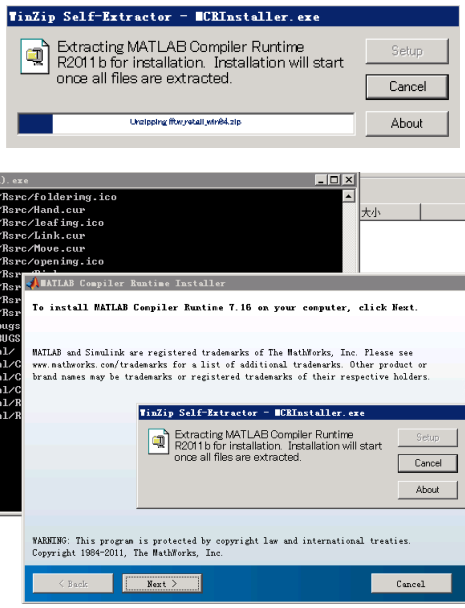
The efc software has two versions of exe (opened in windows system) and pcode (opened in MATLAB). There is no need to install MATLAB when using exe version. However, the MATLAB Compiler Runtime 7.16 is required to be installed for efc software developed by MATLAB 2011b. If that version of MATLAB is installed, there is no need to install compiler. If other MATLAB version is installed, it is required to contact with developer for the relevant version of efc software. The pcode version requires any version of MATLAB software.

4.1.1 exe version

Double-click the installation package (see figure below). If MATLAB 2011b or Matlab Compiler Runtime 7.16 not installed, please double-click the installation package of "efc install (MCR) _64bit.exe". If you have installed MATLAB or relevant version of the compiler, double-click the installation package of "efc install 64bit.exe".

After double-clicking on the installation package of "efc install (MCR) _64bit.exe", the following installation information will be displayed. After the

extraction of MCRInstaller.exe, the dialog box of installation MATLAB Compiler Runtime 7.16 will be mentioned to be installed and click the "Next" until the installation is completed.



Meanwhile, the working path where the efc package is located will display the following files or folders:

If the MATLAB Compiler Runtime is not automatically installed, you can double-click on MCRInstaller.exe until the installation is completed.

Double-click efc_v1.0.exe, and the black window (for some versions of Windows systems) will display. Please do not close that because the window mainly submits the task given by efc.v1.0.exe to the compiler to perform the operation and displays possible error messages. If the right image below, which is the interface of main function list of the efc software, is shown, it means the installation is successful. For an unlicensed version, the module function buttons is grey and cannot be clicked.

With the above steps, the authorized efc.exe can be used normally.

4.1.2 pcode version

For users who are familiar with MATLAB, the software also provides MATLAB p file code, as shown in Figure 15. First, copy the efc_p folder to the toolbox directory of MATLAB; then, use "set path" in MATLAB to set the efc_p folder and its subfolders to working path; finally, enter EFC in MATLAB Command Window to open the interface main function list.

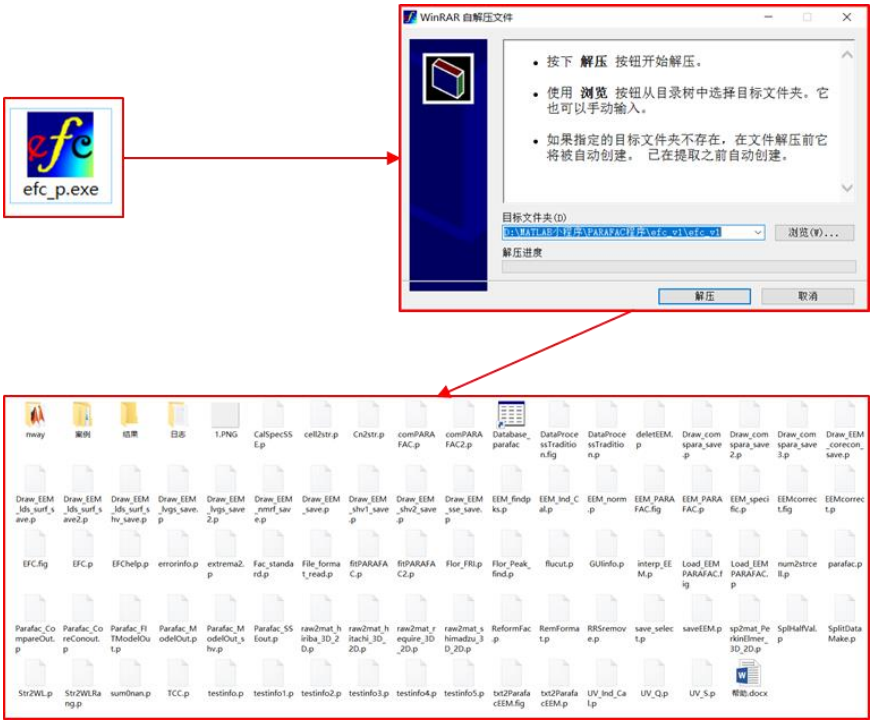


Figure 15 p-code files in working paths

4.1.3 Authorization

The efc software is developed for purpose of scientific application. However, to balance the development costs and scientific applications, this software may charge a little fee for commercial company and require cooperation for nonprofit organizations such as colleges and universities. In order to protect the relevant code, the software developers will customize user product according to the user ID of their PC. Thus, prior to application for use, the user needs to provide the product ID of the computer. As shown in figure below, users first type "systeminfo" in the command prompt (CMD), and the product ID are four groups of numerals with "-" as the connection string. After get the product ID, the requirement of this software can be emailed to the developer. Every email will be replied carefully and users can email without any hesitation.

4.2 Data conversion of EEM dataset

For the EEM conversion module, its core function is corresponding sample data to blank data. The sample and blank data are required to have the same excitation and emission wavelengths and intervals. Although the efc software can independently convert sample data or blank data, it is still suggested to import the sample and blank data simultaneously. The following three cases will help to show the operation ways of corresponding the samples and the blanks: (1) all the samples correspond to the only one blank; (2) some samples correspond to one blank, and others correspond to another blank; (3) the number of samples is the same with that of blanks, and they are one-to-one-corresponding.

4.2.1 Case 1: all samples to a single blank using Perkin-Elmer LS data

LS data are in the folder as shown in Figure 5. When sorting data, please copy data files correspondingly in sub-folders of "Samples" and "Blanks" folders within path of "/Examples/M1-Data conversion/2-LS55 (sp)/". Thus, it facilitates importing sample and blanks data latter. As shown in Figure 16, the operation process of this case is:

(1) Select radio-button "PerkinElmer (LS55)";

(2) Select imported data type of "EEM (3D)";

(3) Load sample data, blank data and data output path;

(4) Set Ex start as 250, end as 500, and gap as 5;

(5) Set Em start as 280, end as 550, and gap as 0.5;

(6) Click "run" to convert those data;

(7) The table in the "Information" panel will show the corresponding relations between samples and blanks.

The result files and folders, commonly named EEM_dataset_ time, can be found in the output path.

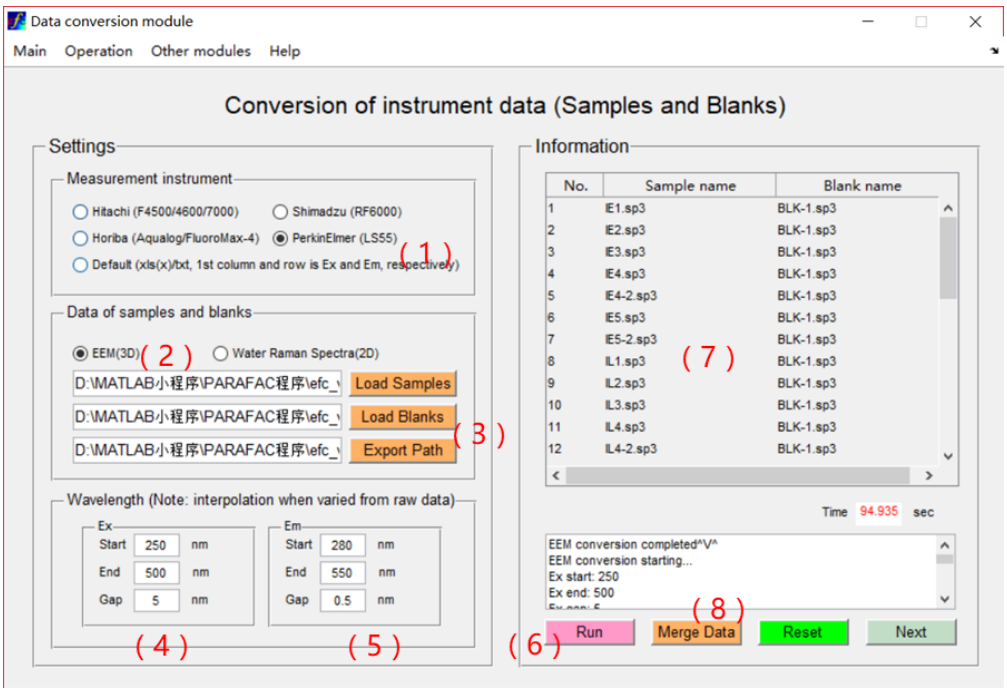


Figure 16 Operation procedure of converting data with all samples corresponding to a single blank using Perkin-Elmer LS data

4.2.2 Case 2: some samples to one blank, other samples to another blank using Hitachi F series data

When sorting data, please copy data files correspondingly in sub-folders (2-some-to-one) of "Samples" and "Blanks" folders within path of "/Examples/M1-Data conversion/1-F4500(txt)/". The EEM (3D) and Raman scan (2D) data can both be converted in this example data. The general operation steps can refer to section 4.2.1. The different steps are as follows:

(3) When loading data, the data format should be txt, otherwise the data cannot be observed. The samples with prefix of 03L correspond to blanks with suffix of L. The samples with prefixes of 03R1 and 03R2 correspond to blank with suffix of R1R2. The samples with prefixes of 03R3, 03R4, and 03R12 correspond to blank with suffix of R3R4R12. Then three pairs of blanks and samples are formed.

(4) The start, end, and gap of Ex is set as 200, 450, and 5.

(5) The start, end, and gap of Em is set as 220, 600, and 2.

(8) After clicking "Run" for three pairs of blanks and samples. Three EEM_datase_time.mat can be obtained. Click the "Merge Data" button to combine all those three pairs to one merged data as shown in Figure 17.

Figure 17 The key operation procedures of converting data with all samples corresponding to a single blank using Hitachi F series data

For PerkinElmer LS series data in case II, the data folders should be placed into separate folders as shown in Figure 18. For other measured data, there is no need to place data into different folders. The user just remember the corresponding relationships between samples and blanks. However, the arrangements in Figure 15 can make this step much clearer.

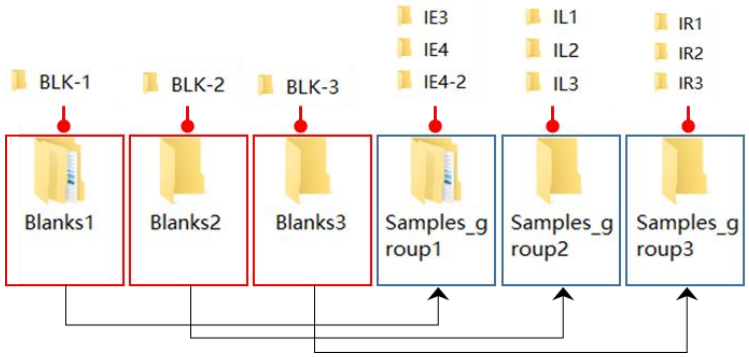


Figure 18 The data arrangement when converting data with all samples corresponding to a single blank using Perkin-Elmer LS series data

When converting the Raman scan data, the basic operation refers to section 4.2.1. The different steps are as follows:

(2) Select imported data type of "Water Raman spectra (2D)".

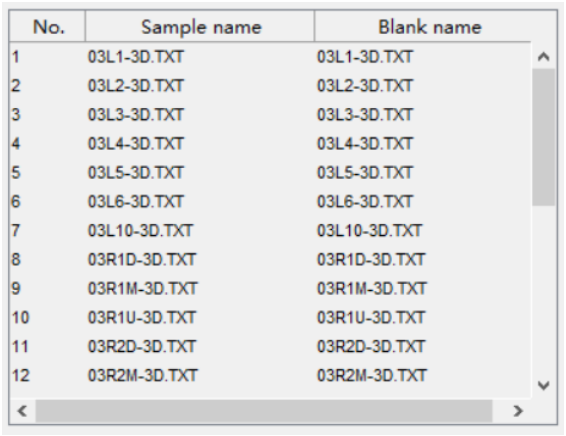
(3) Ex is set as 350.

(4) The start, end, and gap of Ex is set as 365, 450, and 0.2.

4.2.3 Case 3: one sample to one blank using Hitachi F series data

When sorting data, please copy data files correspondingly in sub-folders (3-one-to-one) of "Samples" and "Blanks" folders within path of "/Examples/M1-Data conversion/1- F4500(txt)/". The EEM is processed in this case. The operation of water Raman scan data is the same with EEMs'. Referring to section 4.2.1, the number of blanks' data files is the same with that of samples' and their

corresponding relationships are shown in Figure 19.



No.	Sample name	Blank name
1	03L1-3D.TXT	03L1-3D.TXT
2	03L2-3D.TXT	03L2-3D.TXT
3	03L3-3D.TXT	03L3-3D.TXT
4	03L4-3D.TXT	03L4-3D.TXT
5	03L5-3D.TXT	03L5-3D.TXT
6	03L6-3D.TXT	03L6-3D.TXT
7	03L10-3D.TXT	03L10-3D.TXT
8	03R1D-3D.TXT	03R1D-3D.TXT
9	03R1M-3D.TXT	03R1M-3D.TXT
10	03R1U-3D.TXT	03R1U-3D.TXT
11	03R2D-3D.TXT	03R2D-3D.TXT
12	03R2M-3D.TXT	03R2M-3D.TXT

Figure 19 The corresponding relationships between samples and blanks in case 3

4.3 EEMs' correction

The general steps of EEMs correction are as follows and shown in Figure 20:

(1) Load raw data. The data here, which is called EEM_dataset_time.mat is from the previous module. Sets the exported data's name and path.

(2) Select the way of Raman normalization. If "All-to-one" or "One-to-one" is selected, the Raman data should be loaded. For the selection of "Auto", the excitation wavelength and the emission wavelength ranges should be set or the default values are used.

(3) Choose whether the spectral correction will be done or not. If the spectral correction is selected, the correction factor data needs to be imported.

(4) Set the wavelength ranges and intervals. Generally, the software will automatically read the wavelength ranges and intervals of the import EEM data. If some modification is done here, the interpolation may happen. The required wavelength settings is that ranges should not exceed that of the imported data and the interval should be from 0.1 to 5.

(5) Choose whether to carry out the inner filter effect correction. For the imported absorption spectrum in this software, the absorbance at 254 nm should less. If it cannot be satisfied, the correction is only a reference.

(6) Select the way of blank subtraction. See the case introduction in different cases.

(7) Choose whether to conduct the quinoline sulfate (QS) correction. If the QS correction is not conducted, the exported results will not contain EEM data with QS unit.

(8) Click the "Run" to perform the correction. The correction time and correction processes will be shown in the information box.

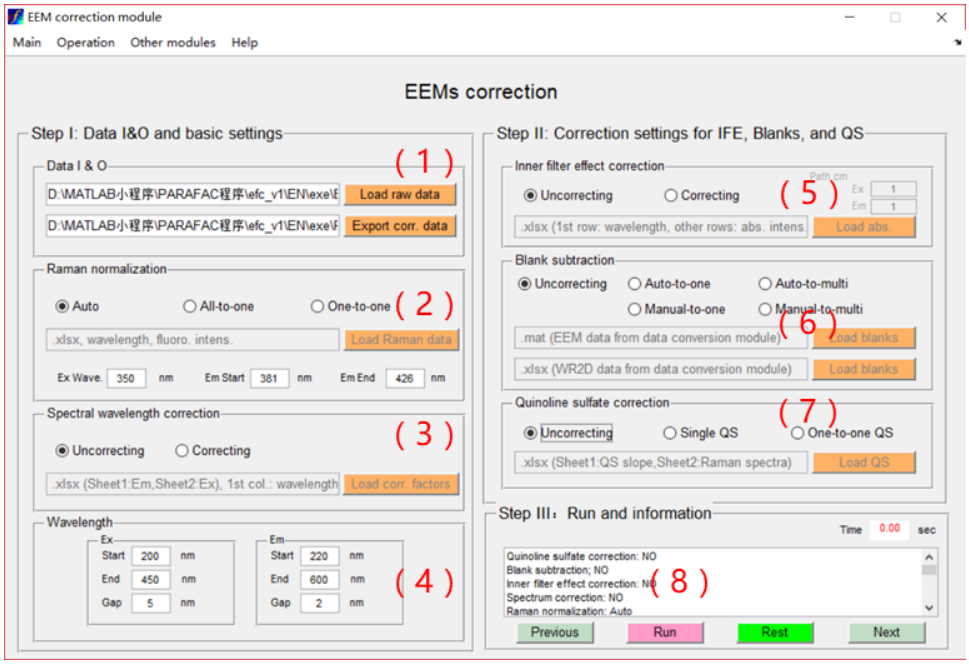


Figure 20 The general operation procedure of EEMs' correction

4.3.1 Case 1: non-correction

After loading the EEM dataset (EEM_dataset_101500.mat) and set the output path and file name, he non-correction case is set as shown in Figure 20. The 1st item will be selected in each radio-button groups. For Raman normalization, the excitation wavelength should be within the imported EEM dataset, otherwise the nearest wavelength might be employed. If the emission wavelength ranges is not included in the imported data, the nearest inside data will be employed. The imported data in the latter two interfaces is from this interface, therefore the EEM data, which has been corrected through blank subtraction, scattering elimination, spectral correction, and inner filter effect by instrument, is required to loaded into the present interface and choose the non-correction way. The exported EEM dataset only contains EEMs with the Raman unit (R.U.), which is a widely employed right now.

4.3.2 Case 2: simple correction

If the corresponding blanks or/and QS slopes of samples or/and the Raman spectra are not measured, the simple correction can be performed in this situation. Based on the procedures in section 4.3.1, the different steps should be conducted instead as follows:

(2) Select "Auto" and use default parameters.

(3) Select "Uncorrecting" for non-correction of spectra here;

(5) Select "Correcting" to conduct inner filter effect correction with default optical path. The imported absorption spectra data file is "UV-VIS.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

(6) Select "Manual-to-one" and load blank data file "WR2D_blanks2.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

(7) Select "Single QS" and import QS data file "QS corr1.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

4.3.3 Case 3: strict correction

If the EEM correction is strictly performed, each correction item needs a measured data. Ideally, this strict correction is conducted as follows based on the procedures in section 4.3.1:

(2) Select "One-to-one" and use default parameters. The imported Raman spectra data file is "WR2D_samples.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

(3) Select "Correcting" and import data file "Spectra Corr factor.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

(5) Select "Correcting" and use the default parameter. Imported absorption spectra data file is "UV-VIS.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

(6) Select "Manual-to-multi" and load the blanks' EEM data file "EEM_dataset_101500" and blanks' Raman spectra data file "WR2D_blanks3.xlsx".

(7) Select "One-to-one QS" and load QS data file "QS corr2.xlsx" in the subfolder of "M2-data correction" of folder "Examples".

4.4 Calculation of adsorption and fluorescence indicators

The calculation process of absorption and fluorescence indicators is carried out as follows and shown in Figure 21:

(1) Load fluorescence data (EEM) and/or absorption spectrum data.

(2) Check the absorption spectra indicators which is required to be calculated.

(3) Pretreat EEMs' scattering peaks and choose one of three ways of "Unclear" for non-elimination, "Clear" for elimination and interpolation, and "Cut" for only elimination without interpolation.

(4) Check the fluorescence spectra indicators which is required to be calculated.

(5) Click "Run" to calculate;

(6) Display the contour and surface plots through the drop-down box or slider. Click the "Save" button and select saved images' format (png, bmp, emf, eps, jpg, pdf, tif, fig, ai, and pcx formats). The present figure or all the figures of EEMs are exported will be chosen after clicking "Save" button.

4.4.1 Case 1: calculation of indicators

If only calculation of indicators is the purpose, the steps (1), (2), (4), and (5) can give the results. In the menu "Operation", the "All abs. indicators" and "All fluorescence indicators" can check all the indicators or remove all the indicators.

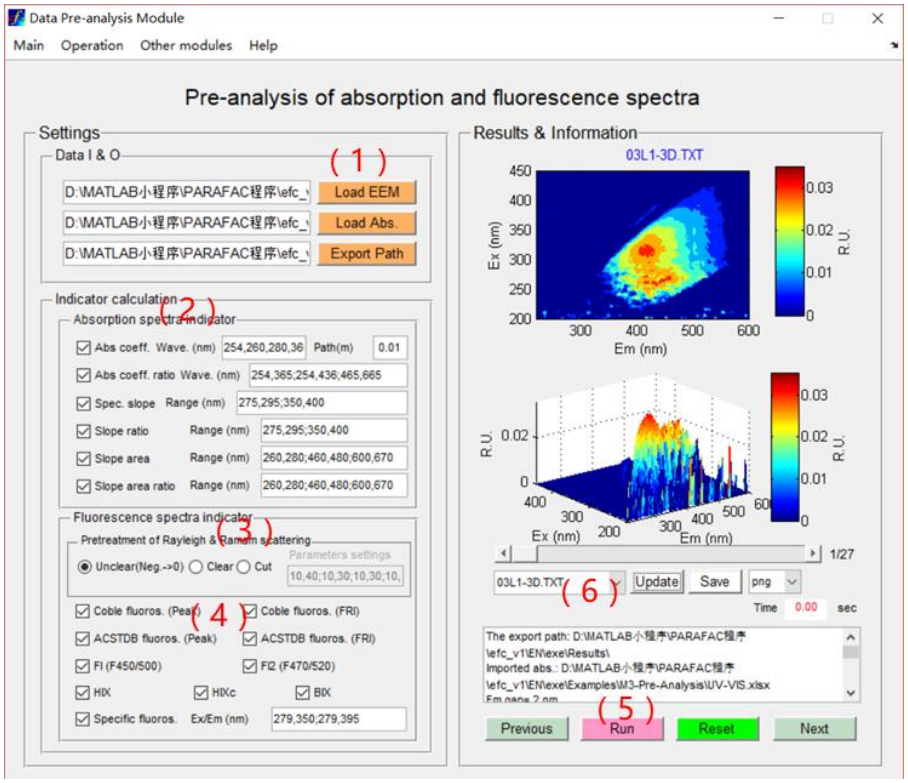


Figure 21 The calculation procedure of adsorption and fluorescence indicators

4.4.2 Case 2: elimination of scattering peaks

In order to get the best parameters for elimination of EEMs' scattering peaks, after adjusting the parameters x1, y1, x2, y2, x3, y3, x4, and y4, perform the Steps (5) and (6), and check whether the elimination is appropriate. Repeat the operations above several times, the best parameters will obtained. The parameters' control on the elimination of scattering peaks is shown in Figure 22. Those parameters can be used in the latter EEM-PARAFAC interface.

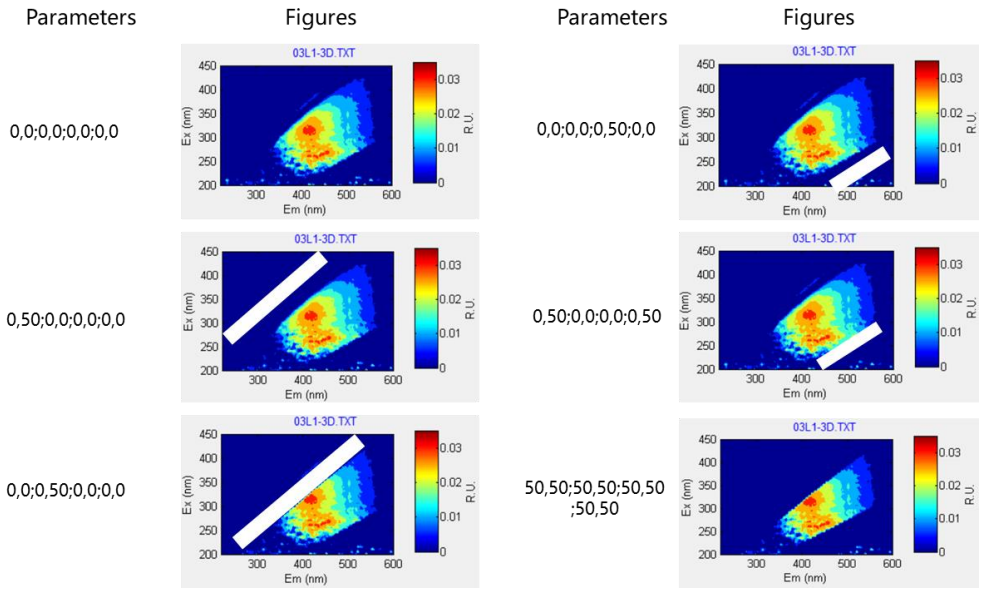


Figure 22 Parameter settings of scattering peaks' elimination and the eliminated figures using different parameters

4.4.3 Case 3: save of EEM figures

After loading EEM dataset and set the export path using Step (1), the present EEM figures and all the EEM figures can be exported using Step (6). Of all the image formats, figures in fig and pdf formats are generally very clear. However, fig files should be opened by MATLAB. The png figures have less storage space and muc clearer than images in tif and bmp formats. Figures in ai format can only be oponed by software such as Adobe Illustrator and Photoshop.

4.5 EEM-PARAFAC analysis

EEM-PARAFAC analysis follows the basic procedures in Figure 12. The specific operation steps is as follows and shown in Figure 23:

(1) Load the EEM data, which is the result data from the interface of EEM correction.

(2) Fill in the pretreatment parameters of Rayleigh and Raman scattering peaks. The determination of the best parameters refers to section 4.4.3. We can also take the parameters obtained in the previous interface.

(3) Fill in the basic setting parameters of PARAFAC model. When the fitting accuracy requirement is not that high, user can increase the PARAFAC convergence value, such as 1e-4, or reduce the maximum number of iterations. The meanings of the parameters refer to previous introduction.

(4) Choose whether the EEM data is normalized or not. If the "Normalization" is chosen, all data will be normalized to the maximum value of the fluorescence

intensity.

(5) After entering the models' total components, click the "Extract" button to perform the preliminary PARAFAC analysis, and there will be dialog box to indicate whether the task is completed or not.

(6) Through the preliminary analysis, the outliers will be identified. After entering the outliers' series number, click the "Exclude" button, a second PARAFAC analysis without outliers will be performed, and there will be dialog box to indicate whether the task is completed or not.

(7) Click the "SSE" button to summarize SSE results.

(8) Click the "Core consistency" to summarize core consistency results.

(9) Click the "Split half validation" button to make split half data and conduct the PARAFAC analysis using split half data. Then the split half validation results will be summarized.

(10) Through Steps (7), (8), and (9), the best PARAFAC models will be found. Then enter the best model's component number (single number). After that, click the "Rand. initial. analysis" button to obtain the best PARAFAC model and to calculate the score matrix of each component for the best model.

(11) Choose the "User model" or the "Library model" as the comparison object. If the "User model" is selected, the user should imported a PARAFAC model, otherwise the 37 models in library will be selected. Click the button "Model comparison" to start the models' comparison.

(12) The model importing is similar to the Step (11).Click the button "EEM fitting" to realize fitting the user's EEM data with the PARAFAC models. Then the scores of each component in the relevant model will be obtained.

(13) Display different output results by the list control or slider. Save the figures of concern.

(14) Click the "Save data" button to save the data of all important results. Generally, Excel format data will be exported, and the processing data will be stored in the mat format data.

(15) Click the option "Load PARAFAC results" in the "Expansion" of the menu to call the previously imported data, display results, and re-do some steps.

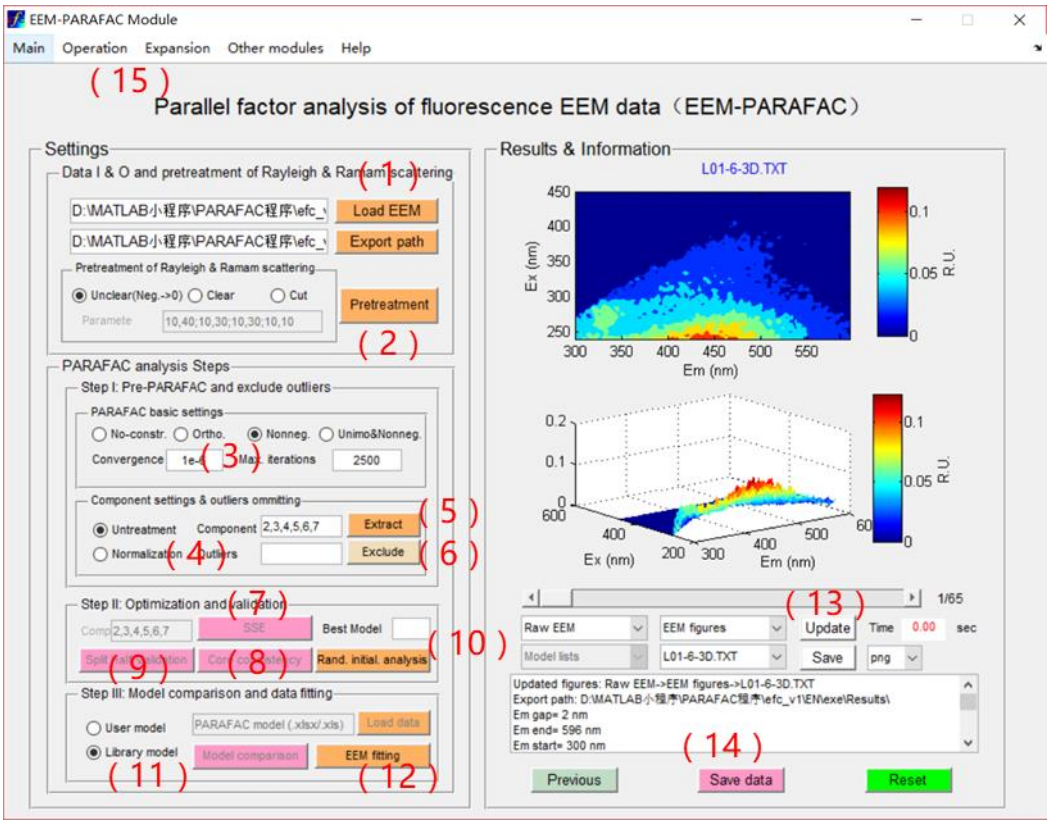


Figure 23 The basic operation procedure of EEM-PARAFAC analysis

4.5.1 Case 1: an intact PARAFAC analysis

An intact PARAFAC analysis includes elimination of scattering peaks, outlier removing, core consistency analysis, SSE analysis, and split half validation. Generally, the split half validation is a proposed step to identify the best model. However, sometimes the EEM dataset has only a few samples or its heterogeneity is too big. Then the SSE, explained variations, and core consistency can be used to determine the best model. To conduct an intact PARAFAC analysis, the Steps (1) to (10) as well as (13) will be followed. Finally, the results can be exported by Step (14).

To find out the outliers, which should be omitted, the list controls "Extract models" – "Leverage of Ex, Em, and samples" – "2 to N" will give the leverage of samples in different models (Figure 24). The potential outliers is No. 5, 21, 30, and 53 judged by the leverage for the example data. After plotting the contour and surface figures of those EEMs (Figure 25), it is easy to find that only No. 5 and 30 EEMs have measuring problem, and those EEMs should be measured again or omitted in the latter EEM PARAFAC analysis. The other two EEMs have good measurement but a little heterogeneity compared with other EEMs. We can keep those two EEMs.

After entering the Outliers 5 and 30, the PARAFAC analysis is conducted again. The optimization and validation, including SSE, core consistency, and split half

validation, will be performed to give results in Figure 26. And the model with 3 components is the best model. Then the EEM-PARAFAC analysis with random initialization is conducted for the 3-component model to optimize the models' loadings and scores. The data are saved when carrying out the Step (14).

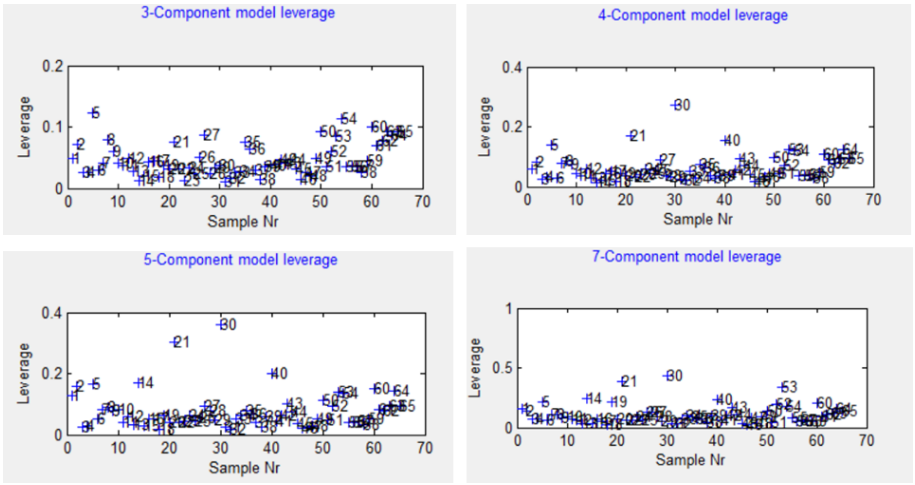


Figure 24 Leverages of PARAFAC models with various components (3, 4, 5, and 7)

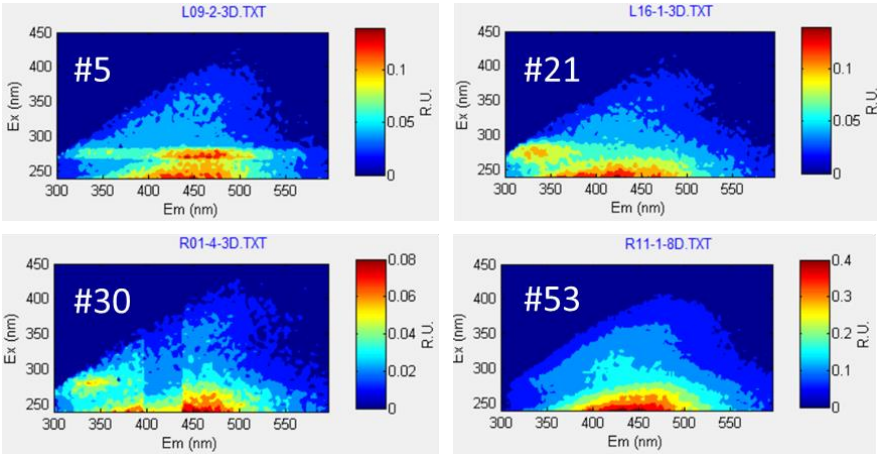


Figure 25 The contour figures of potential outliers

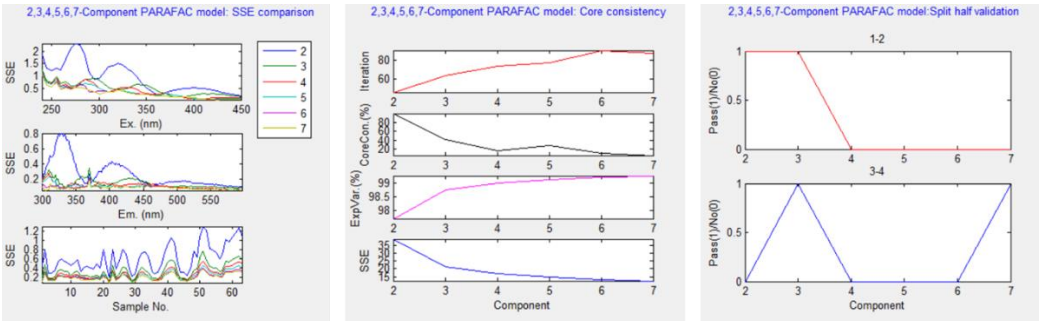


Figure 26 The SEE, core consistency, and split half validation results

4.5.2 Case 2: comparison with PARAFAC models in literatures

To quantitively figure out the meanings of the best model's components, the Step (11) is performed in this case based on the intact EEM PARAFAC analysis in section 4.5.1. After comparison with the model in library as shown in Figure 27 through the Step (13), we can know the meanings of each component. For the example data, the component 1 is similar to component 2 of Williams 2010 PARAFAC model with mTCC value of 0.953. The box at the left-bottom shows the type of the component. The component 1 is a king of humic-like, terrestrial, and bio-refractory substance. The maximum excitation and emission wavelength of each component is also shown. The users is Ex 310/Em 426, whereas Williams 2010's is Ex 316/Em 418. After checking the contour plots, both two components are very similar.

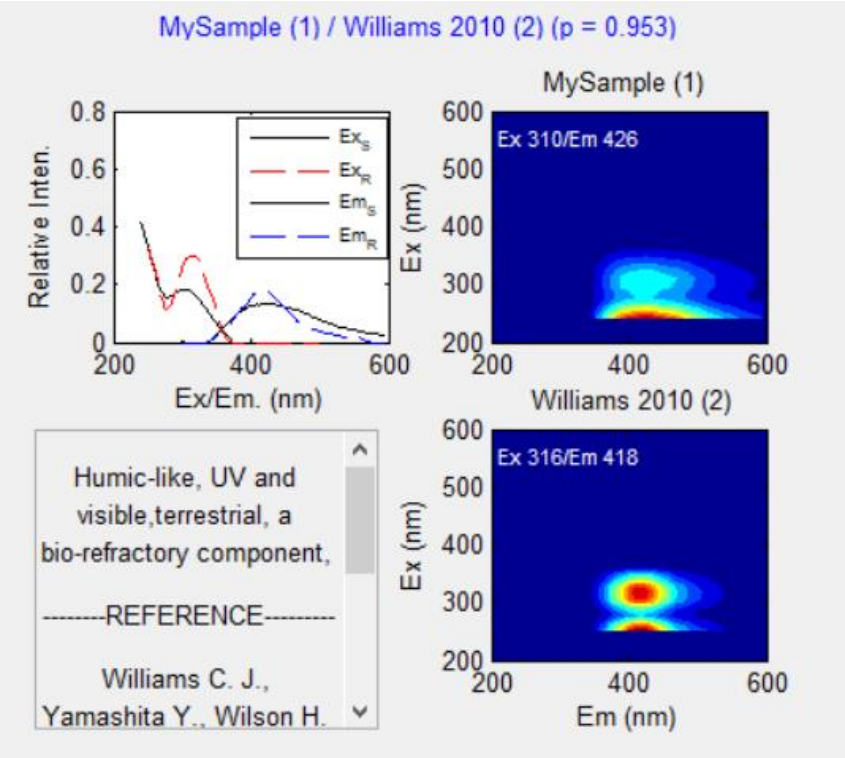
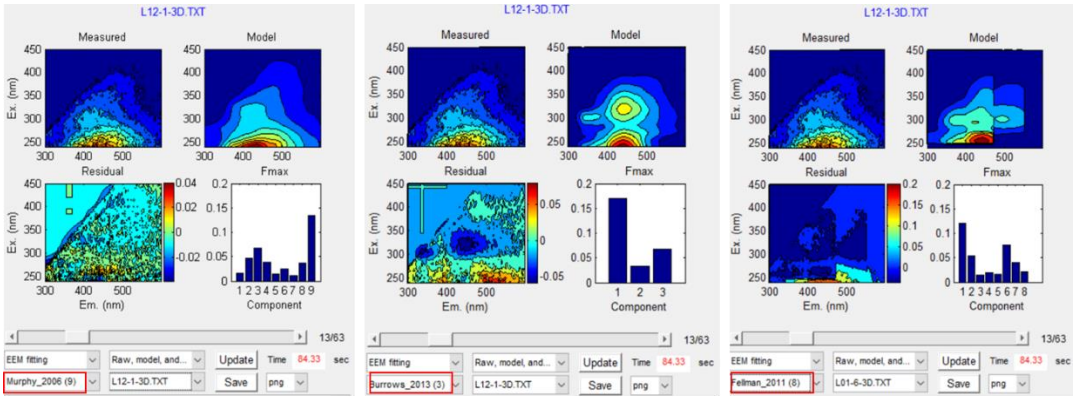


Figure 34 Results of model comparison

4.5.3 Case 3: fitting data using PARAFAC models in literatures

To fit the EEM data using other researchers' models, the Step (12) is required. Through the Step (13), the EEMs' fitting will be completed. The data can be saved by the Step (14). Some results are shown in Figure 28. The error (Err), explained variations (ExpVar), SSE, and the RSSE are all very small for Murphy (2006)'s model of those three models. There is few differences between the measured and model EEM for Murphy (2006)'s model as shown in Figure 34.



Models	Iteration	Err	Consistency	ExpVar	SSE_mean	SSE_Std	RSSE_mean	RSSE_std
Murphy_2006 (9)	2	33.5	0.1	98.0	0.53	0.33	2.15	0.41
Burrows_2013 (3)	2	254.6	19.9	85.1	4.04	2.84	15.13	1.83
Fellman_2011 (8)	2	705.7	-0.1	58.6	0.41	0.25	1.71	0.38

Figure 28 Results of data fitting with models in library

4.6 Import PARAFAC model into library

The basic steps for importing PARAFAC model into library are as follow and shown in Figure 29:

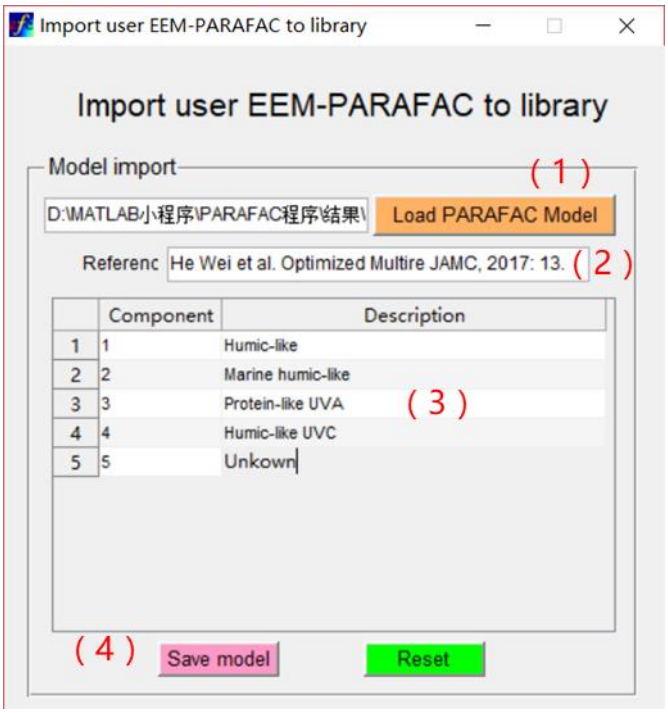


Figure 29 The loading procedure of PARAFAC model into library

(1) Load PARAFAC model, which is saved data in interface of EEM-PARAFAC analysis, such as "Extract models_Model5.xlsx". In that Excel file, there sheets,

namely Ex Loadings and Em Loadings, are used as in this interface.

(2) Enter the reference of the imported PARAFAC model. If the reference cannot be provided, some other information such as places and media. After the library model is expanded, we can know the information or sources of the imported model when operate the model comparison or EEM fitting in the future.

(3) After automatically reading the model component, the tables with component will be shown in Figure 36. It is required to fill the represented substances of each component in the table.

(4) Click the "Save model" button to update the model library.