**Metabolic Confounding Effect Elimination Tool**

**(MCEE) V1.0**

**User Guide**

Shanghai Jiao Tong University Affiliated Sixth People’s Hospital

Contents

[1 Introduction 3](#_Toc510105478)

[**1.1** **Hardware environment** 3](#_Toc510105479)

[**1.2** **Software environment** 3](#_Toc510105480)

[**1.3** **Programming language and version** 3](#_Toc510105481)

[**1.4** **Main functions and technical features** 3](#_Toc510105482)

[**1.5** **Potential users** 3](#_Toc510105483)

[2 Installation instructions 4](#_Toc510105484)

[3 Software function and operation steps 4](#_Toc510105485)

[**3.1** **Software startup** 4](#_Toc510105486)

[**3.2** **Software interface and operation process** 4](#_Toc510105487)

[**3.2.1** **Import data** 4](#_Toc510105488)

[**3.2.2** **Parameter setting** 6](#_Toc510105489)

[**3.2.3** **Program running** 6](#_Toc510105490)

[**3.2.4** **Check the PCA scores plots** 7](#_Toc510105491)

[**3.2.5** **Close the software** 7](#_Toc510105492)

1. **Introduction** 
   1. **Hardware environment**

Personal computer, internal storage 512Mb or above, CPU main frequency 1.66GHz or above.

* 1. **Software environment**

Windows XP or higher operating system, Matlab2014a or above, Office Excel 2003 or above.

* 1. **Programming language and version**

Matlab language, Matlab 2014a.

* 1. **Main functions and technical features**

Metabolomics is an important part of systems biology and translational medicine. It refers to a discipline that studies the metabolic network of organisms by analyzing the structures and changes of the metabolites produced by stimuli or disturbances. It is now widely used in a variety of studies, such as disease diagnosis, new drug development, drug effect and toxic mechanism.

It is well recognized that physiological and environmental factors such as race, age, gender, and diurnal cycles often have a definite influence on metabolic results that statistically manifests as confounding variables. Currently, removal or controlling of confounding effects relies heavily on experimental design. There are no available data processing techniques focusing on the compensation of their effects. We therefore proposed a new method and software, Metabolic Confounding Effect Elimination (MCEE), to remove the influence of specified confounding factors and make the data more accurate. The method consists of three steps: metabolites grouping, confounder-related metabolites selection, and metabolites modification. The software is based on MATLAB platform and is simple with friendly interface.

Technical features:

1. The elimination of confounding factors in a metabolomics dataset;
2. Applicable to multi-platform data, such as gas phase mass spectrometry or liquid phase mass spectrometry data.
3. This method, based on the generalized linear regression model (GLM), can eliminate the bias of one confounding factor at a time.
4. An adjustable parameter “*ss*” is designed to control the processing strength.
5. It does not affect the number of metabolites or samples.
6. PCA scores plots derived from datasets before and after processing are provided to show the processing performance.
7. Input and output are .xls documents, convenient and quick operation.
   1. **Potential users**

This tool is suitable for researchers in medical, biological, chemical, or related fields who have little basis of data analysis and software operation.

1. **Installation instructions**

This is a green software with no need for installation, and can be used directly in Matlab platform.

1. **Software function and operation steps** 
   1. **Software startup**
   2. Start Matlab.
   3. Set up the work path.
   4. In the Current Folder window, find MCEE.m, right click, and select run (Figure 1).

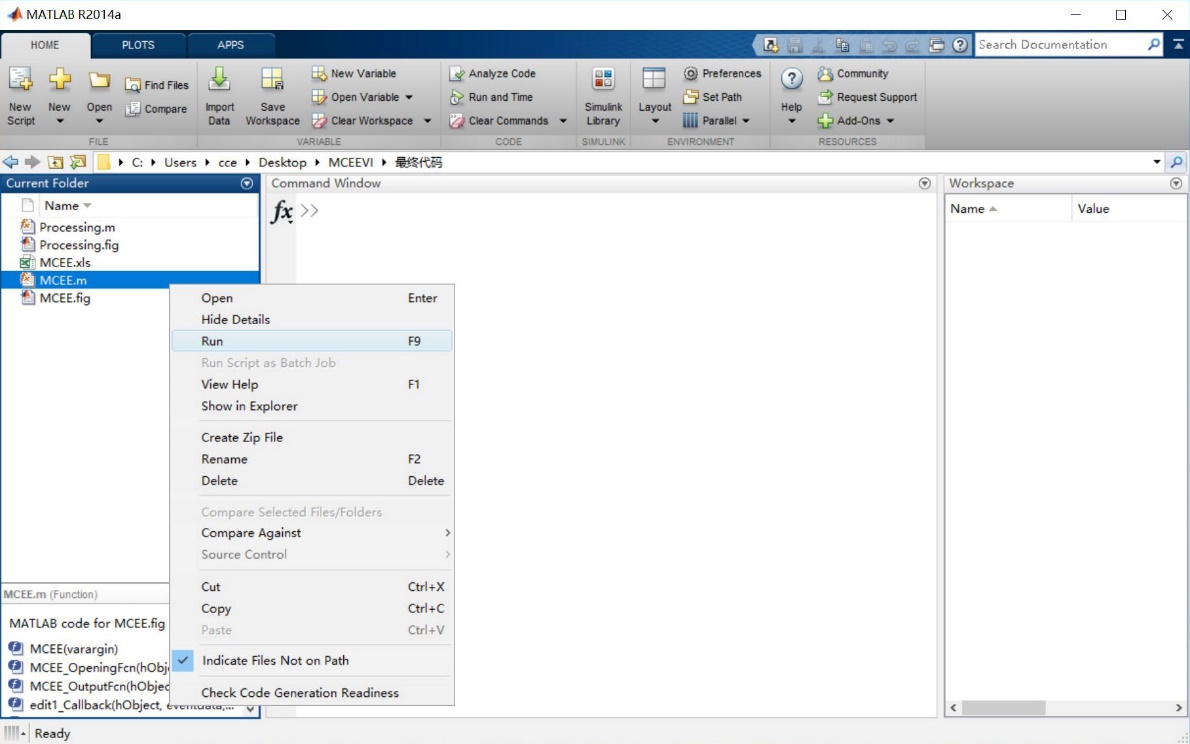


Figure 1. Start MCEE in Matlab

* 1. **Software interface and operation process** 
     1. **Import data**

The initial interface of MCEE is shown as figure 2.

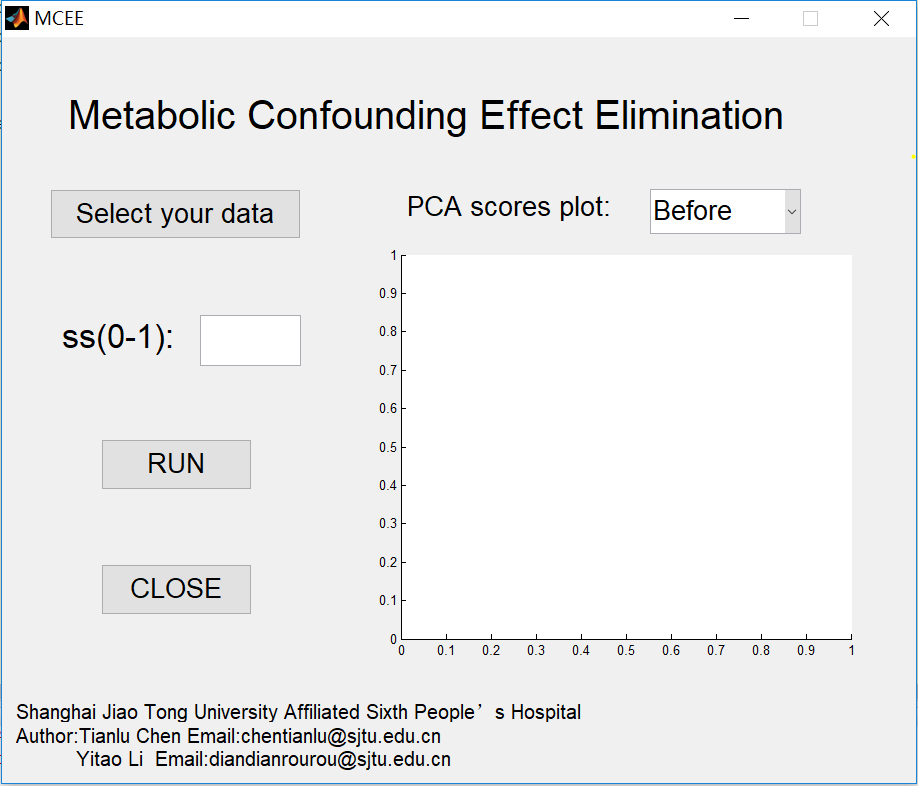


Figure 2. The initial interface of MCEE

Click "Select your data" and select the data file in the pop-up window and open it.

Input data format (Figure 3)

1. Acceptable file format: .xls.
2. The file should be a pure digital content, and text content such as the name of the metabolites is not allowed.
3. In the sheet, each row represents a sample, and each column (except the first two columns) represents a metabolite.
4. The first column of the data represents groups, such as the normal group and the disease group, which are expressed in 1, 2, 3, … respectively.
5. The second column represents the confounding factor you want to remove.

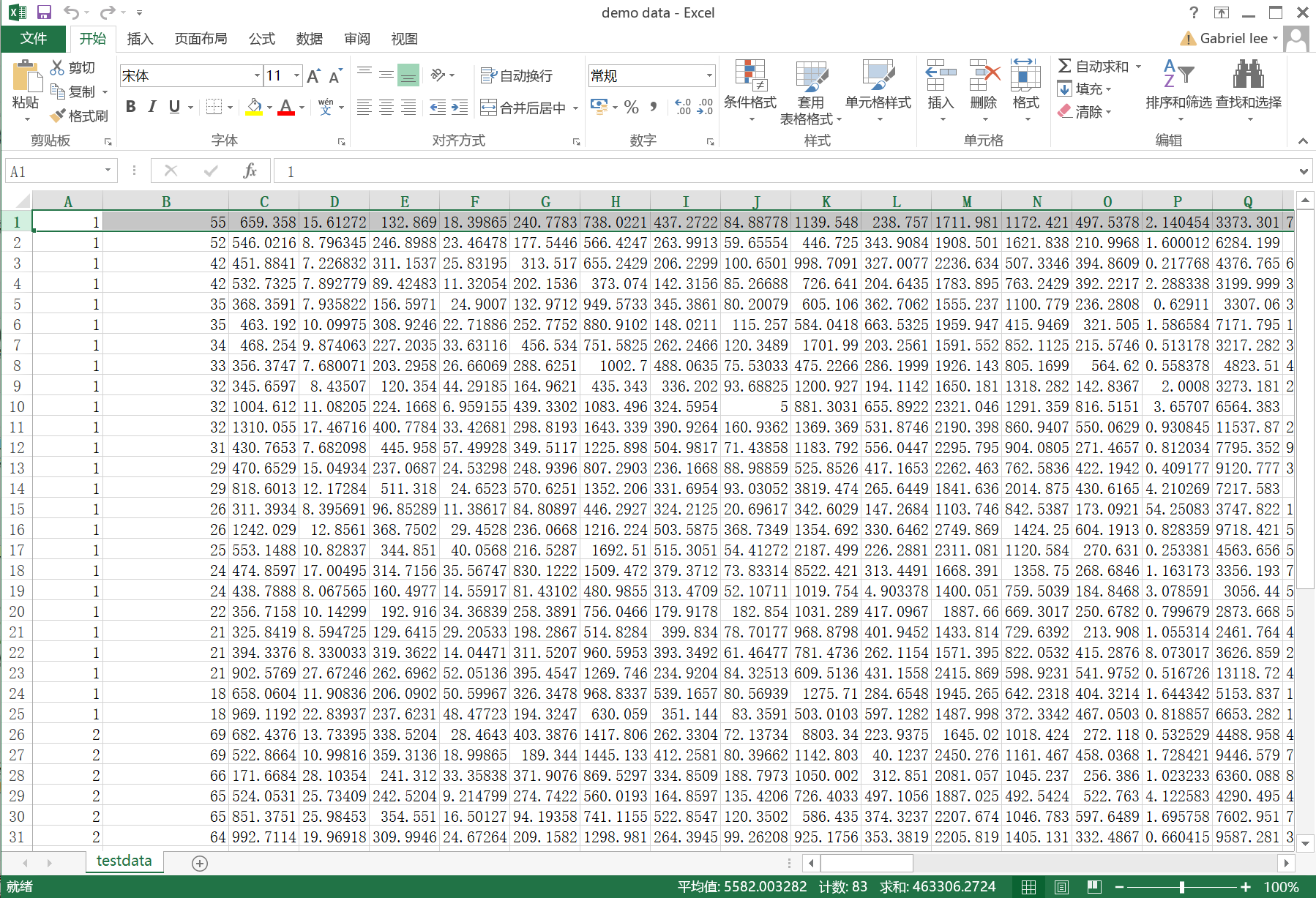


Figure 3. Input data format

* + 1. **Parameter setting**

Enter the value (0-1) of “ss” in the input box as needed.

* + 1. **Program running**

Click “RUN” to start the processing. A Processing interface will pop up. When the processing is finished and the corrected data named MCEE.xls is saved, the text will display in the window ” The result has been saved to the folder where the software is located.” (Figure 4). Click “CLOSE” to close this interface.

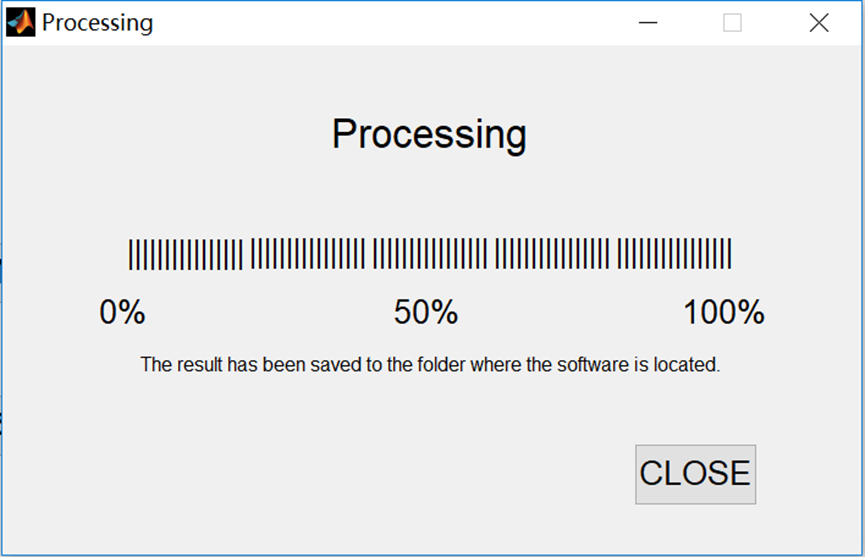


Figure 4. The interface of Processing

* + 1. **Check the PCA scores plots**

When the processing is finished, the PCA scores plot derived from raw data will appear. You may check the plot derived from processed data via the drop-down menu (Figure 5).

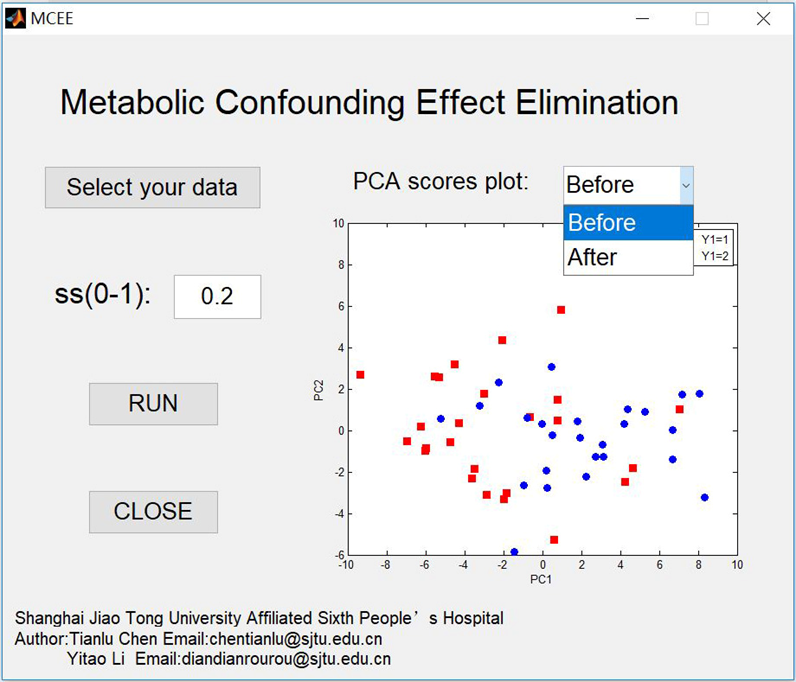


Figure 5. Drawing the PCA scores plots

* + 1. **Close the software**

Click the "CLOSE" button to close the software.