

Full name of software: **A preprocessing approach for metabolic  
confounding effect elimination**

Software abbreviated: **MCEE**

Software version: **V2.0**

## **User Guide**

2019-3-20

Shanghai Jiao Tong University Affiliated Sixth People's Hospital

[Wjia4226@hotmail.com](mailto:Wjia4226@hotmail.com)

[chentianlu@sjtu.edu.cn](mailto:chentianlu@sjtu.edu.cn)

# Contents

<b>1 Survey .....</b>	<b>1</b>
1.1 Hardware environment.....	1
1.2 Software environment .....	1
1.3 Programming language and its version.....	1
1.4 The number of source code .....	1
1.5 Main functions and technical features.....	1
1.6 Purpose .....	2
<b>2 Software introduction .....</b>	<b>2</b>
2.1 Overview .....	2
2.2 Software features .....	2
2.3 Interface design .....	2
2.4 Introduction of main functions .....	2
<b>3 Installation instructions.....</b>	<b>2</b>
<b>4 Software function and operation steps .....</b>	<b>2</b>
4.1 Software startup .....	2
4.2 The operation process and interface of software .....	3
4.3 Summary .....	7

# **1 Survey**

## **1.1 Hardware environment**

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

## **1.2 Software environment**

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

## **1.3 Programming language and its version**

Matlab language, Matlab 2014a.

## **1.4 The number of source code**

931

## **1.5 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 structure and change 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. The development and improvement of metabolomics technology is related to the collation, correction and analysis of metabolomics based on high-throughput instruments and / or biological samples. The software for the metabolomics data correction software, can eliminate and reduce the effects of confounding factors for metabolomics data, to help people find the potential confounding factors which affected by confounding factors.

Technical features:

- a) The elimination of confounding factors in metabolomics;
- b) Applicable to multi-platform data, such as gas phase mass spectrometry or liquid phase mass spectrometry data.
- c) Two methods, generalized linear regression model and canonical correlation discrimination, for users to choose.
- d) Effects of multiple confounding factors can be eliminated simultaneously.
- e) Sufficient parameters are provided, and recommended parameters can be calculated according to user requirements.
- f) It does not affect the number of metabolites or samples.
- g) PCA scores plots are provided hoping to evaluate the data preprocessing performance.
- h) Inputs and outputs are excel documents, convenient and quick operation.

## **1.6 Purpose**

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

## **2 Software introduction**

### **2.1 Overview**

The software is based on MATLAB language. It can be used to eliminate the influence of all kinds of specified confounding factors.

### **2.2 Software features**

Same as section 1.5.

### **2.3 Interface design**

The interface is simple and easy to operate.

### **2.4 Introduction of main functions**

It is used to adjust the metabolites in the metabolomics data in order to eliminate the influence of confounding factors.

## **3 Installation instructions**

This software does not need to be installed, and can be used directly in Matlab platform.

## **4 Software function and operation steps**

### **4.1 Software startup**

- 1.1 Launch Matlab.
- 1.2 Set up the work path.
- 1.3 In the Current Folder window, find MCEE.m, click right, and select run (Figure 1).

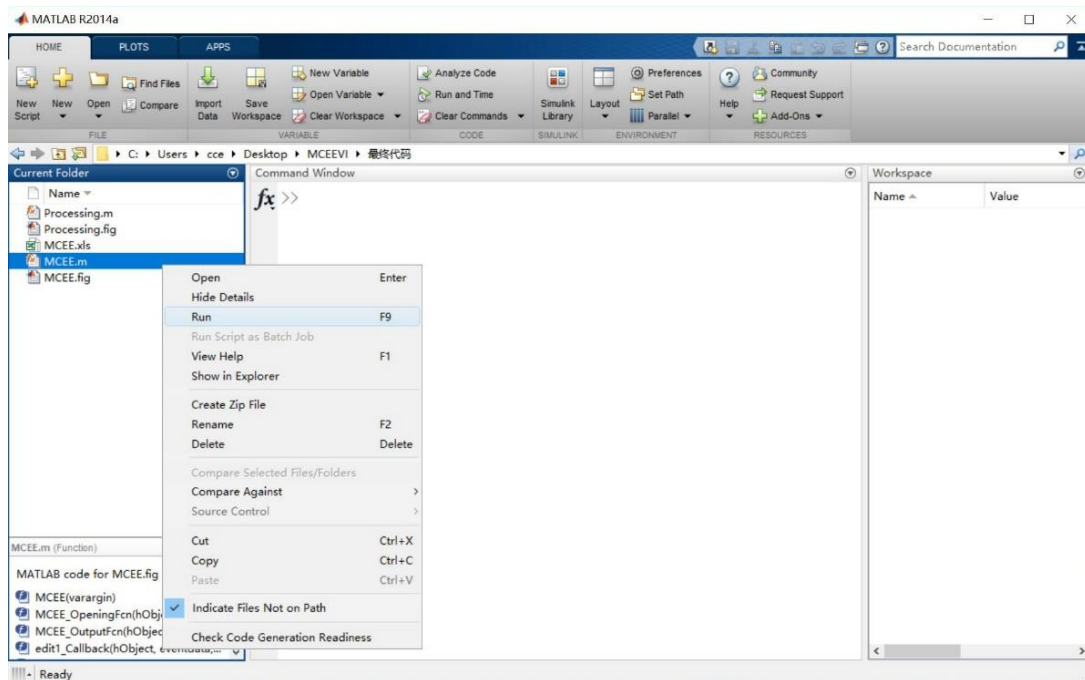


Figure 1. Start MCEE in Matlab

## 4.2 The operation process and interface of software

### 4.2.1 Import data

The initial interface of MCEE is shown in figure 2.

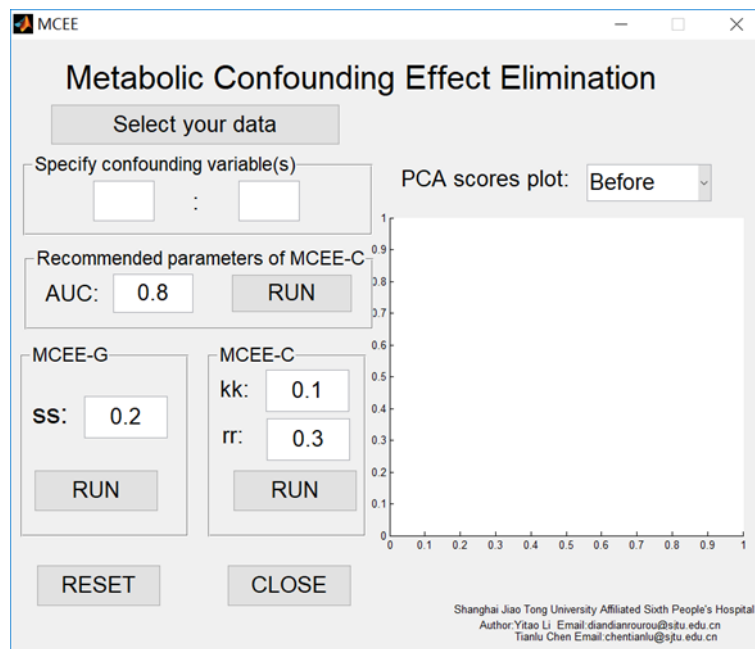


Figure 2. The initial interface of MCEE

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

Input data format (Figure 3)

- Acceptable file format: .xls.
- The first row of the table is the name of the metabolite
- In the sheet, each row represents a sample, and each column (except the first two or more columns) represents a metabolite.
- The first column of the data represents groups, such as the normal group and the disease group, which are expressed in 1 and 2 respectively.
- The second or more columns represent the confounding factor(s) and metabolites.

Group	Age	BMT	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19
1	55	16.22	659.36	15.61	132.87	18.40	240.78	738.02	437.27	84.89	1139.55	238.76	1711.98	1172.42	497.54	2.14	3373.30	73.04	124.43	660.37	15213.4
1	52	22.72	546.02	8.80	246.90	23.46	177.54	566.42	263.99	59.66	446.72	343.91	1908.50	1621.84	211.00	1.60	6284.20	88.84	178.36	929.59	10114.2
1	42	23.46	451.88	7.23	311.15	25.83	313.52	655.24	206.23	100.65	998.71	327.01	2236.63	507.33	394.86	0.22	4376.77	67.97	231.16	603.03	6375.2
1	42	23.05	532.73	7.89	89.42	11.32	202.15	373.07	142.32	85.27	726.64	204.64	1783.89	763.24	392.22	2.29	3200.00	39.17	199.44	609.81	13816.6
1	35	20.93	368.36	7.94	156.60	24.90	132.97	949.57	345.39	80.20	605.11	362.71	1555.24	1100.78	236.28	0.63	3307.06	32.53	39.28	204.98	7935.5
1	35	22.98	463.19	10.10	308.92	22.72	252.78	880.91	148.02	115.26	584.04	663.53	1599.95	415.95	321.51	1.59	7171.80	114.69	204.99	882.54	12299.4
1	34	24.61	468.25	9.97	227.20	33.63	456.53	751.58	262.25	120.35	1701.99	203.26	1591.55	832.11	215.57	0.51	3217.28	33.78	212.34	477.44	6123.5
1	33	22.22	356.37	7.68	203.30	26.66	288.63	1062.70	488.06	75.33	475.23	286.20	1926.14	805.17	564.82	0.56	4824.51	40.15	303.81	457.25	7180.6
1	32	21.85	345.66	8.44	120.35	44.29	164.96	435.34	336.20	93.69	1300.93	194.11	1650.19	1318.28	142.94	2.00	3273.19	24.06	173.90	809.13	6235.1
1	32	20.96	1004.61	11.08	224.17	6.96	439.33	1083.30	324.60	5.00	881.30	655.89	2321.05	1291.36	816.32	3.66	6564.38	35.28	146.85	821.56	27490.4
1	32	22.86	1310.03	17.47	400.78	33.43	298.82	1643.34	390.93	169.94	1369.37	531.87	2190.40	860.94	550.06	0.93	11337.87	29.20	411.32	832.02	17077.7
1	31	24.38	430.77	7.68	445.96	57.50	349.51	1225.90	504.98	71.44	1183.79	556.04	2295.80	904.08	271.47	0.81	7795.35	98.96	250.39	601.22	7088.6
1	29	21.15	470.65	15.06	237.07	24.53	248.94	807.29	236.17	88.99	525.85	417.17	2262.46	762.58	422.19	0.41	9120.78	31.53	313.11	946.97	8062.6
1	29	23.32	818.60	12.17	511.32	24.65	570.63	1352.21	331.70	93.03	3819.47	265.64	1841.64	2014.88	430.62	4.21	7217.58	71.94	445.79	829.34	21024.4
1	26	27.78	311.39	8.40	96.85	11.39	84.81	446.29	324.21	20.70	342.60	147.27	1103.75	842.54	173.09	54.25	3747.82	17.14	135.89	936.53	12923.2
1	26	21.95	1242.03	12.56	368.75	29.45	236.07	1216.22	503.99	368.73	1354.69	330.65	2749.87	1424.25	604.19	0.83	9718.42	57.38	737.12	2310.02	16697.7
1	25	21.78	553.15	10.83	344.85	40.06	216.53	1692.51	515.31	54.41	2187.50	226.29	2311.08	1120.58	270.63	0.25	4563.66	57.71	167.04	645.09	6090.4
1	24	23.89	474.56	17.00	314.72	35.57	830.12	1509.47	979.37	73.83	8322.42	313.43	1688.39	1338.73	260.68	1.16	3356.19	79.89	239.02	697.50	9232.1
1	24	22.64	438.79	8.07	160.50	14.56	81.43	480.99	313.47	52.11	1019.75	4.99	1400.05	759.30	184.85	2.08	3656.44	56.69	152.80	1046.56	11819.4
1	22	20.34	356.72	10.14	192.92	34.37	238.39	756.05	179.92	182.85	1031.29	417.10	1887.66	669.30	230.68	0.80	2873.67	50.25	283.73	1414.90	5494.0
1	21	21.22	325.84	8.59	129.64	29.21	198.29	514.83	399.83	78.70	968.88	401.95	1433.81	729.64	213.91	1.06	2461.76	40.61	177.04	813.24	5538.8
1	21	22.20	394.34	8.33	319.36	14.04	311.52	960.60	393.35	61.46	781.47	262.12	1571.40	822.05	415.29	8.07	3626.86	23.98	428.20	424.62	12498.7
1	21	19.16	902.58	27.67	262.70	52.05	395.45	1269.75	234.92	84.33	609.51	431.16	2415.87	598.92	541.98	0.52	13118.72	44.15	480.22	1469.76	14240.7
1	18	22.32	658.06	11.91	206.09	50.60	326.35	968.83	539.17	80.57	1275.71	284.65	1945.26	642.23	404.32	1.64	5133.84	113.89	609.31	848.77	20836.9
1	18	17.51	969.12	22.84	237.62	48.48	194.22	630.06	351.14	83.36	503.01	597.13	1488.00	372.33	467.05	0.82	6653.28	19.27	566.79	1481.40	26294.1
2	69	25.83	682.44	13.73	338.52	28.46	403.39	1417.81	262.33	72.14	8803.34	223.94	1645.02	1018.42	272.12	0.53	4488.96	44.34	243.90	883.83	6923.5
2	69	21.97	522.87	11.06	358.31	19.00	183.84	1445.13	412.26	80.40	1142.80	40.12	2450.28	1161.47	458.04	1.73	9446.58	76.78	422.25	682.09	14834.7
2	66	27.04	171.67	28.10	241.31	33.36	371.91	869.53	334.85	188.80	1050.00	312.85	2091.08	1045.24	236.39	1.02	6360.09	85.11	272.73	1116.77	14364.4
2	65	24.22	524.05	25.73	242.52	9.21	274.74	560.02	164.86	135.42	726.40	497.11	1887.02	492.54	522.76	4.12	4290.50	46.37	403.09	837.92	11916.1
1	65	26.78	851.38	25.98	354.55	16.50	94.19	741.12	522.85	120.35	586.44	374.32	2207.67	1046.78	597.65	1.70	7602.95	71.32	663.16	326.18	36441.1
2	64	24.49	992.71	19.97	309.99	24.67	209.16	1298.98	264.39	99.26	925.18	333.38	2203.82	1405.13	332.49	0.66	9387.28	38.77	489.44	1409.78	19434.3
2	63	25.47	554.43	14.57	305.09	27.77	274.01	978.69	290.04	101.98	925.18	279.05	2305.15	1046.78	360.31	1.21	7818.40	71.73	415.62	810.88	11935.5

Figure 3. Input data format

## 4.2.2 Specify the range of confounding factor(s)

The user inputs the range of start and end columns of confounding factor(s), specifying the scope of confounding variables.

## 4.2.3 Recommended parameters of MCEE-C

The recommended value of parameter  $ss$ , for MCEE-G, is 0.2. The optimum values of  $rr$  and  $kk$ , for MCEE-C, should be predicted according to the data set and user specified AUC value (default is 0.8).

Input expected AUC in the box, and click run. If the value of AUC is suitable, the recommend parameters will be calculated and filled into the boxes of “ $rr$ ” and “ $kk$ ” automatically, in about 1 minute. The processing window displays the success of recommended parameters calculation (Figure 4). If the value of AUC is not suitable, the processing window shows the failure information (Figure 5).

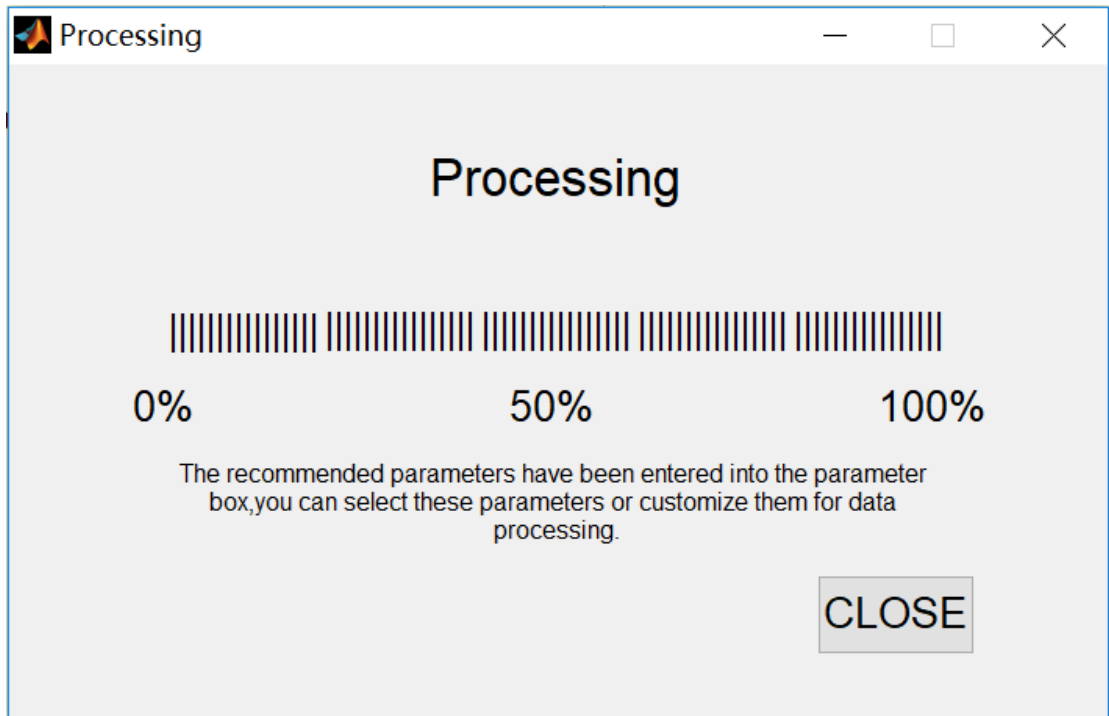


Figure 4. The processing window of finding the recommended parameters successfully.

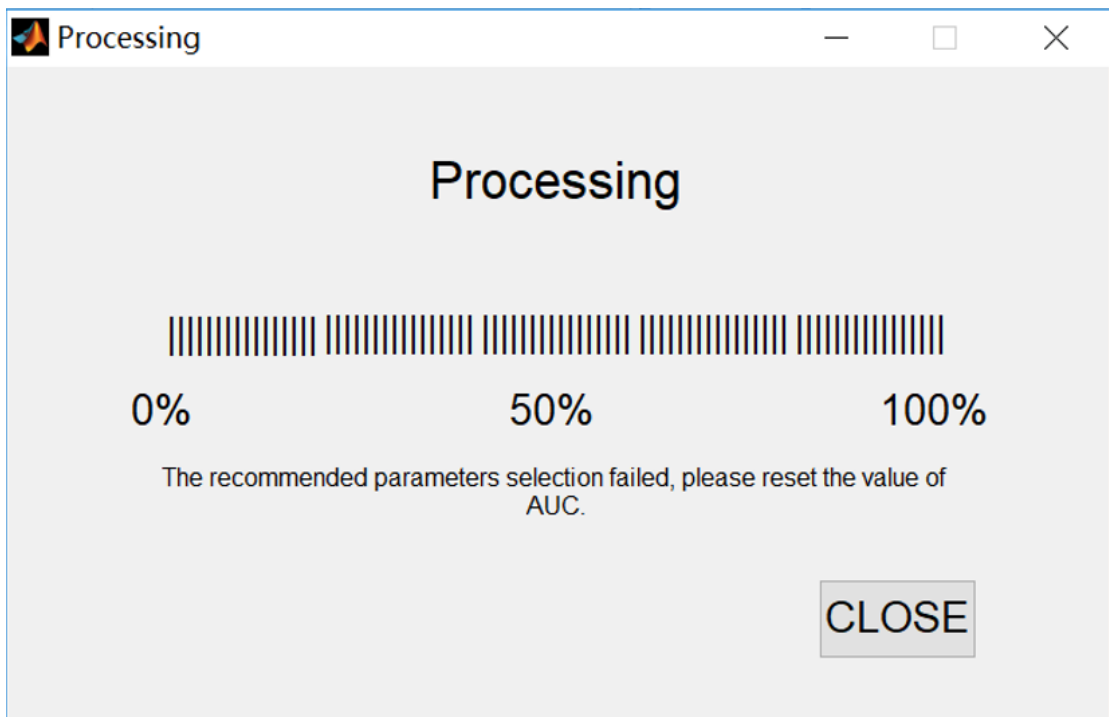


Figure 5. The processing window of finding the recommended parameters unsuccessfully.

#### 4.2.4 Parameter setting

According to the user's demands, select the appropriate method. If the user selects the MCEE -G method, the value of the adjustable parameter  $ss$  (0-1) (the

initial value is 0.2) shall be specified, and the ss value (0-1) shall be entered in the input box. If the user selects the MCEE-C method, users can choose to use the recommended parameters obtained in the previous step or specify the values of the adjustable parameter kk (0-1) and the adjustable parameter rr(0-1) by themselves.

#### 4.2.5 Program running

Click "RUN" to start the software, this time will pop up a Processing interface to display the running schedule. When the program is finished, the window will prompt "The result has been saved to the folder where the software is located." (Figure 6). The corrected data named MCEE.xls and saved to the folder. Then click "CLOSE" to close the interface.

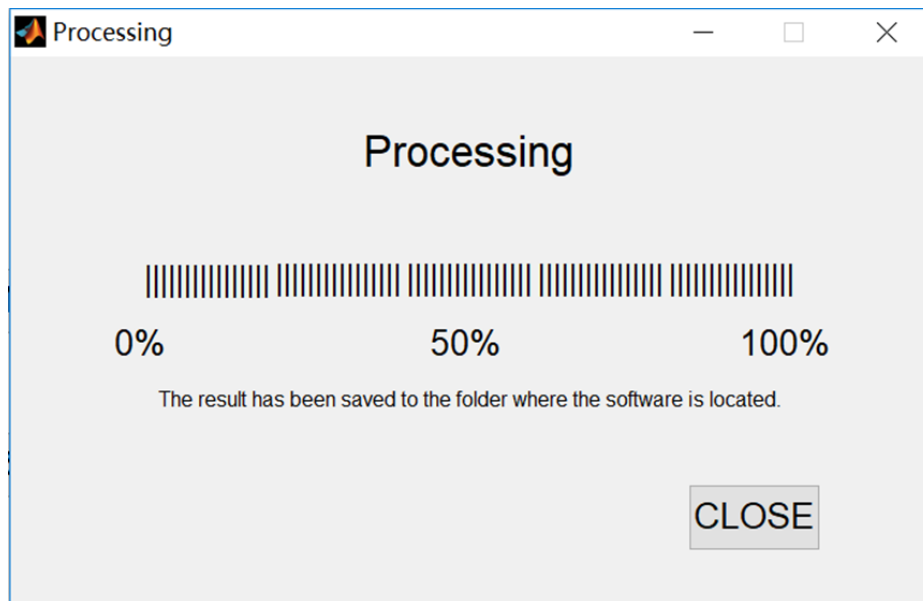


Figure 6. The interface of Processing

#### 4.2.6 Draw the PCA scores plots

When the program has finished, the type of PCA scores plots can be selected and drawn (Figure 6).

#### 4.2.7 Restore default parameters

Click the "RESET" button to restore the default parameters.

#### 4.2.8 Close the software

Click the "CLOSE" button to close the software.



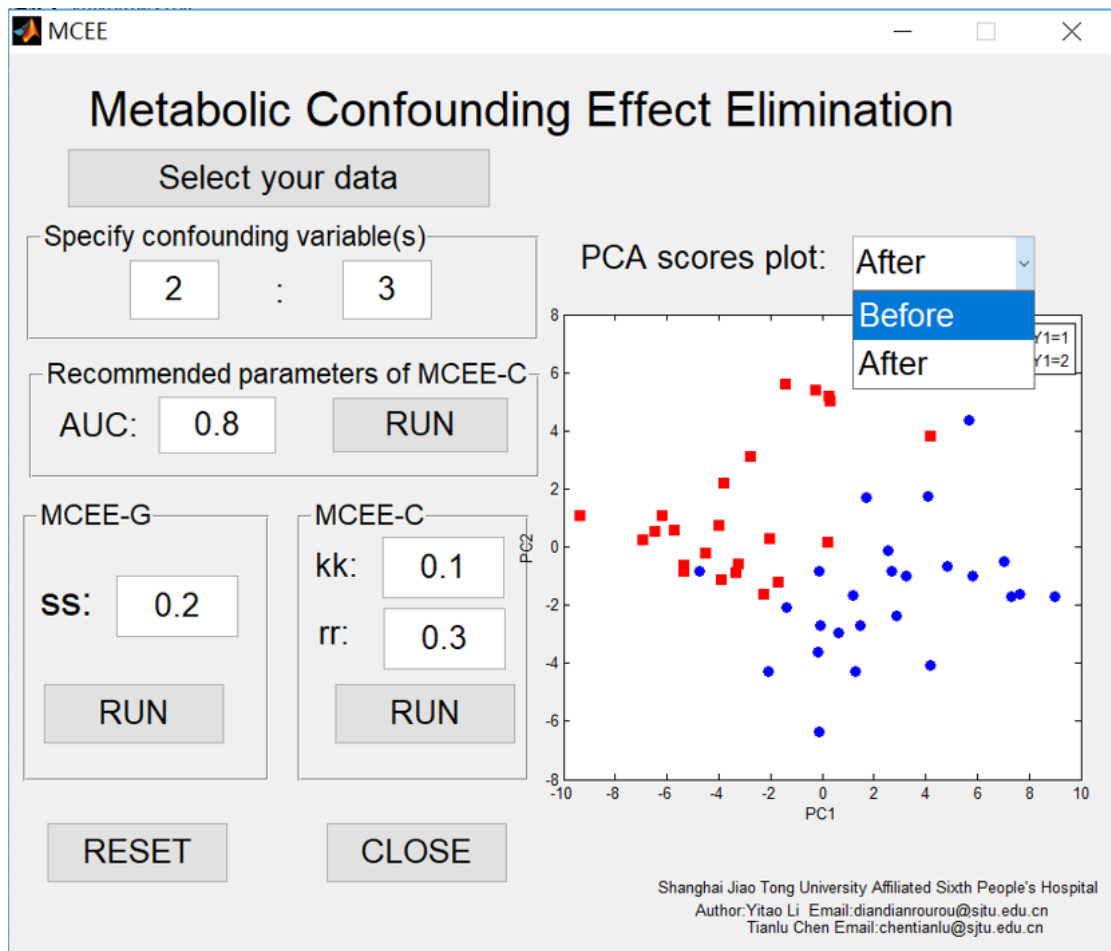


Figure 7. Drawing the PCA scores plots

### 4.3 Summary

This software is suitable for eliminating the bias caused by confounding factors.