

Calibration by QCs

Manual

2013-05-24

1. Introduction

Various approaches have been used to reduce systematic non-biological error which is frequently occurs within and between batches in metabolomics data sets. Direct scaling, using the total signal intensity, labeled internal standards for every analyte, or a representative Internal Standard, tends to suppress the sensitivity of the analyses and thus may lead to a loss of information. Reducing systematic nonbiological error by periodically analyzing pooled samples (PQCs) along with the subject samples is gaining acceptance as a quality control strategy in metabolic profiling. Despite their success, PQC-based calibration methods have a limited capacity to adjust batch and injection order effects lacking the pretreatment of the PQC data beforehand. Additionally, the calibration process is laborious and prone to human error. Therefore, we developed a seven-step workflow and software to calibrate metabolomics data sets by pooled samples as quality controls (PQCs). The software can adjust both within- and between-batch variance and perform rough, medium, and precise calibrations. Calibrated data sets derived from various platforms can be combined directly. It is user friendly with an integrated GUI, has lots of options, and is independent to data sources.

2. Operation procedure

1) Startup

In the “current directory” panel of Matlab environment (7.1.0246+SP3), right click the file “calibration.m” and select run (Fig. 1). The main window will appear (Fig. 2).

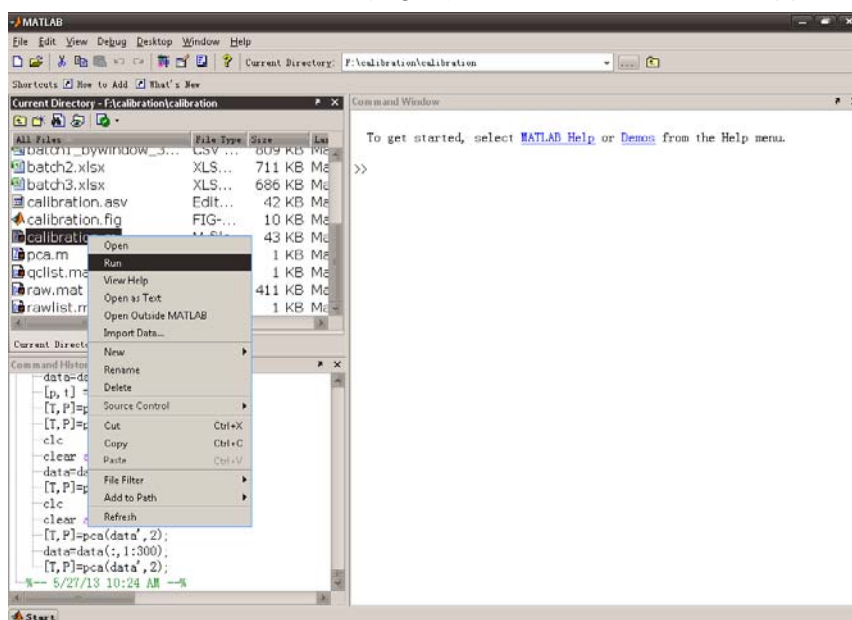


Fig1. Software startup

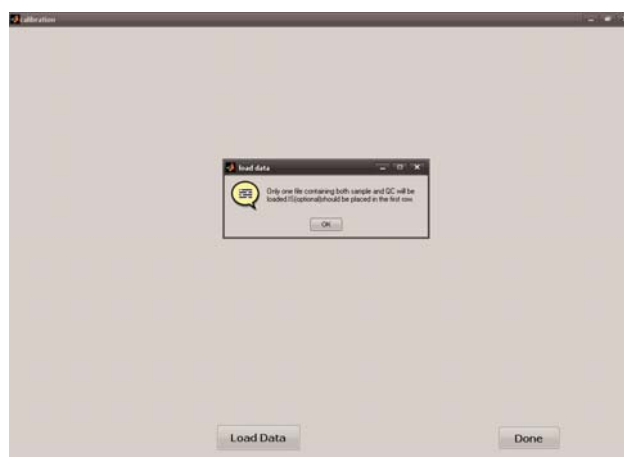


Fig. 2 The initial window

2) Import dataset

Dataset with both samples and QCs (quality controls) is required. QCs should be pooled samples. They contain the same compounds as the subject samples and are supposed to reflect the average metabolite concentrations within a study. QCs are pretreated according to the same protocols as the subject samples and are evenly injected throughout the analyses. The performances of the pretreatment and the

analytical platform can be assessed using the QCs.

Click the button “load data” to select a file for processing. Acceptable file formats are .xls, .xlsx, and .csv.

Each row is a variable (metabolite/peak) and each column is a sample or QC. Rank all the samples and QCs respectively, as instrument detection order. Samples at first. The first row and first column are supposed to be names of variables or samples (Fig. 3).

Compound	S1	S2	S3	S4	QC1	QC2	QC3	QC4	QC5	QC6	QC7	QC8	QC9	QC10
1. Compound														
2. 2-chlorophenylalanine(13)	2478106	1202614	641913	212846.2	2456.373	463896.8	188830.3	777.0322	46276.78	11591.13	977			
3. 2-Hydroxyproline	294610	1147807	590952.8	211703.1	1554.064	246843.1	206553.7	297.3393	40328.78	11443.13	976			
4. Benzenesulfonamide	2952268	1040298	541685.8	50745.57	2403.333	294393.6	176853.2	116.642	41377.8	3247.892	433			
5. Phosphate	3501456	1193730	590395.9	8802.522	3253.872	256483.5	263597.6	1437.525	56185.43	40758.74	15			
6. 2,3-Dihydroxybutanoic acid	2974405	1151680	608467.5	283838.5	6200.13	378372.9	243274.1	304.5301	60083.08	4352.29	482			
7. Threonine	3463439	1224029	734827	247189.6	5634.521	403208	379679.7	356.4437	77943.96	12008.97	8			
8. Methionine	2620688	1237143	623274.8	704650.35	2464.465	394885.6	160361.3	381.4905	44020.26	12443.87	579			
9. D-Ascorbic acid	3080931	118941	778732.5	62437.58	2077.834	297171.9	161099.5	977.337	44458.12	13352.04	943			
10. L-Histidine	2844053	1150858	576925.7	83323.48	1706.844	176634.6	193659.1	123.4793	50056.2	17724.82	677			
11. D-Glutamine	3284508	1009703	595544	224917.2	3131.471	377004.5	196169.2	1552.076	52474.01	11023.4	813			
12. Cysteine	3474761	1297766	765473.7	191613.2	2060.169	330218.9	197006.5	25.78607	52923.61	12360.87	15			
13. Glutamic acid	4189312	1134228	805050	104962	231.7189	340524.4	310031.5	1635.313	65826.36	6297.468	7			
14. Phenylalanine	2964497	1073492	624015.8	170484.1	2131.816	330382.1	1099235		149332.43	5651.582	341			
15. D-Xylose	3334538	1241152	837107.7	184980.3	4028.963	423965.7	302651.2	305.7908	65095.1	11425.38	534			
16. L-Glutamine	3315564	1182225	678890.5	335191.4	4890.125	411800.5	892883.7	423.2839	47563.64	6132.944	676			
17. Hypoxanthine	3012032	1121897	750743.5	144073.5	636.3862	381095.6	1466663	514.5351	59046.28	6757.299	6			
18. Ornithine	3424377	1227559	816208.9	189785.2	2591.601	610482.6	268246.5	187.1955	70726.52	118186.7	17			
19. Citric acid	3045478	1149236	681376.4	277199.3	1864.428	286047.2	1538923	1722.966	54848.35	11328.29	739			
20. D-Erythritol	3461056	1118642	731864.2	381144.7	5601.036	549725.2	232413.5	2683.254	52098.41	17328.24	14			
21. Fructose	3320807	1130953	635015.2	255234.5	3474.387	331958.6	558608.3	1908.799	66602.3	18231.41	938			
22. L-Lysine	3129840	1214263	623492.4	90141.85	3884.541	267836.8	267447.1	1259.8194	61927.61	12274.79	702			
23. D-Gluconic acid	3321146	1044070	640275.5	244205.3	1962.019	378679.6	475325.9	2025.631	76150.09	9805.23	613			
24. L-Tyrosine	3526140	1091690	713534.2	129801.5	2827.625	674580.3	452014.5	1049.625	102279.7	5152.997	616			
25. D-Mannopentose	3677429	1033994	786496.7	147728.2	3244.325	379910.8	374089.1	1264.4794	92331.09	28107.38	12			
26. Palmitic acid	3887980	1209904	898857.7	135446.2	4517.146	687324.1	276322.7	258.7643	65826.67	20218.69	10			
27. Myoinositol	3860521	1139944	867074.5	558482.7	4539.887	774943.9	629137	1223.924	143294.2	18208.94	11			

Fig. 3 format of dataset

3) Set up parameters

In the “Parameter Setting” area, change the parameters manually or as it is (Fig.4).

Click “OK/Initialization” to continue.

Number of QCs: number of QCs

Number of Samples: number of samples

QC average: QC average mode (batch or window)

Batch/Window Size: size of QC average

Delete ratio (%) of RSD (QC): an RSD threshold. Variables with RSD>threshold will be removed.

Delete ratio (%) of zero (QC): a threshold for percentage of zero values. Variables with zero values>threshold will be removed.

IS Calibration: calibrate by IS

QC Calibration: calibrate by QCs

Z-Score: Z-Score transformation

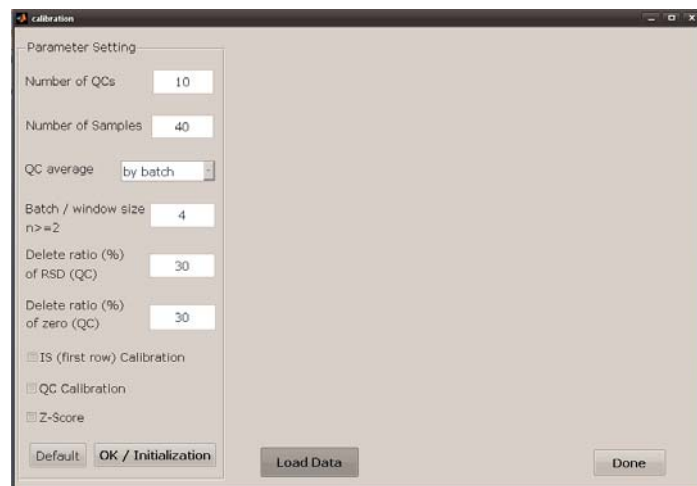


Fig. 4 set up parameters

4) QC outlier selection

After parameters setting, the “QC Outlier Selection” panel will be visible (Fig. 5). Select and exclude QC outliers manually from the “QC List” with the help of various PCA scores plots. The plots will be saved as figures(.tiff, .ai, .fig) in specified folders. “Auto Selection” mode is recommended which will exclude QCs deviated from $3 \times \text{SD}$ in the PCA scores plot derived from all QCs (Fig. 6). Click “OK” and continue.

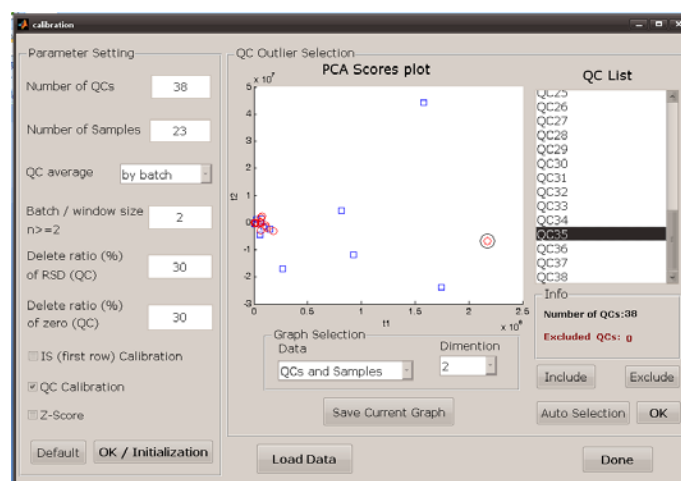


Fig. 5 QC outlier selection panel

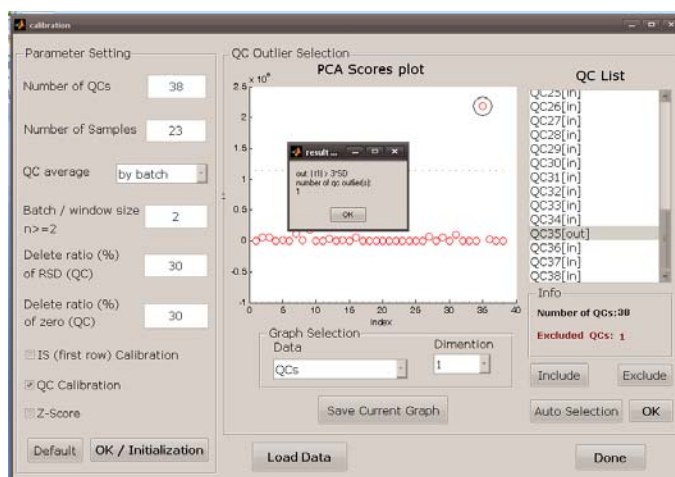


Fig. 6 Autoselection of QC outliers (QC35)

5) Calibrate the data

Click “Run and save result” and the software will calibrate the data as the parameters (Fig. 7). The workflow is shown as Fig. 8.

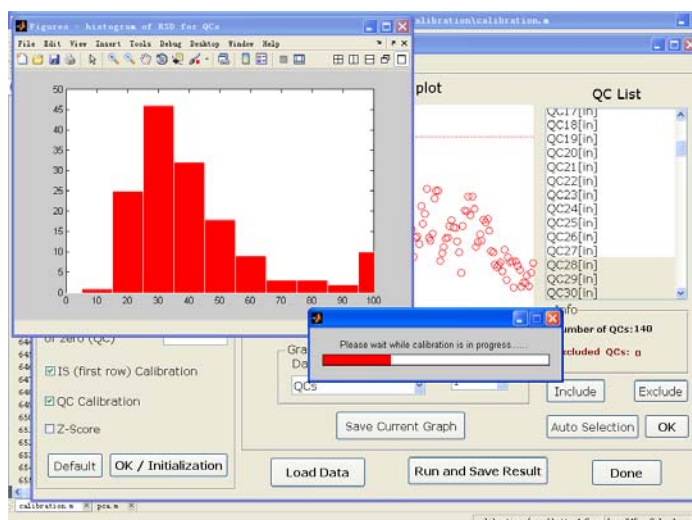


Fig. 7 Calibration in progress

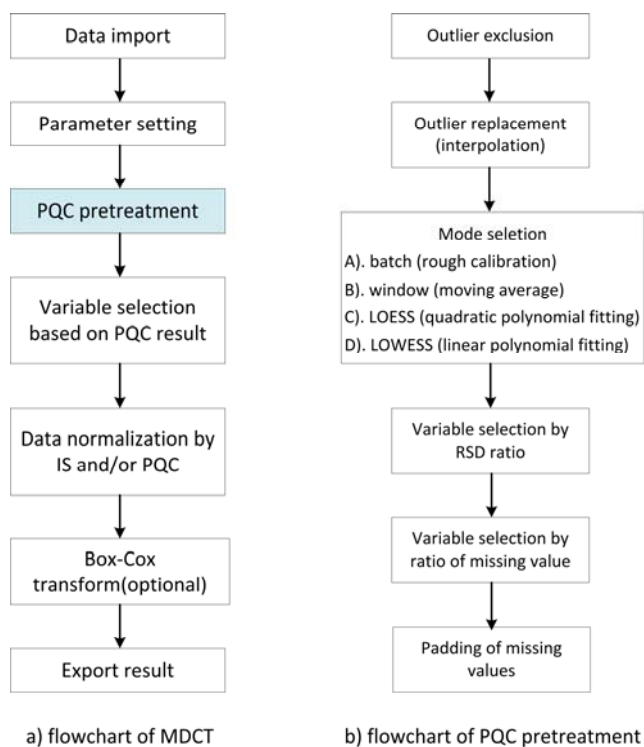


Fig. 8 workflow of calibration

6) Result output

The dataset after calibration (a .xls or.csv file) will be saved specified folder. Some relevant information will be displayed (Fig. 9).

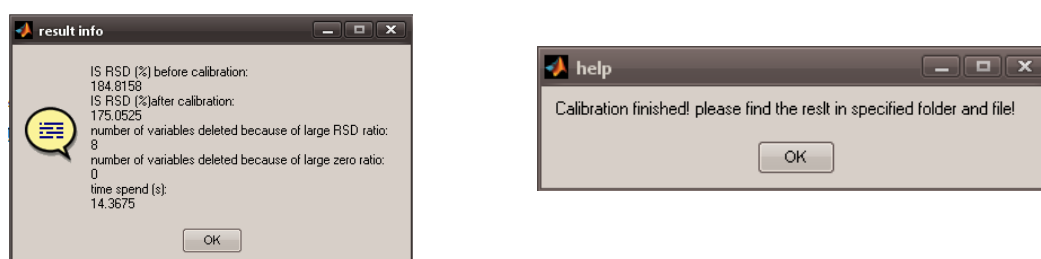


Fig. 9 Relevant information of the calibration

7) Re-calibration

Click “Done” to close the window. Or, change the parameters and click “Run and save result” button to calibrate the dataset once again.