

Automated Parsing of a LabSolutions Batch Results File (ASCII Output) When Using a Spreadsheet or Statistical Package to Summarize Data

1. Introduction

Long retention times (more than 30 minutes) are common when analyzing samples using traditional HPLC methodology. Long retention times combined with large sample batches (many samples) can result in long analysis times. Baseline drift and changes in detector response during long runs can cause significant deviations in calculated concentrations over the course of a single run. Cutting large sample sets into smaller ones can alleviate these issues, but also adds more preparatory work for investigators. Another common solution is bracketing samples with standard sets, allowing for large sample sets to be analyzed while intermittently running standards and creating new, updated calibration curves that more realistically reflect the current instrumental conditions and detector response.

It is sometimes desired to use a spreadsheet or statistical software package to summarize or further analyze data, and this can be challenging when using batch analysis in LabSolutions because the software outputs many fields of data that may not be needed and can only be pared down in broad categories. This note highlights a tool that can be used to automatically parse and summarize the data exported during batch analysis in LabSolutions.

2. Realtime Analysis Batch Processing

Each injection results in the creation of its own data file (*.lcd), which contains sample and peak information obtained using the peak identifying parameters provided by the method file and which are used to calculate and assign concentrations for each analyte. These are the concentrations that are reported in Postrun when you open a data file. Upon initial creation of the data file, the concentration field will be empty.

LabSolutions has a built-in feature that allows samples to be bracketed between standards and have multiple calibrations occur over the run. For example, a batch file can be made to have the following structure:

- Standard Group 1
- Unknown Group 1
- Standard Group 2
- Unknown Group 2
- Standard Group 3
- Unknown Group 3
- Standard Group 4

LabSolutions includes three different types of bracketing, described below.

Overlap Bracketing

This is the most common application of standard bracketing. Each set of unknowns is quantitated using a calibration curve created from two sets of standards - the standards both preceding and following the group of unknowns. It requires that the following standard injections be completed before the unknowns are quantitated. The following standards from the first series of unknowns become the preceding standards for the second group of unknowns, and the preceding standards from the first set of unknowns are replaced with the standards following the second set of unknowns. In the structure outlined above, Unknown Group 1 would be quantitated using Standard Groups 1 & 2, and Unknown Group 2 would be quantitated using Standard Groups 2 & 3. This process of removing the earliest calibration points from the calibration curve and adding the most recent calibration points is repeated until the end of the batch. At the end of the batch the method (*.lcm) includes the calibration curve created from the last two Standard Groups.

Sequence Bracketing

All unknowns are quantitated using a calibration curve created from all of the standards, and no concentrations will be assigned to unknown samples until data has been acquired and integrated from all standard samples in the batch file. At the end of the batch, the method (*.lcm) includes a calibration curve created from all standard injections.

Average Bracketing

Unknowns are quantitated using calibration curves based on all standard samples that have been acquired prior to the unknown samples, as well as the Standard Group directly following it. In the structure outlined above, Unknown Group 1 would be quantitated using Standard Groups 1 & 2, and Unknown Group 2 will be quantitated using Standard Groups 1, 2, & 3.

3. Data Processing in Postrun

When opening a data file in Postrun, the displayed concentrations and calibration curve are those that are saved to that particular data file. The method, including the calibration curve, as it existed when the file was last processed is saved as part of the data file. The current method file, including the current calibration, is not referenced. This means that the reported concentrations on an unaltered or unmodified data file will reflect the values calculated during a batch analysis, and will be correct as the appropriate calibration groups were used to produce those concentrations. If it is desired to use a spreadsheet for additional data processing or summary, an entire batch can be easily exported as a delimited text file, but the data is extensive and can be difficult to parse.

4. Data Processing in Quant Browser

Quant Browser can be used to provide a concise summary of data files, displaying only the fields required by the analyst, by adding individual data files or opening a batch file (*.lcb). The method of the first data file, as well as the calibration curve as it exists in the current version of that method, is applied to quantitate the unknowns, so bracketing is not applied. It does, however, offer the opportunity to quickly apply various methods to a set of unknowns simply by opening a new method. If it is desired to use a spreadsheet for additional data processing or summary, the quantitative results view can be exported as a delimited text file to provide a concise data set, but the results can differ from those processed using batch processing in Postrun because bracketing has not been applied. It is therefore not recommended for summarizing batch data where overlap or average bracketing has been used.

This process is not an issue with:

- Data files or batch files that have 1 Standard Group
- Sequence Bracketed datasets, as the sequence bracketing method does not result in periodic changes to the calibration curve
- An overlap or average bracketed dataset that only has one sample group and two Standard Groups.

It does, however, cause issues where you have overlap or average bracketing calibration sets for more than one group of unknowns. For example, in the case of overlap bracketing, correct concentrations are reported for the final Unknown Group (group 3 in the structure above) but the concentrations reported for Unknown Groups 1 and 2 are now based on Standard Groups 3 and 4 instead of the correct Standard Groups. Depending on the analysis type and the amount of deviation in instrument response over the course of a run, this can negatively affect the results a little or can have major implications on the accuracy of the reported values.

5. Resolutions using the tools available in LabSolutions

One option to obtain the correct results is the use of summary reports. A problem posed by this solution is that the output format is a .pdf, which is difficult to import into Excel or statistical packages such as R, SAS, and SPSS. A more suitable solution is to modify the batch file options to output the calculated values to an ASCII file by selecting "output each batch" under the ASCII Conversion Tab in the batch file options. The result is a long tab or space delimited file that contains the output from each injection sequentially. It is up to the investigator to then parse the

resulting file for their values. Performed by hand, this task is both time consuming and tedious, providing ample opportunity for human error.

It should be noted that the database and client/server versions of LabSolutions include a feature called Multi-Data Report. This feature automatically carries out all of the manipulation described above and provides a nice, easily readable and usable output for all samples in a batch.

6. Resolution using a VBA Script to parse the text file created during batch analysis

A VBA script has been written to automate the process of cleaning up and summarizing the output file (ASCII) from a batch in LabSolutions. At the time of creation of this Technical Note, the summary that is produced includes the Sample Name, Sample ID, and concentrations of each component identified in the Compound table. Also included is the ability to process many batch output files at once, creating one large table for all samples in all batches. Note, however, that all batches must use methods that contain the same compounds listed in the same order.

As the script results in concentration values being taken directly from an ASCII file created during the batch analysis, it will work with any of the 3 bracketing methods, as well as any quantitative method type available within LabSolutions (for example, External Standard, Internal Standard, or Standard Addition).

The script requires a Microsoft Excel workbook with two worksheets. On the first worksheet, it imports the batch ASCII output file. If multiple files are in the selected directory, it will append the data from each file to the first worksheet. It then extracts Sample Name, Sample ID and the concentrations of each identified component and creates a simple table on the second worksheet with just that information. An example of the output on the second worksheet is provided below, and a complete example of both input and output data is provided at the link at the end of the note.

SampleName	SampleID	Na+	NH ₄ ⁺ -N	K+	Mg ²⁺	Ca ²⁺
singleinjection	1	0	0	0	0	0
singleinjection	2	0	0	0	0	0
singleinjection	3	0	0	0	0	0
Cation Std Set	1	0	0	0	0	0
Cation Std Set	2	0.103	0.156	0.626	0	0
Cation Std Set	3	2.056	3.125	12.512	1.251	1.25
Cation Std Set	4	4.124	6.233	24.999	2.503	2.5

7. Summary/Discussion

While there are several ways to export data in LabSolutions so that it can be further processed or summarized in a spreadsheet or statistical software package, the resulting data set can be either difficult to parse in the best case or, in the worst case, erroneous if using Quant Browser when standard bracketing was intended. In this note, we have identified the reasons why Quant Browser should not be used when bracketing standards, and we have created and shared a tool for automatically parsing the data exported from batch processing.

This tool is verified to work in LabSolutions Version 5.54 SP5 or earlier, although it should continue to work in future versions as well. It is provided free of charge and can be freely modified or extended by any interested party. Extensions to the code might include adding concentration units and additional fields such as retention time, area, and column performance parameters such as peak tailing, as well as the ability to select for only unknowns and optimizing the user interface.

The solution, including the script, instructions for use, and supporting materials is available at:

<https://github.com/saleslab/ParsingLabSolutionsASCIIOutp>

This technical note was written by Jacob R. Price of Drexel University and edited by Jeff Parish, Shimadzu Technical Support. The referenced VBA script and associated documentation was written by Thomas J. Thompson of Drexel University.



First Edition: January 2016

SHIMADZU Corporation
www.shimadzu.com/an/

SHIMADZU SCIENTIFIC INSTRUMENTS
7102 Riverwood Drive, Columbia, MD 21046, USA
Phone: 800-477-1227/410-381-1227, Fax: 410-381-1222
URL: www.ssi.shimadzu.com

For Research Use Only. Not for use in diagnostic procedures.
The contents of this publication are provided to you "as is" without warranty of any kind, and are subject to change without notice. Shimadzu does not assume any responsibility or liability for any damage, whether direct or indirect, relating to the use of this publication.

© Shimadzu Corporation, 2016