SAND2014-15965D

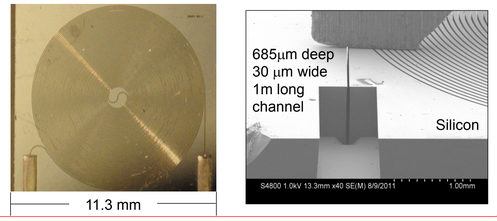
Comprehensive Automation of Data Analysis for High-Speed Micro Gas Chromatography

Nathaniel. B. Pfeifer
Biosensors and Nanomaterials (1714)
Sandia Staff Member Advisor: Douglas H. Read, Microsystems Enabled Detection (1716)

Introduction

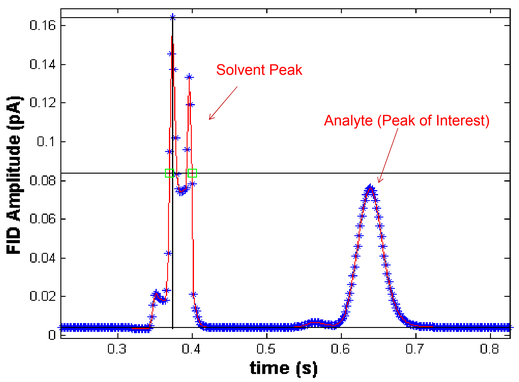
High-Speed Gas Chromatography (HSGC) is an analytical method used to detect and discriminate individual chemicals in a mixture very quickly (< 1 min.). Microscale gas chromatography columns are fabricated and polymer-coated in house for this application. The objective of this project is to automate the labor-intensive analysis of raw GC data to produce the column performance metric known as the van Deemter plot. We have developed MATLAB code and an advanced graphical user interface (GUI) that performs this multi-level data reduction and analysis.

Microfabricated GC column (μGC)

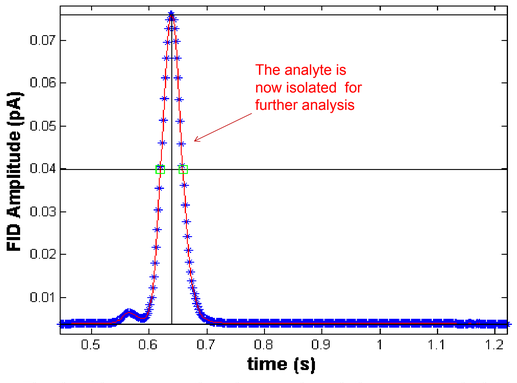


Solvent Peak Remover

Sample Chromatogram



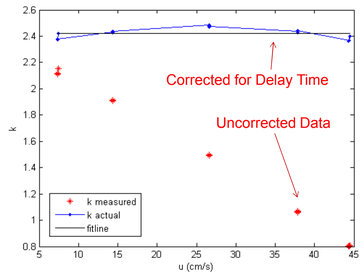
- The analyte and solvent are separated by the GC
- The solvent peak is unneeded and prevents the algorithm from analyzing the correct peak
- The peak remover function deletes the solvent peak



The algorithm removes the solvent peak, and the program calculates the correct values:

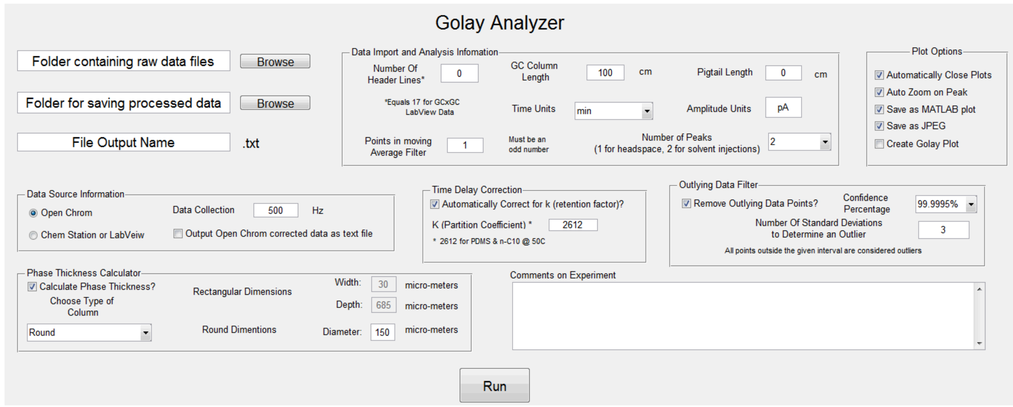
- Retention time (t_R)
- Height Equivalent to a Theoretical Plate (HETP)
- Number of plates
- Full Width Half Maximum (FWHM)
- Peak Area

Delay Time Correction



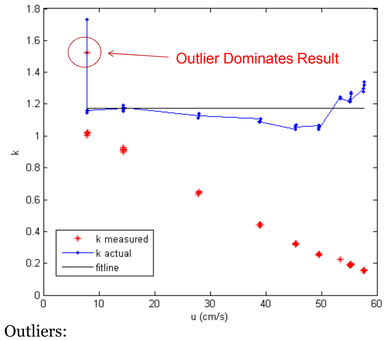
- The timing error in the instrument causes an unphysical result in the retention factor , k
- Delay time algorithm finds the instrument delay time by forcing the slope of $k(u)$ to zero (See *Theory*)

GUI



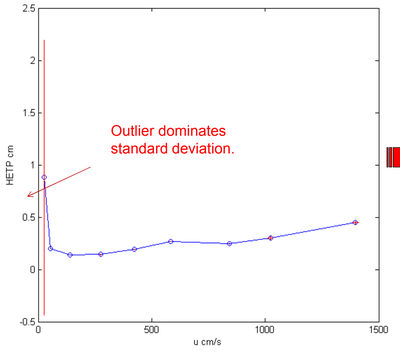
Our program uses a graphical user interface to simplify its operation. The user can customize the way that most operations are performed and bypass others if desired.

Unfiltered Data



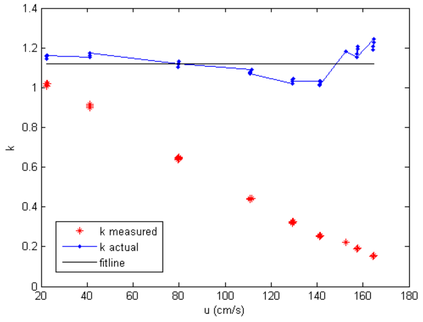
Outliers:

- Produce an unphysical result
- Introduce error into future calculations
- Take time to remove manually

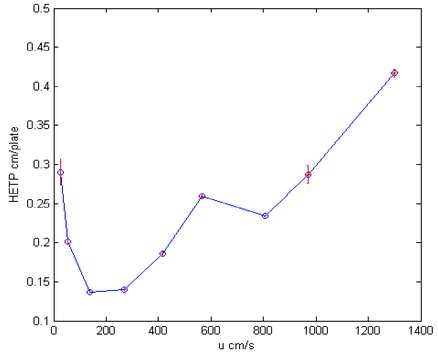


- Excessive deviation makes the first data point unreliable
- Actual trends within the data are hidden because of scale

Filtered Data



- Filtered data represent the trends set by the entire data set
- The values calculated later are more accurate



- The HETP data can now be interpreted easily and accurately

All points removed by the filter are recorded. It is the operator's responsibility to inspect all data points that were thrown out to ensure they are true outliers and not statistically significant points.

Theory

Outlier Filter

The data is filtered for outliers using a Modified Thompson Tau test to identify statistically outlying data points. The equation for τ is given in Eq. 1.

$$\tau = \frac{t\alpha_{/2}(n-1)}{\sqrt{n}\sqrt{n-2+t\alpha_{/2}{}^2}} \qquad Eq.1$$ Ref.1

The distance tau from the bi-weight mean is calculated using the critical Student's t value ($t\alpha_{/2}$) and the number of data points in the set (n). If the furthest point from the mean is further than $\pm\tau$ away, then it is labeled an outlier and discarded. The mean is then recalculated and retested until no outliers are identified.

GC-Specific Delay Time (Instrument timing error)

Time delays in the instrument data acquisition create an error in each time series. This leads to errors in GC performance and characterization metrics. For example the retention factor, k, is corrected using Eq.2. Time delay, t_{delay}, is iterated until k is a constant value (which matches theory). This delay is then subsequently removed from each recorded time.

$$k = \frac{t_R - t_M}{t_M - t_{delay}} = \frac{2 \cdot K \cdot d_f}{r_{col}} \qquad Eq.2$$ Ref.2

t_{delay}	➡ Time delay (instrument DAQ error)	d_f	➡ Stationary Phase Polymer Thickness
t_R	➡ Retention time of analyte	r_{col}	➡ Radius of GC column
t_M	➡ Retention time of unretained	k	➡ Partition Coefficient
k	➡ Retention factor		

This data processing program has cut analysis time from 3 hours per experiment to 5 minutes! Many more data points can also be added for better statistical analysis with no additional time cost to the researcher! **BETTER DATA, LESS TIME & MONEY.**

References
[1]: Cimbala, John M. September 2011 https://www.mne.psu.edu/me345/Lectures/outliers.pdf
[2]:Grob, Robert L. and Eugene F. Barry *Modern Practice of Gas Chromatography* 4th ed. 2004 John Wiley & Sons

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

The project is sponsored in whole by the Department of the Defense, Defense Threat Reduction Agency (DTRA).