Poster Reprint

ASMS 2023
Poster number TP 062

AI Peak Integration for MassHunter software automates manual peak integration during the data analysis process in GC/SQ

Thomas Bispham, Tamas King, Winnie Chau, and Ruoji Luo

Agilent Technologies, Inc.

Introduction

GC/MS data analysis is often manual, time consuming, and error prone (see Fig 1). Even though many automated integrators are available, a chemist may still need to manually correct integration peaks and baselines. A new quantitative analysis software automates peak integration during the data analysis process for analytical testing laboratories. A Machine Learning (ML) model is custom-trained through a user's standardized data analysis workflow, observing manual peak integration events. Machine Learning then replaces manual peak integration with adaptable artificial intelligence (AI)-assisted peak detection and integration with continuous learning.

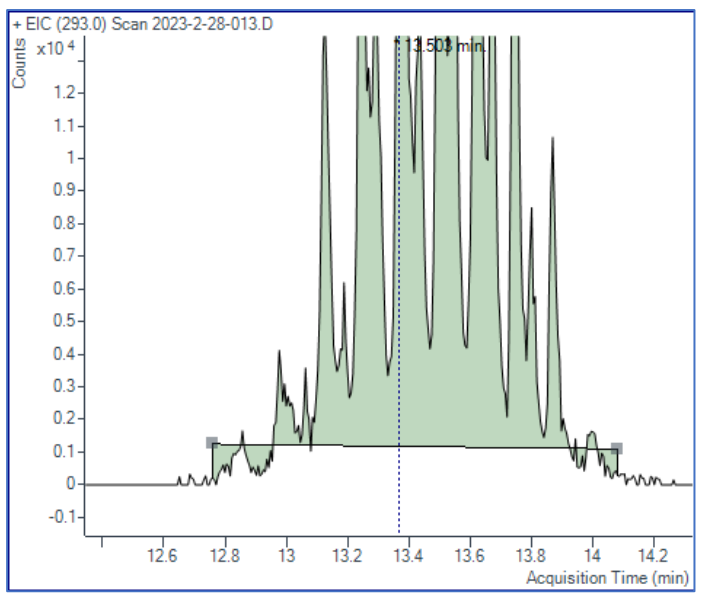


Figure 1: Manual integration often results in an inconsistent baseline, even for an expert chemist.

The novel peak integration software helps improve process automation, maximize sample throughput, and produce consistent and high-quality data while facilitating ease of adoption for laboratories.

Phthalate compounds are described as an "everywhere chemical" around the world [1], as they are heavily used in packaging and plastic materials.[2,3] Manual integration of the compounds' high molecular weight proves difficult, even for an experienced analyst (see Fig 2).

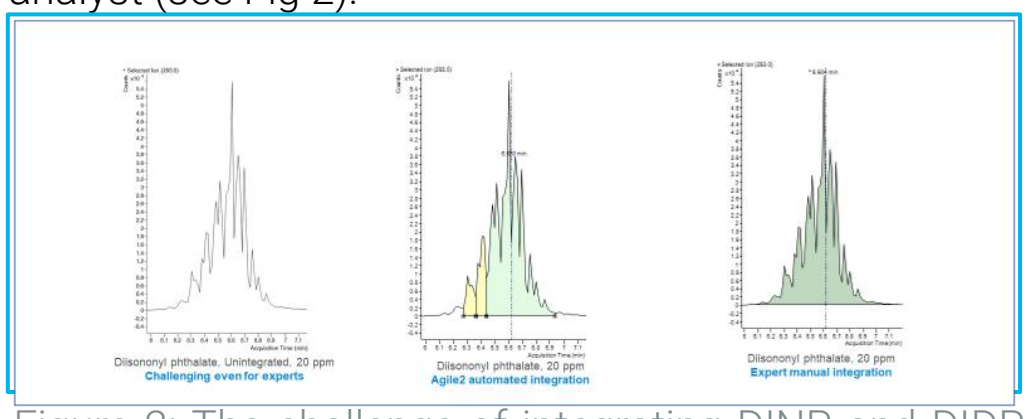


Figure 2: The challenge of integrating DINP and DIDP in phthalates with single quad GC/MS – DINP in Phthalates have GC/MS peaks that are partially co-eluted (left), DINP Peak integrated by automated integrators (center) and DINP Peak manually integrated by an expert chemist (right).

Experimental

Equipment and Software

Agilent GC/MS (Single Quad) instruments were used to acquire samples. Agilent MassHunter (Version 10.2) software workstation was used for data acquisition and quantitation.

An add-on to MassHunter Quant was setup to communicate to the AI Peak Integration database, which was cloud-based. MassHunter Quant functions with a default parameter-less integrator, Agile2, to analyze sample data. This was used as a basis for the Machine Learning model.

Modeling and Data

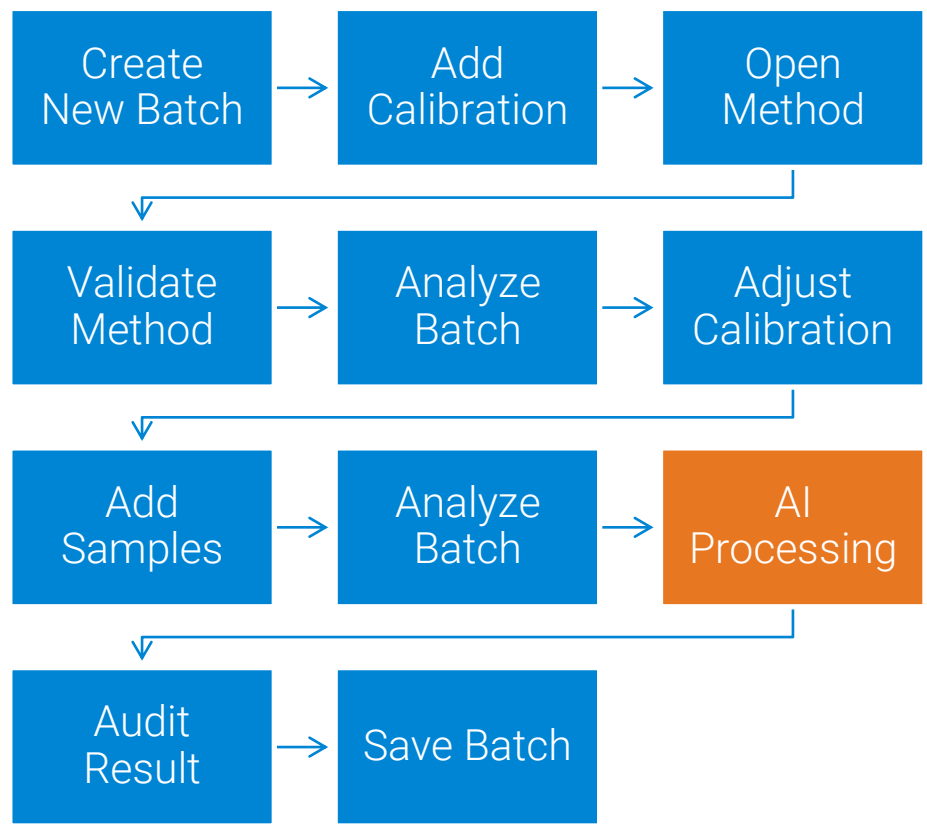
A Machine Learning model was developed for the analysis of phthalate compounds in consumer products. Data from 1247 real-world consumer samples were used to "train" a Machine Learning model in phthalate detection and quantitation. The sample set contained 800-1000 positively-identified phthalate peaks.

Data Info

Targeted Compounds	14 phthalates, 1 Internal Standard (ISTD)
Target Acquisition	SIM/Scan
Default Automatic Integrator	MassHunter Agile2 (parameter-less)

Data Analysis Workflow

Samples and Method are combined into a MassHunter Batch. A chemist analyzes a batch in a consistent sequence, but manual integration is replaced by AI Peak Integration processing.



Speed

Under ideal conditions with an expert analyst, reviewing data and performing manual integration requires an average of 60-120 seconds per chromatogram.

The AI Peak Integration solution outpaced manual integration after just a few positive samples. The minimum cloud processing time was about 30 seconds.

In preliminary testing, AI Peak Integration software performed peak integration on 100 samples in fewer than 25 minutes, where an expert chemist (under the ideal conditions) would take 2 hours to manually integrate peaks in the same batch, thus enabling at least a **4-times gain in productivity** (see table).

Phthalate Samples	Manual chemist integration estimate (hours)	AI Peak Integration via Cloud estimate (hours)
25	0.44	0.10
50	0.88	0.19
100	1.76	0.39
1000	17.6	3.9
10000	176	39

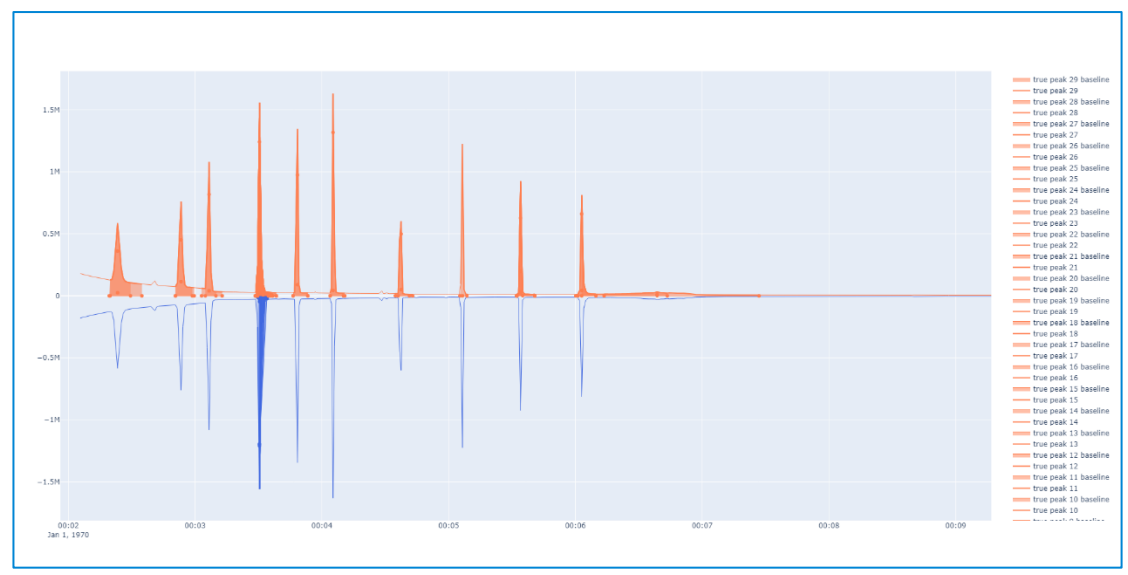
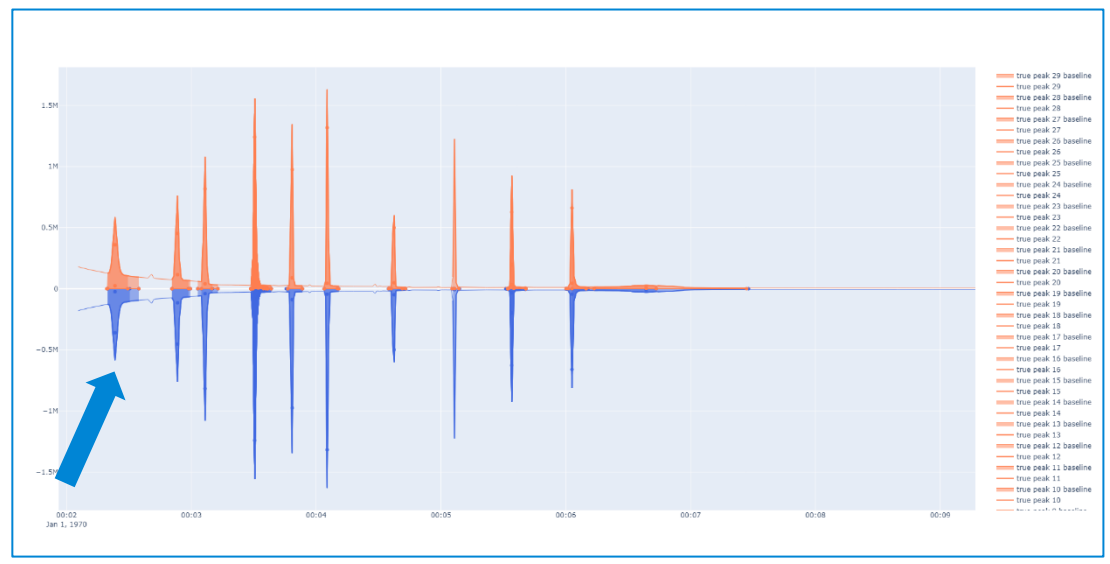


Figure 3: Comparison of a data-deficient Machine Learning model (above) and a sufficiently-trained model (below) with numerous positive samples.

Solid orange peak is "truth", based on chemist manual integration.

Blue is predicted integration and identification of substance by AI Peak Integration.



In addition to speed, using a sufficiently-trained Machine Learning model improves integration quality when compared to the default integrator or manual integration.

Accuracy

Machine Learning model accuracy metrics improve with more sample submissions and more diverse data. Prediction of less-common phthalate compounds was shown to improve over time (see Figure 4).

Data randomization was used to remove inherent bias when a ML model is trained on new or supplemental data.

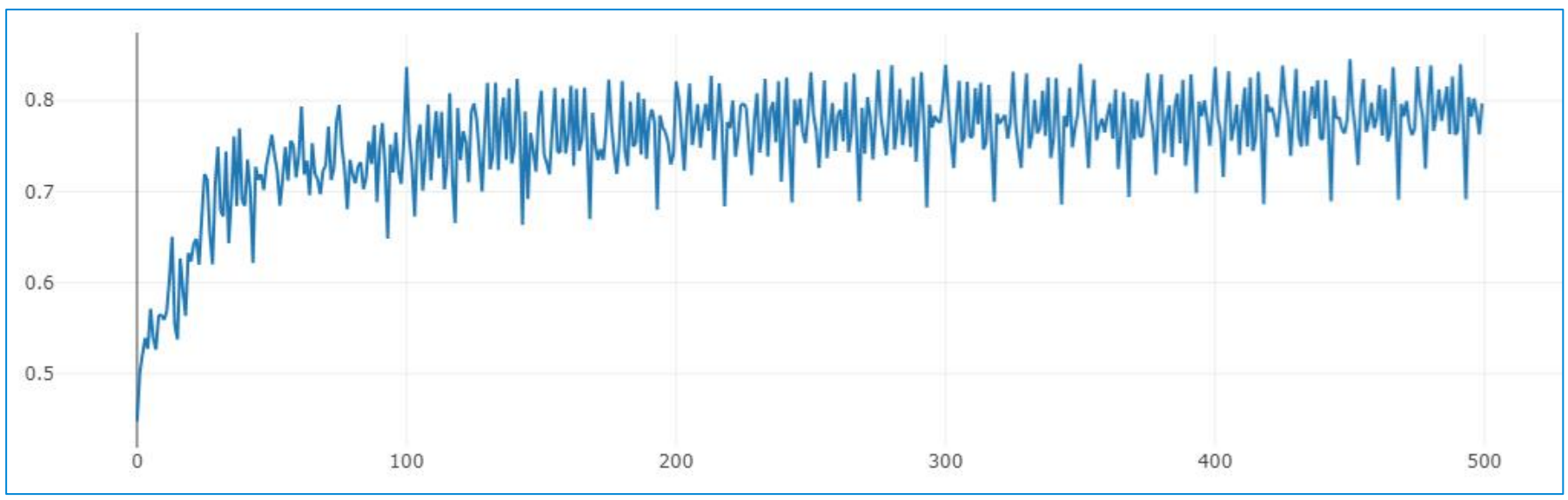


Figure 4: Critical Success Index (maximum of 1) of ML model measured on a held-out data as training progresses.

Accuracy (continued)

When comparing a typical chemist's manual integration of peaks and baselines, the AI Peak Integration yielded greater precision, especially with uneven baselines (see Figure 5).

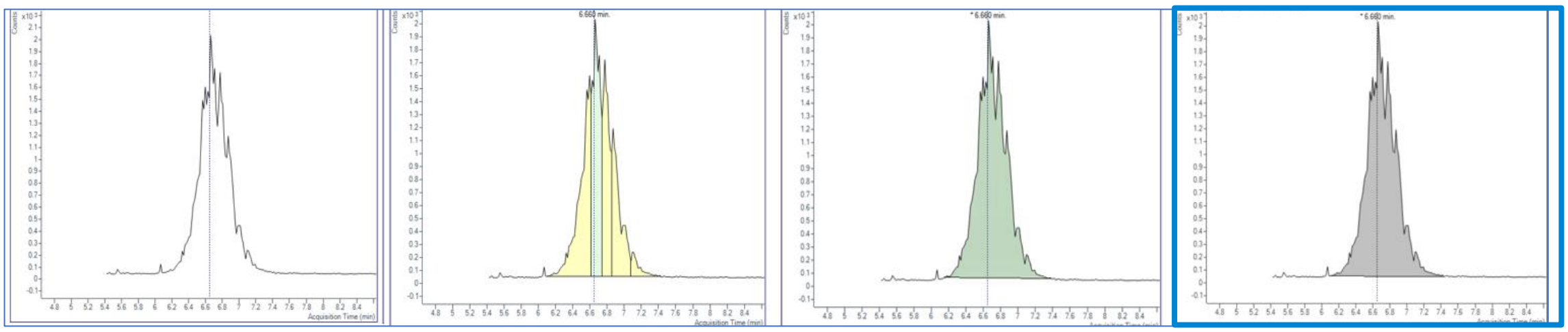


Figure 5: Comparison of peak detection of Diisodecyl Phthalate (DIDP). Left-to-right: a) Not integrated, b) Integrated using MassHunter default parameter-less integrator (Agile2), c) Integrated manually by chemist, d) Integrated using AI Peak Integration for MassHunter.

Future

Machine Learning modelling holds promise for several applications in GC/MS. Extending the solution to other substances, such as regulated compounds in oil and gas, and additionally agriculture (pesticides) can benefit from reducing manual integration time of non-conventional peak shapes.

Environmental semi-volatile organic compounds are candidates for future experimentation. Other substances, like short chain chlorinated paraffins (SCCPs, see Fig 6) are also difficult to integrate.

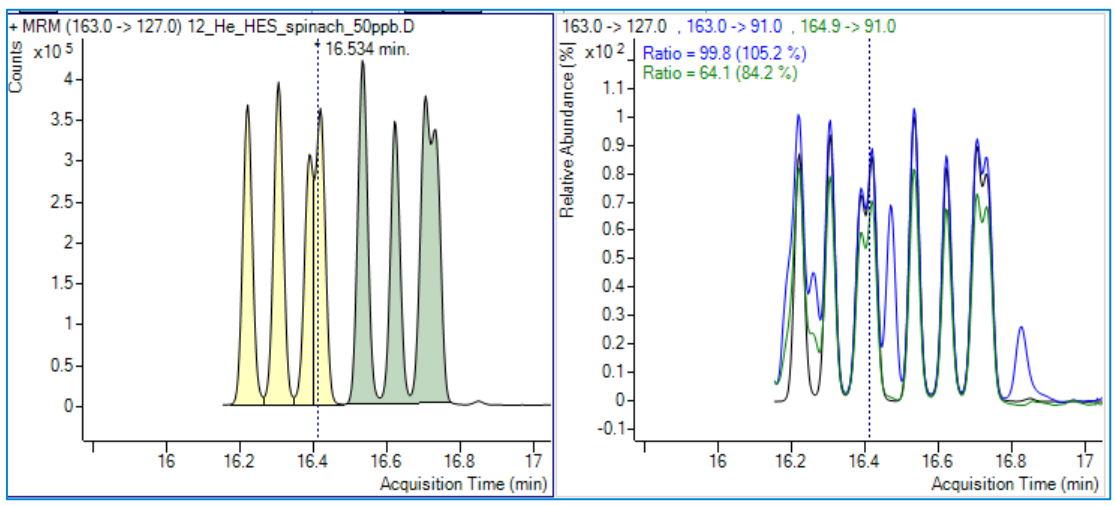


Figure 6: Peaks of short chain chlorinated paraffins could benefit from AI Peak Integration.

In addition to more analyte detection, other Machine Learning applications to chromatography might include calibration curve automation, and predictive maintenance.

Conclusions

Chemist time can be spent more valuably

- Reproducible results
- Universal consistency of analytical results
- Ease-of-use for expert and new lab analysts
- Reduced turnaround time
- Improved overall lab efficiency

References

[1] Phthalates: The Everywhere Chemical https://www.niehs.nih.gov/research/supported/assets/docs/j_q/phthalates_the_everywhere_chemical_handout_508.pdf

[2] Phthalates and Cumulative Risk Assessment: The Tasks Ahead https://pubmed.ncbi.nlm.nih.gov/25009926/

[3] Phthalate and novel plasticizer concentrations in food items from U.S. fast food chains: a preliminary analysis. Edwards, et al. 2021. doi: 10.1038/s41370-021-00392-8.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9119856/

https://www.agilent.com/en/promotions/asms

This information is subject to change without notice.

DE65869893

© Agilent Technologies, Inc. 2023
Published in USA, May 31,2023

Agilent
Trusted Answers