Facilitating & Improving Environmental Data Analysis: A Machine Learning Approach

NEMC

The Role of AI in Environmental Analyses

Ruoji Luo Ph.D.
Application Scientist

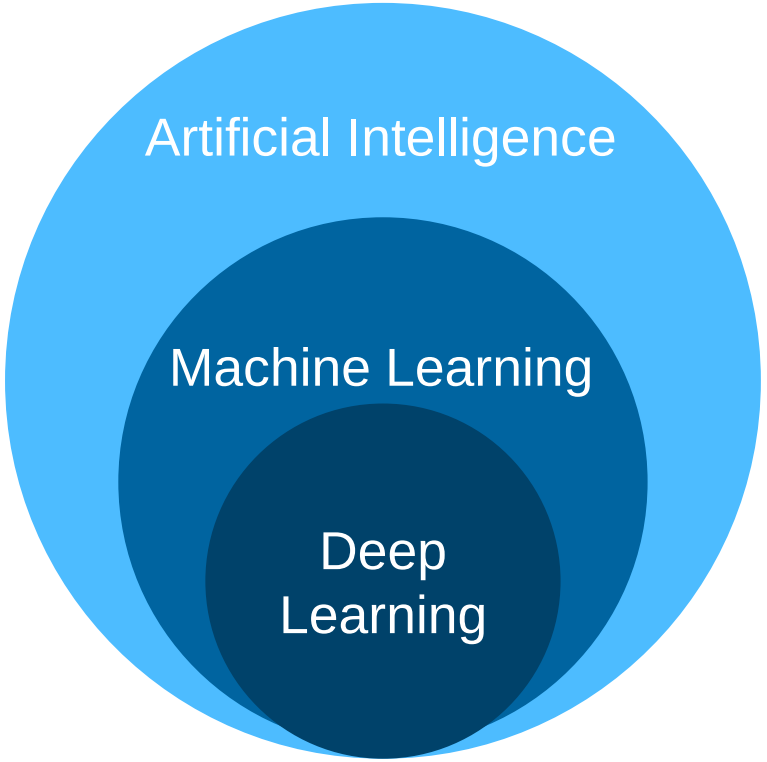
What are AI, ML, and DL?

- AI: Artificial Intelligence

ChatGPT

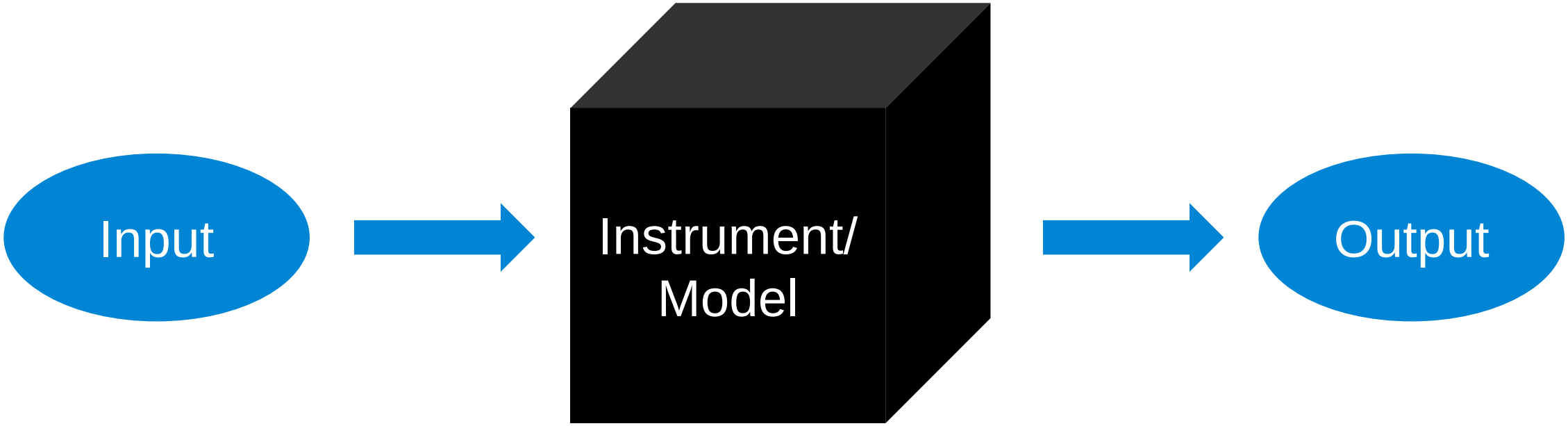
Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think and learn like humans. The goal of AI is to develop systems that can perform tasks that typically require human intelligence, such as visual perception, speech recognition, decision-making, language translation, and problem-solving.

- ML: Machine Learning
- DL: Deep Learning



Analytical Instrument and Machine Learning Model

An Analogy



The Evolution of AI/ML

A Century's Journey

- 1940s-70s:
  Mark I Perceptron machine
  Machine Learning by Arthur Samuel
- 1970s: First AI winter
- 1980s:
  Recurrent Neural Network
  Reinforcement Learning
  Backpropagation
- 1990s: Second AI winter
- 2000s:
  Support Vector Machine
  Random Forest
- 2010s-now:
  Deep Learning
  Generative AI

https://americanhistory.si.edu/collections/nmah_334414

The Evolution of AI/ML

A Century's Journey

- 1940s-70s:
  Mark I Perceptron machine
  Machine Learning by Arthur Samuel
- 1970s: First AI winter
- 1980s:
  Recurrent Neural Network
  Reinforcement Learning
  Backpropagation
- 1990s: Second AI winter
- 2000s:
  Support Vector Machine
  Random Forest
- 2010s-now:
  Deep Learning
  Generative AI

https://americanhistory.si.edu/collections/nmah_334414

Deep Learning Applications and Architectures

- Computer vision
- Speech/audio recognition
- Natural language processing
- Drug discovery
- Protein structure prediction

Prompt: A petri dish with a bamboo forest growing within it that has tiny red pandas running around.
Sora | OpenAI

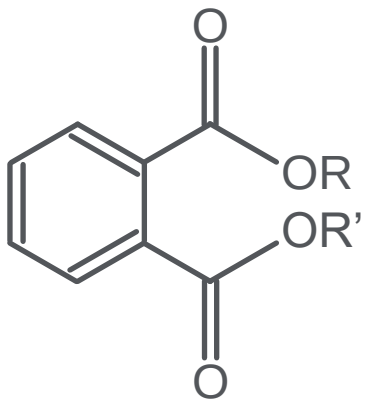
AI Peak Integration for MassHunter Quantitative Analysis

Optimizing your data analysis

Case Study

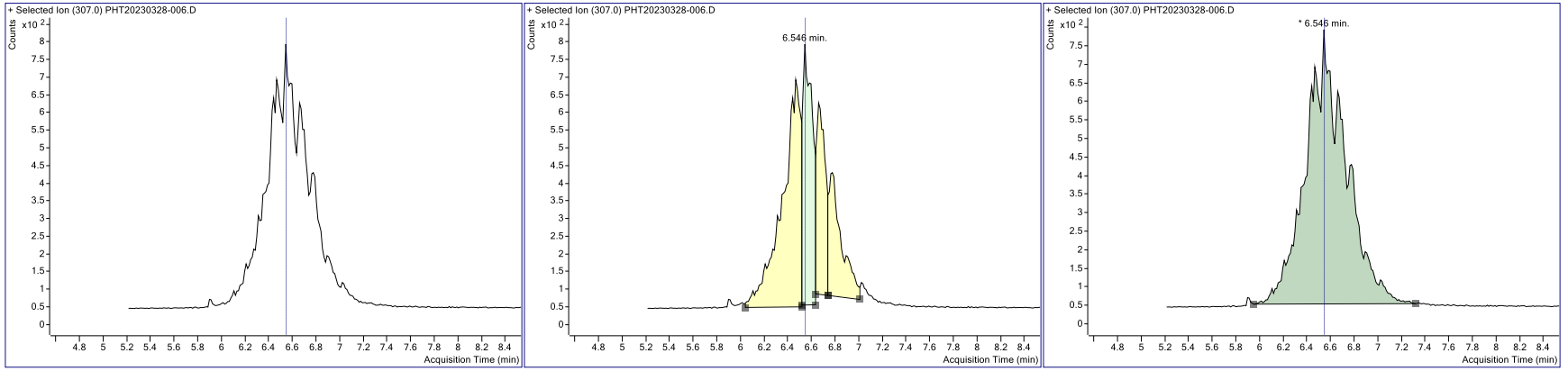
Phthalates: Addressing Data Integration Challenges

- Esters of phthalic acid
- Low (C_3-C_6)/high-molecular-weight phthalates
- Used as plasticizers



- Integration challenges: e.g., for isomers (such as DINP and DIDP)

SIM (*m/z* = 307) of diisodecyl phthalate (DIDP)

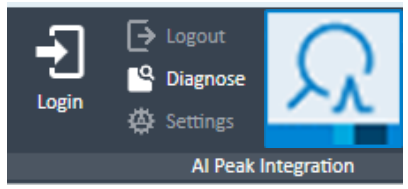


Raw data | MH integration | Manual integration

Agilent AI Peak Integration

General Workflow for MassHunter Quantitative Analysis Software

Training Mode



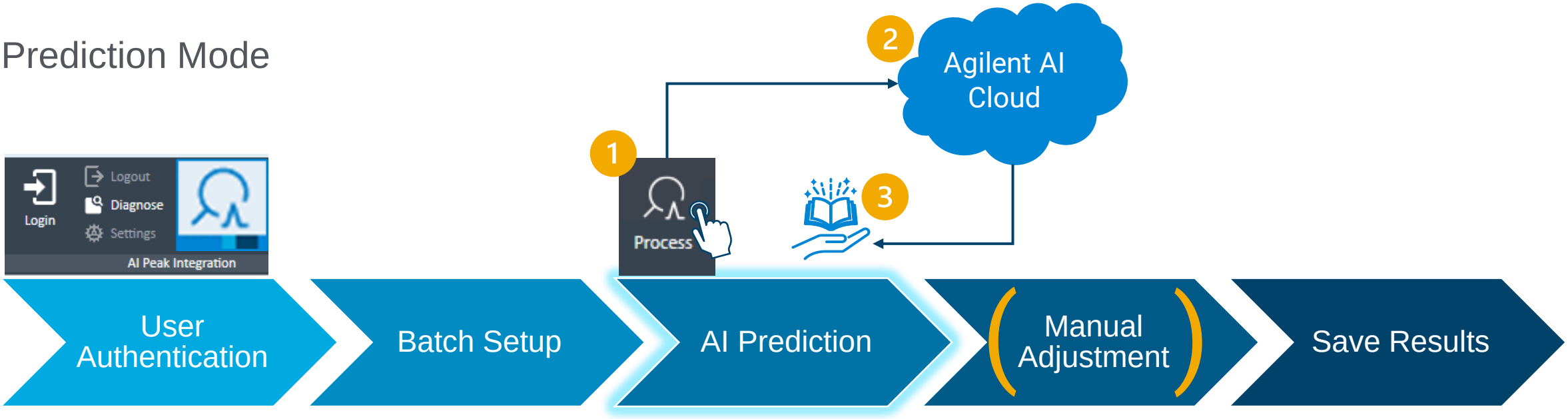
User Authentication → Batch Setup → Batch Processing → Manual Adjustment → Save Results

- Log in to Agilent AI Cloud

Agilent AI Peak Integration

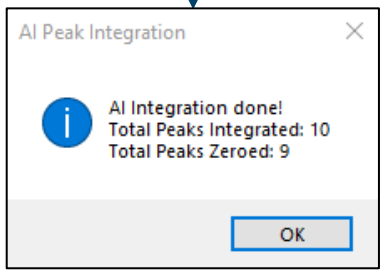
General Workflow for MassHunter Quantitative Analysis Software

Prediction Mode

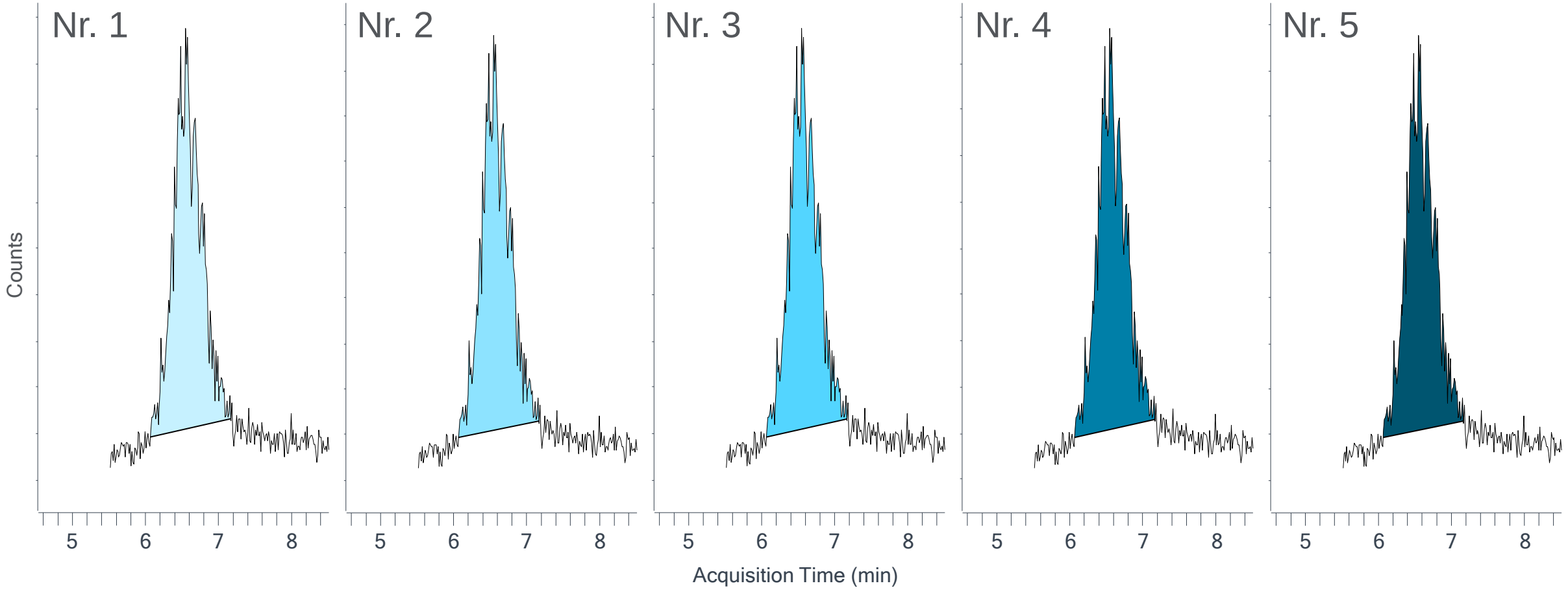


- Log in to Agilent AI Cloud

1. Process batch results in MassHunter
2. AI Peak Integration processing is applied to sample data
3. Results processing complete with AI Peak Integration



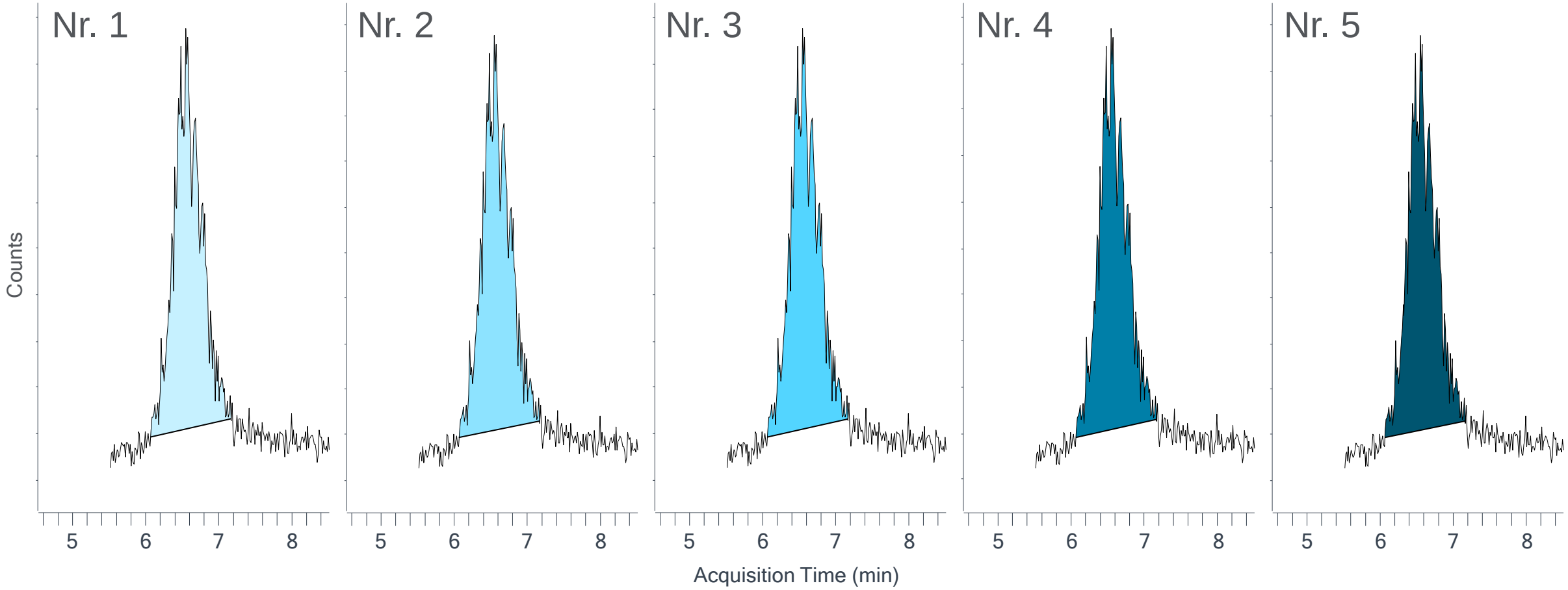
Can You Find the Differences Between the Five SIM Chromatograms?



SIM (*m/z* = 307) of diisodecyl phthalate (DIDP)

Model Reproducibility

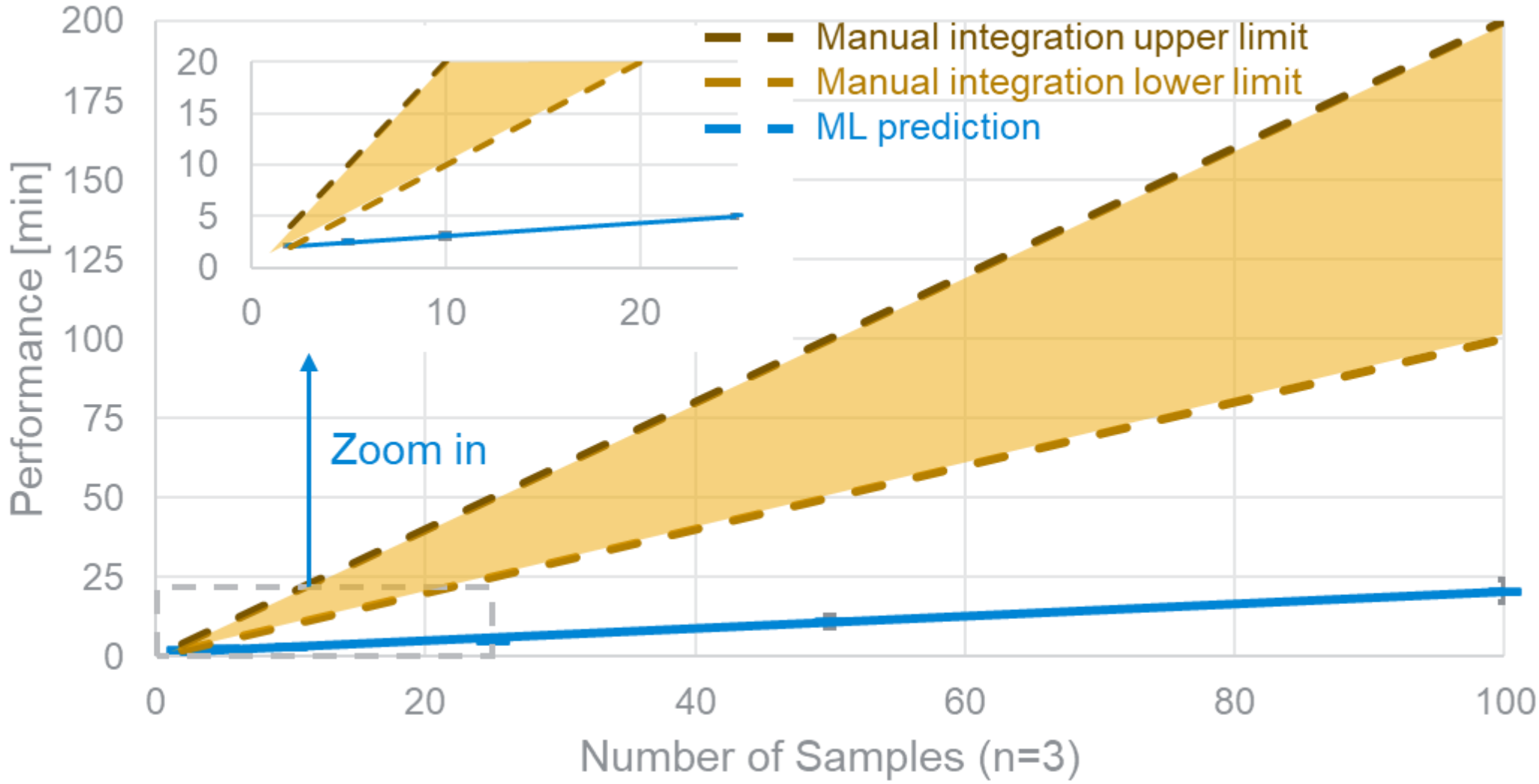
Reproducible ML Integration for the Same Sample Using the Same Model Version



SIM (*m/z* = 307) of diisodecyl phthalate (DIDP)

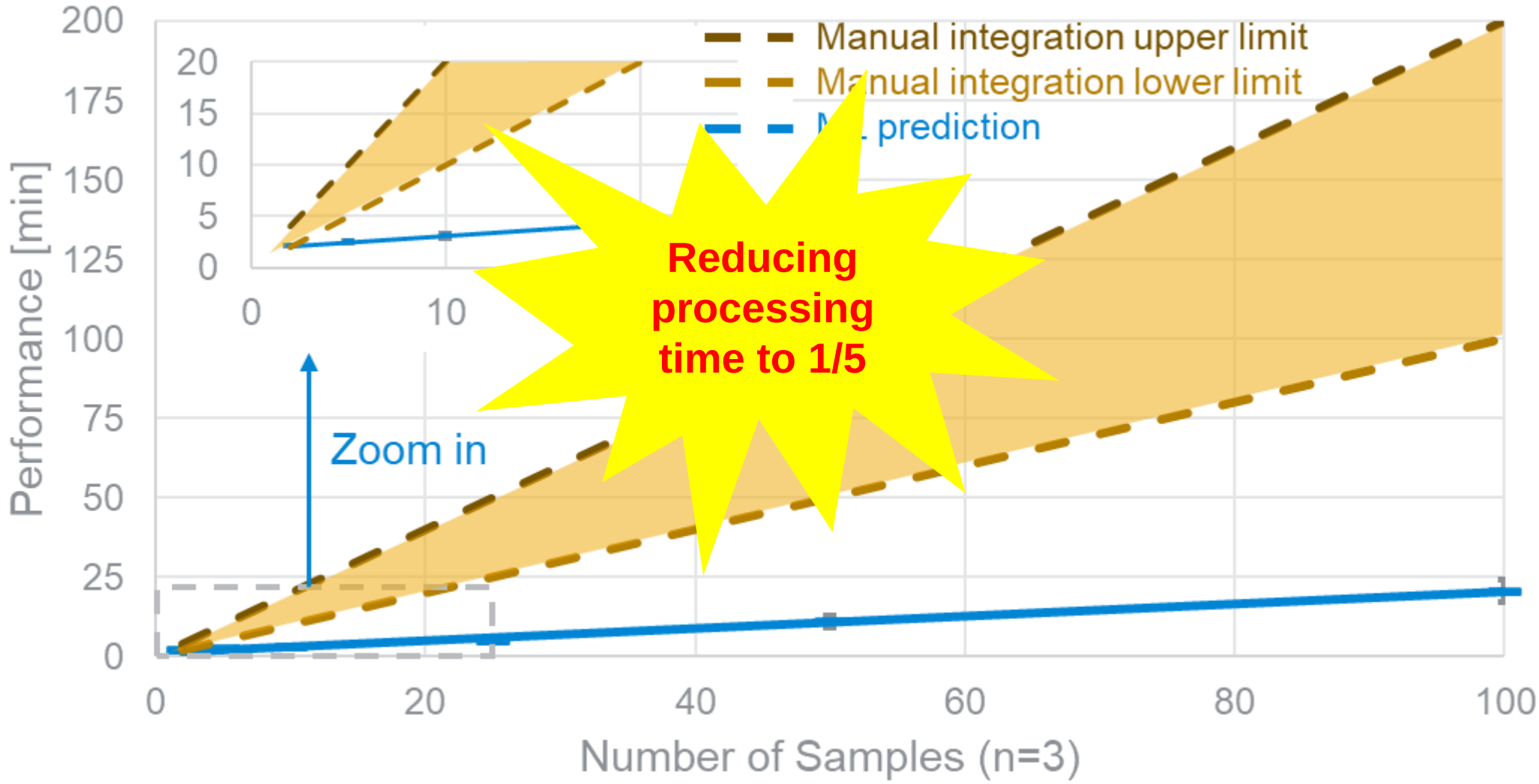
Model Processing Speed

Performance comparison between ML model vs. manual integration



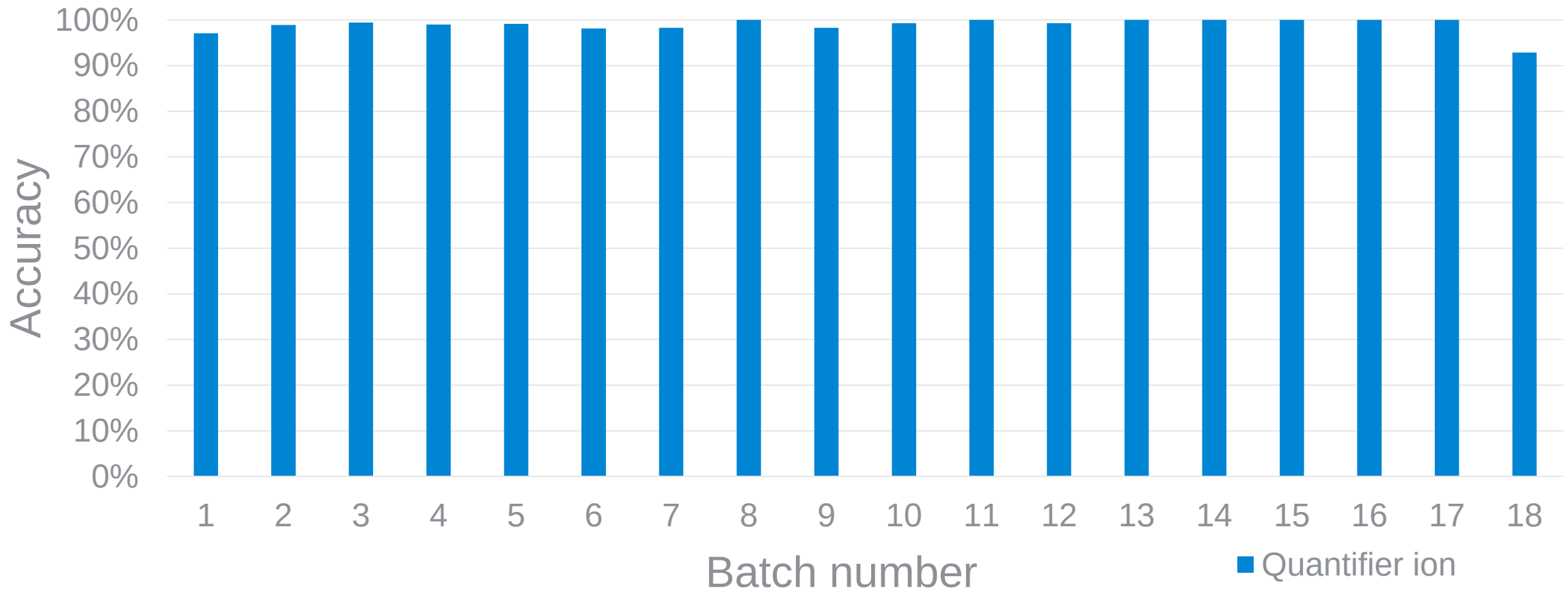
Model Processing Speed

Performance comparison between ML model vs. manual integration



Model Accuracy

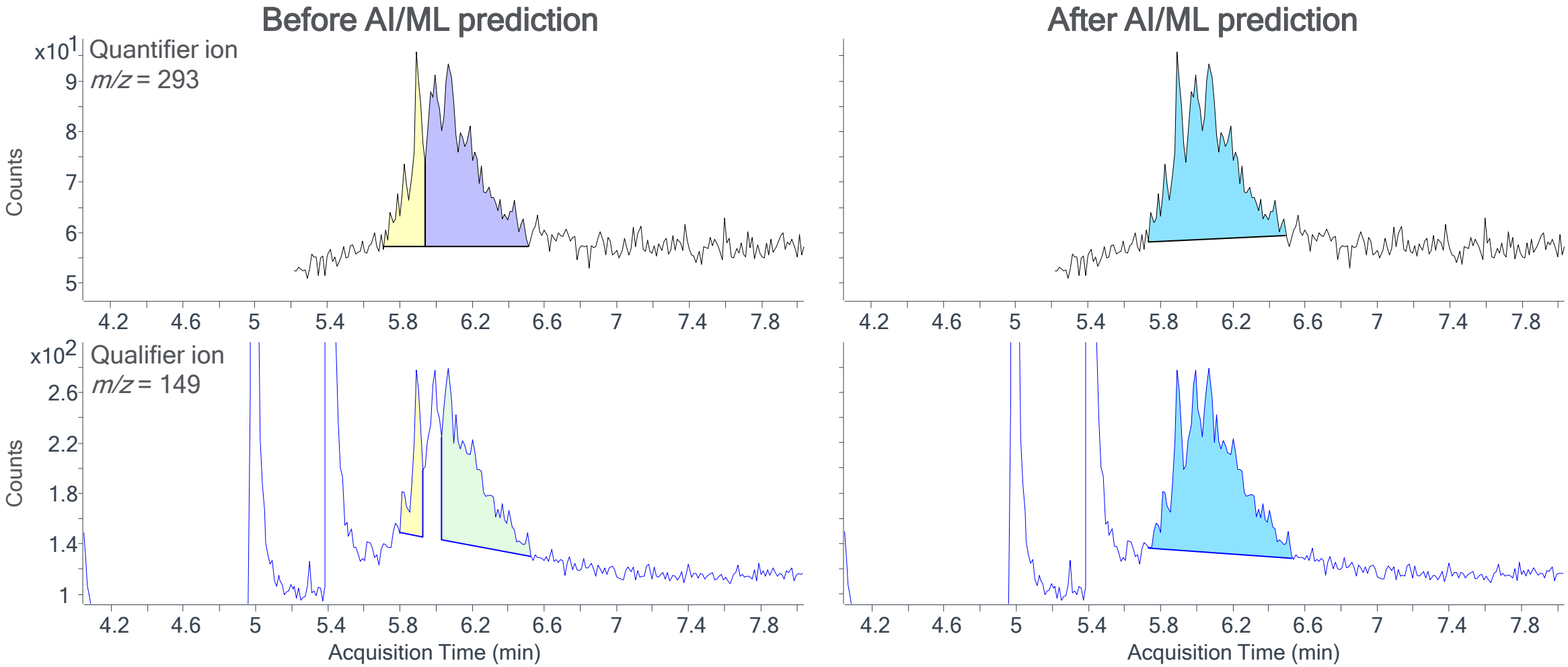
Stable accuracy across batches



Quantifier ions of each phthalate in over 550 samples

Model Accuracy

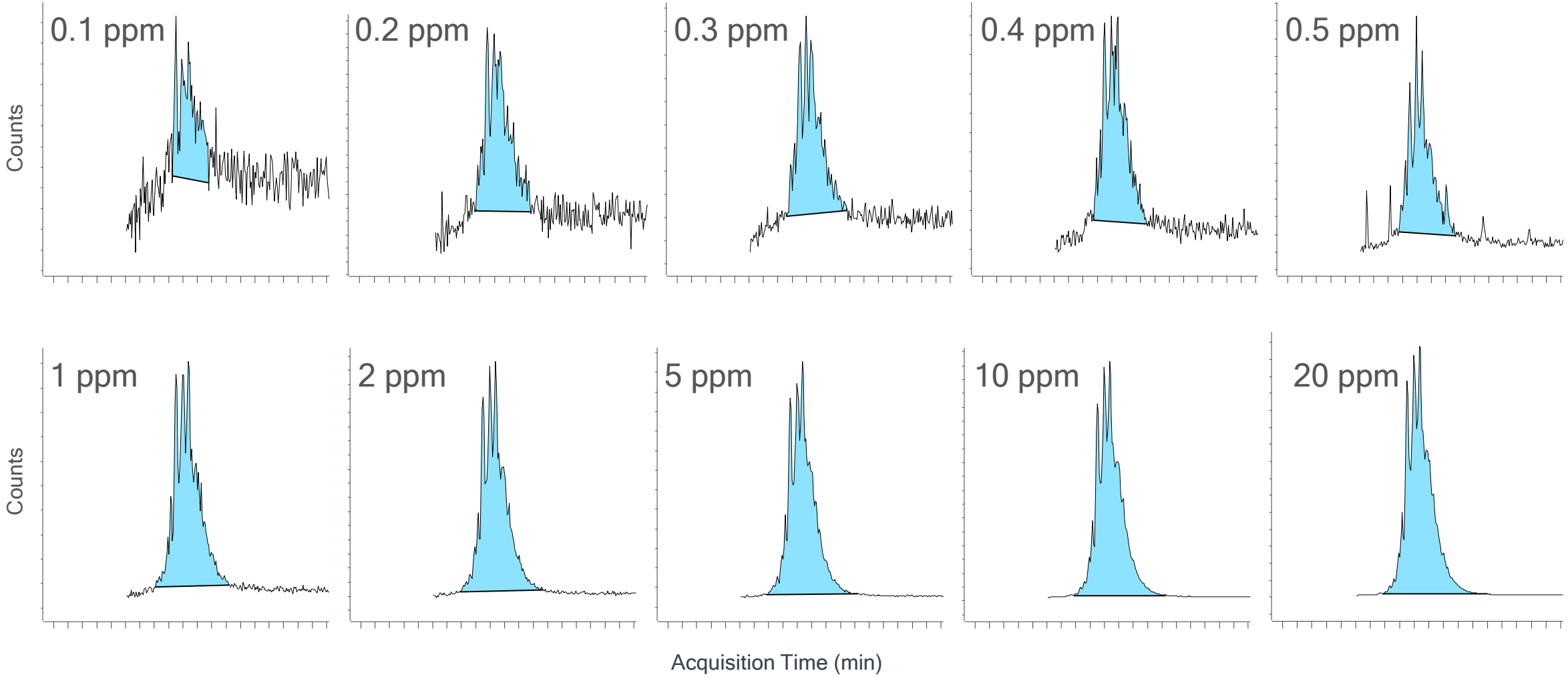
Improvements on integration performance with AI/ML



Diisononyl phthalate (DINP)

Model Accuracy

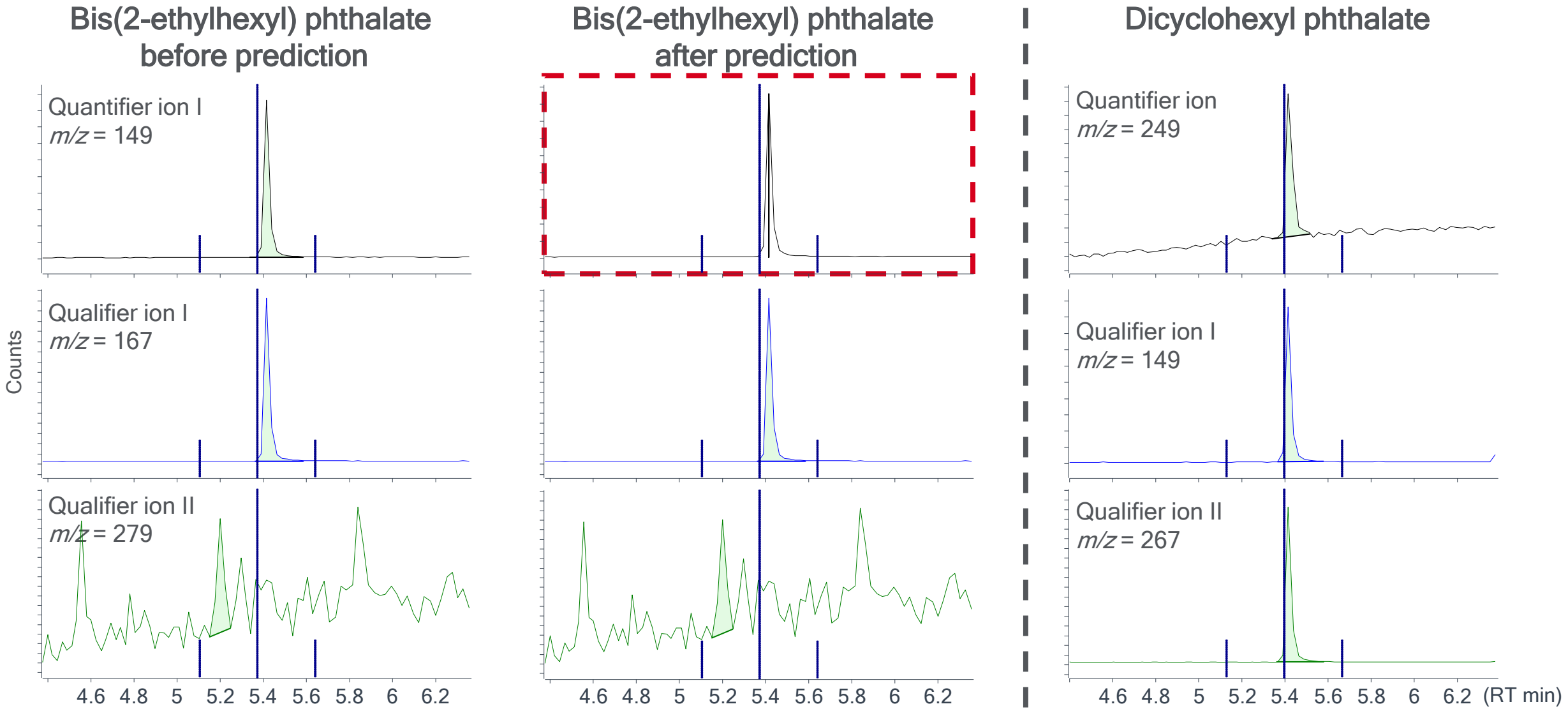
At Different Concentrations



SIM (*m/z* = 293) of diisononyl phthalate (DINP)

Model Accuracy

Correction of False Positive & Negative Peaks



Summary

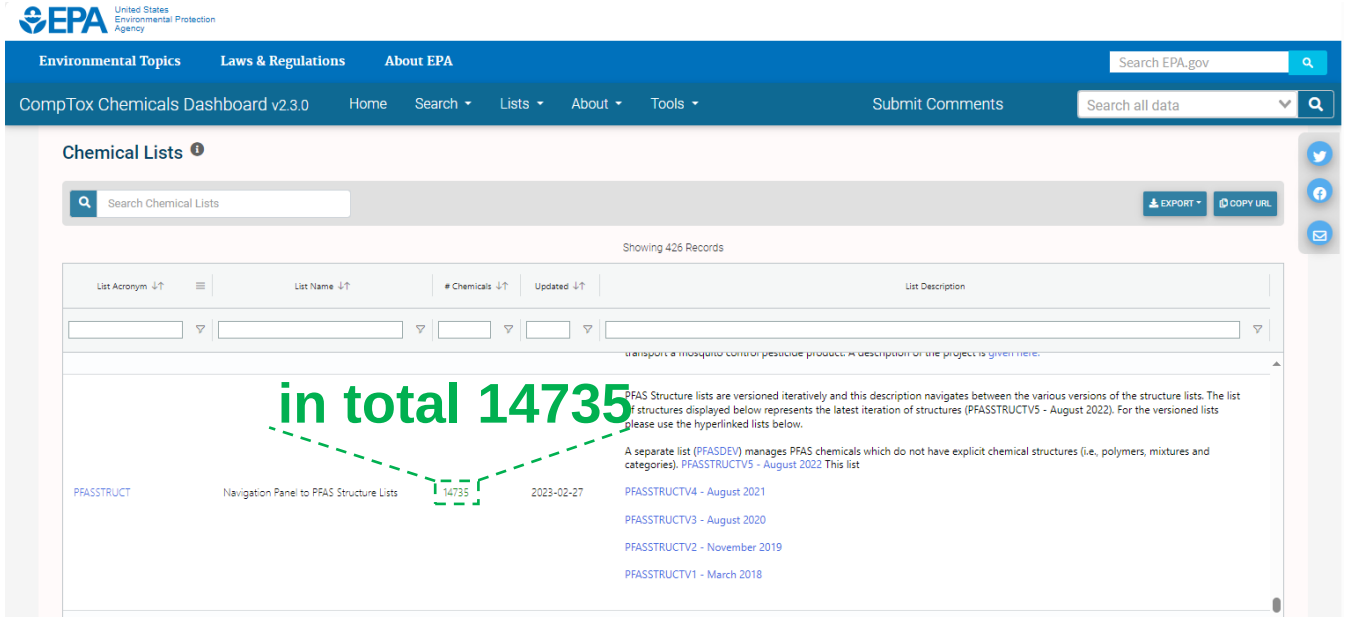
- Demonstration of integrated AI solution into MassHunter Quantitative Analysis software
- Reproducible and reliable peak prediction
- Speed-up of the peak reviewing process by 4 to 5-fold
- Successful handling of traditional manual peak integrations in GC/MS phthalate analysis:
  - ✓ Targeted ions (quantifier & qualifier ions)
  - ✓ Across and beyond calibration range
  - ✓ False positive & negative peaks

Case Study: PFAS

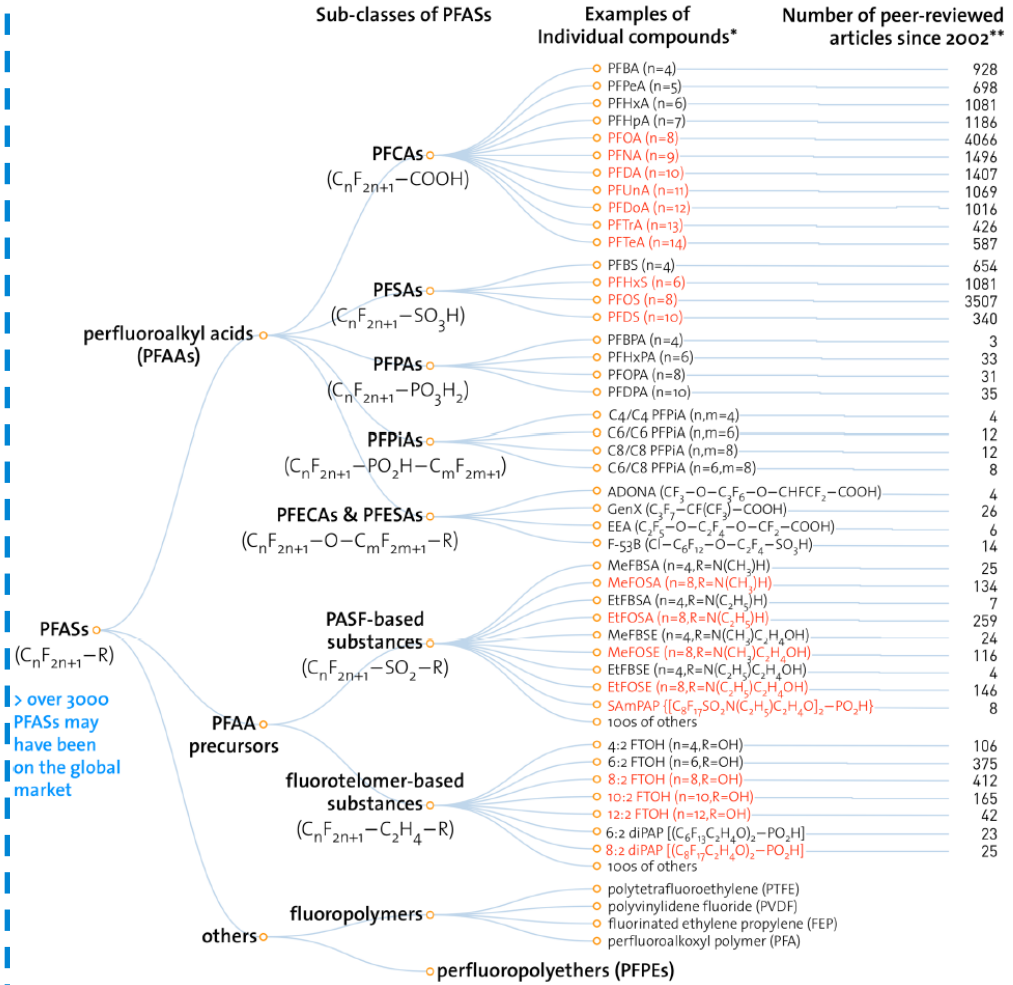
Background

- PFAS: Per- and Polyfluoro Alkyl Substances
- Environmental and health issues
- Analytical techniques:

  LC/TQ, LC/(IM)QToF, GC/MSD, GC/TQ, GC/QToF
- Sample preparation and consumables are critical



Classification of PFAS



Wang, Z *et al.* (2017). *Environ. Sci. Technol.* 51, 2508-2518

Peak Integration Challenges for PFAS Analysis

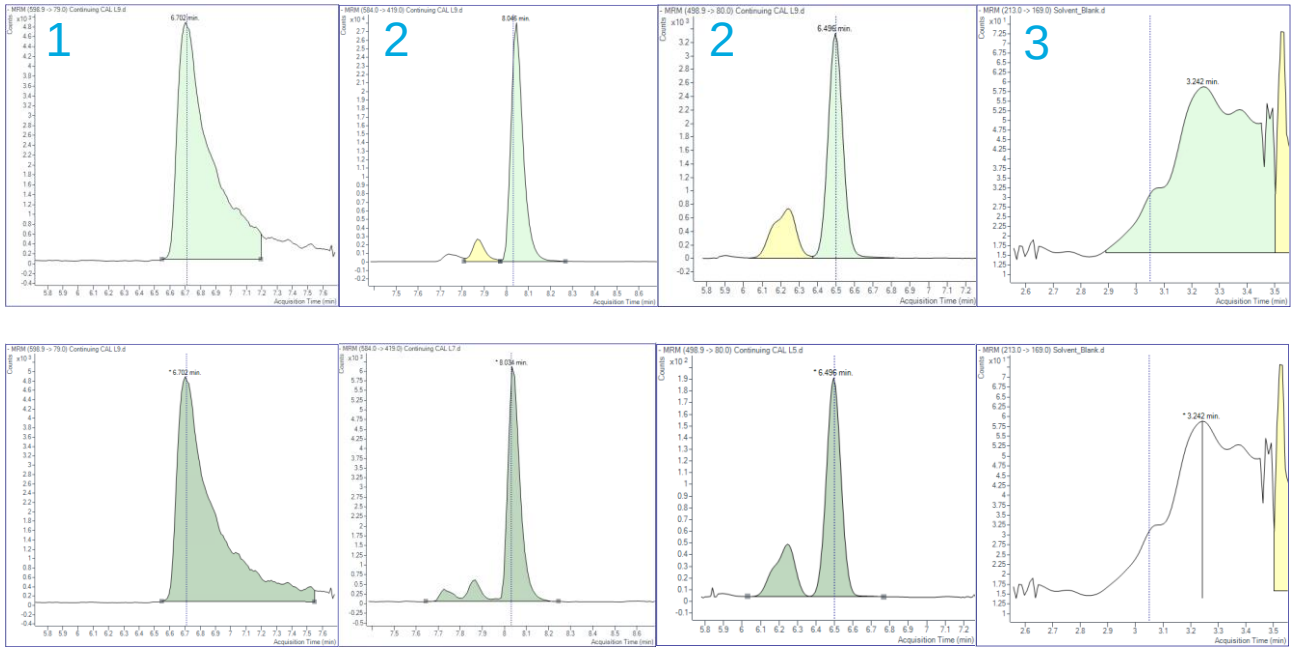
In general:

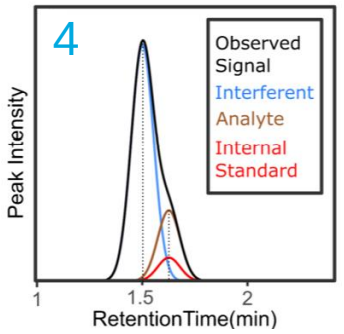
- Baseline correction[1]
- Peak combination/splitting[2]
- False positive/negative peaks[3]
- Background interferences[4]

PFAS specific:

- Early eluting PFAS (bad peak shape)
- Linear and branched PFAS (isomers)
- Varying ratio of linear and branched PFAS

Examples

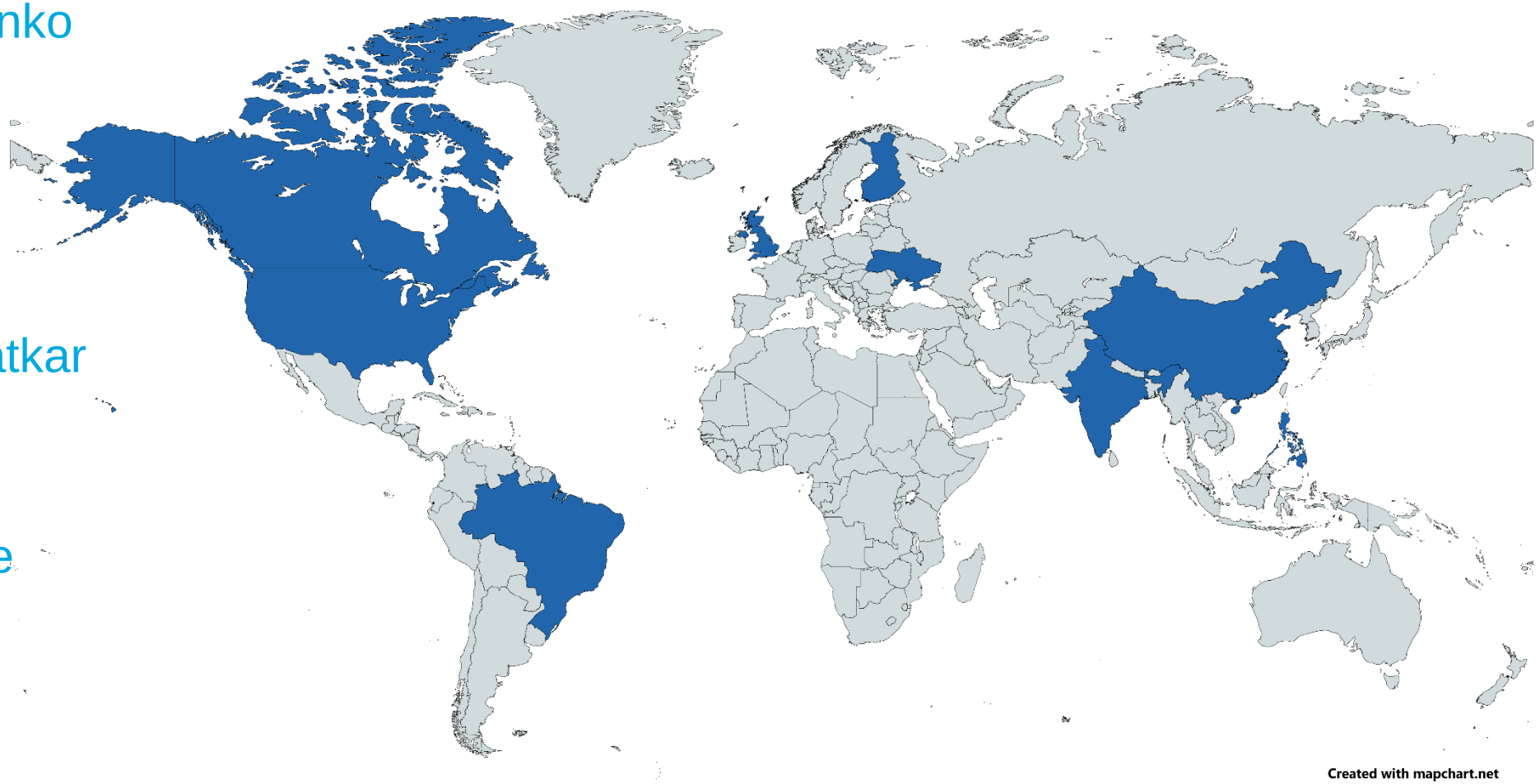




Bangma, J. *et al.* Anal Bioanal Chem 416, 1777–1785 (2024)

Acknowledgement

Tamas King
Alex Sosnovshchenko
Ralph Ramos
Mat Espiritu
Alex Graettinger
Flavio Avila
Charudatta Manwatkar
Edison Cerda
Sunil Rehman
Abhijeet Sonawane
Tarun Anumol
Emily Parry
Matthew Giardina
…



Agilent

Trusted Answers