

Implementing a Digital Transformation in Process Chemistry: Integrated Automation, Machine Learning, and Real-Time Analytics from Lab to Pilot Scale

Elena Braconi*, Jean-Philippe Krieger*, and Thomas Vent-Schmidt*

Abstract: Implementing a digital transformation in the Process Research and Development (PR&D) phase of an active ingredient offers significant opportunities to accelerate the journey from laboratory to manufacturing scale. Here, we report on three distinct initiatives undertaken at Syngenta to address concrete bottlenecks at different PR&D stages. First, Bayesian optimisation enabled efficient navigation of large reaction spaces with minimal experimental effort. Second, laboratory automation combined with multilinear calibration reduced hands-on time by ~85% and laid the foundation for autonomous closed-loop optimisation. Third, advances in Process Analytical Technology (PAT), including improved Multivariate Curve Resolution algorithms and modular Python-based pipelines, enabled real-time reaction monitoring in challenging industrial settings.

Keywords: Automation · Bayesian optimisation · Flow chemistry · Process Analytical Technology (PAT) · Process development



Elena Braconi is a Team Leader in Process Technology for New Active Ingredients at Syngenta Crop Protection AG. She earned her PhD in organic chemistry in 2021 from EPFL under the supervision of Prof. N. Cramer, focusing on asymmetric catalysis with Earth-abundant metals. She joined Syngenta in 2021 as a postdoctoral researcher in Process Research, specialising in sustainability and data-driven reaction

optimisation. Since 2023, she has been developing chemical processes from laboratory to pilot scale, focusing on the integration of digital transformation strategies to accelerate and optimise process development.



Thomas Vent-Schmidt is a Leading Scientist in Process Technology for New Active Ingredients at Syngenta Crop Protection AG. He earned his PhD in 2015 from the Albert-Ludwigs-University in Freiburg/Germany under the supervision of Prof. S. Hasenstab-Riedel. His work focused on the quantum-chemical prediction of f-element metal halides and noble gas compounds as well as their experimental investigation using

matrix-isolation spectroscopy. In Syngenta, he leads the global PAT activities and drives digital projects in the analytical space.



Jean-Philippe Krieger leads the ‘Enabling Processing Technologies’ team in the Process Technology for New Active Ingredients department at Syngenta Crop Protection AG. After obtaining an engineering degree from ESPCI Paris, he earned his PhD in organic chemistry in 2015 under the supervision of Prof. Janine Cossy, focusing on the development of methodologies for macrocycle construction. Following post-

doctoral studies with Prof. Cristina Nevado at the University of Zürich, he joined Syngenta in 2017 as a Team Leader in process development. Since 2020, his work has focused on enabling technologies, including continuous processing and automation.

1. Introduction

Within the journey that brings a new active ingredient from discovery to market, the Process Research and Development (PR&D) phase constitutes a critical bridge. Its primary mission is to identify robust, safe, and cost-effective synthetic routes, ensuring a seamless transition from laboratory to manufacturing scale.

The PR&D phase has been traditionally hampered by significant operational bottlenecks. First, the labour-intensive and repetitive manual execution of experiments, which limits the throughput of an optimisation campaign. Second, the fragmentation of data (where information from various sensors and laboratory platforms remains siloed), which creates barriers to a holistic understanding of the overall process.

In line with broader trends in the chemical industry, the PR&D departments at Syngenta have undertaken significant efforts to address these bottlenecks by integrating digital tools directly into core development workflows. While this is a complex and ongoing journey, concrete progress is already being made across multiple stages of the development process.

In this manuscript, we report on three distinct examples of how digital tools are being integrated into process development workflows at Syngenta. We begin with Bayesian optimisation, applied in the early stages of development to efficiently navigate large reaction spaces. We then describe how laboratory automation reduces manual effort and accelerates flow chemistry optimisation

*Correspondence: E. Braconi, E-mail: elena.braconi@syngenta.com; J.-P. Krieger, E-mail: jean-philippe.krieger@syngenta.com; T. Vent-Schmidt, E-mail: thomas.vent-schmidt@syngenta.com.
Syngenta Crop Protection AG, CH-4333 Münchwilen, Switzerland

campaigns. Finally, we present advances in Process Analytical Technology (PAT) for real-time reaction monitoring, and their deployment at manufacturing scale. Together, these examples illustrate the breadth of opportunities that digital transformation offers across the process development lifecycle.

2. Machine-Learning for Early-Stage Process Development

AI and machine learning are reshaping chemical research. In particular Bayesian optimisation (BO) is emerging as a powerful tool for navigating vast and complex ‘black-box’ chemical spaces.^[1] The interest in BO within the chemistry community has grown rapidly over the past decade, with a ~40-fold increase in publications (Fig. 1). This trend reflects both the method’s maturation and its growing industrial relevance.

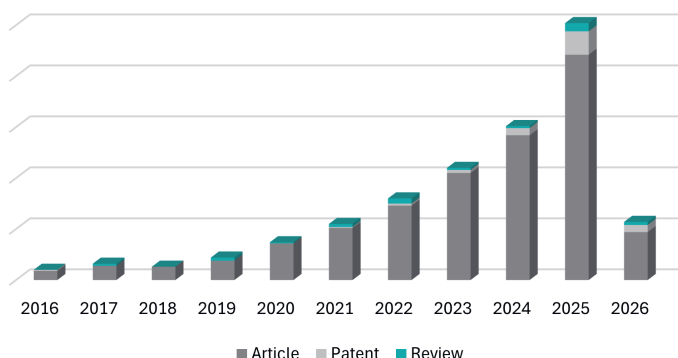


Fig. 1. Growth of Bayesian Optimization literature in chemistry (2015–2026).

BO complements traditional reaction optimisation techniques like One-Factor-at-a-Time (OFAT),^[2] Design of Experiments (DoE)^[3] and High-Throughput Experimentation (HTE).^[4] Conventional OFAT is limited by high resource consumption and an inability to account for synergistic variable interactions. While DoE addresses these interactions, it is primarily suited for continuous variables (categorical variables introduce greater complexity) and suffers from an exponential increase in experiments as the number of factors grow. HTE can expand screening capacity, but it remains insufficient for spaces with millions of combinations. BO fills this gap by accounting for complex interactions in vast chemical spaces with minimal experimental input.

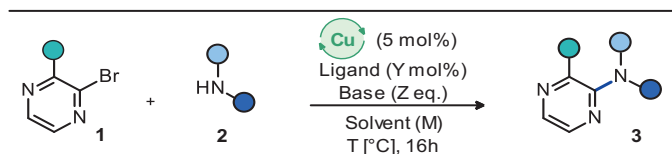
A Bayesian reaction optimisation workflow^[5] begins with the user defining the chemical space (including categorical and numerical variables), alongside the objectives and experimental budget. The optimiser then recommends initial experiments. Through iterative learning, results are fed back to refine the surrogate model and objective function. An acquisition function selects subsequent trials by balancing exploration of new regions with exploitation of known optima, repeating this cycle until the objective is achieved or the budget is exhausted.

In 2023, within our Process Research department, we evaluated the potential of BO for early-stage process development. We aimed to benchmark BO against other optimisation techniques (such as OFAT, HTE, and DoE) to assess their respective advantages and limitations and determine the most suitable R&D stage for each.^[6] In the following sections, we present two case studies: the development of a sustainable Cu-catalysed C–N coupling and the optimisation of a nucleophilic substitution manufacturing process.

2.1 Optimisation of a Copper-Catalysed Ullmann Coupling

The first case study evaluated the utility of BO for the C–N coupling reaction involving sterically hindered bromo-pyrazines

1 and amines **2** (Scheme 1).^[7] Although copper-catalysed versions of this transformation are well-documented in literature,^[8] an analysis of our internal electronic laboratory notebook (ELN) revealed that palladium remains the primary choice, with only 31% of metal-catalysed C–N couplings utilising copper. This preference for palladium stems from its inherent robustness and generality, whereas copper systems often require specific optimisation for different reactant combinations. The objective of this study was to identify sustainable, high-yielding alternatives using Earth-abundant copper catalysts and non-hazardous solvents.



Scheme 1. Cu-catalysed C–N Coupling of pyrazines **1** and amines **2**.

The explored chemical space included 5 copper sources, 12 ligands (each tested at 2 loadings), 8 bases (evaluated at 3 stoichiometries), 6 solvents, 2 molarities, 2 temperatures, and 2 stoichiometries for the amine coupling partner. The combination of these parameters yielded a total of 138,240 possible experimental conditions.

To navigate this chemical space, an experimental budget of 80 reactions was arbitrarily set, divided into ten rounds of eight parallel experiments. Three optimisation campaigns using Bayesian optimisers with different acquisition functions were conducted: EDBO-TS, EDBO-EI,^[9] and AUTOMUS.^[10] These platforms differ primarily in their search strategy:^[1a] EDBO-EI utilizes Expected Improvement to favour experiments with the highest potential for yield improvement; EDBO-TS employs Thompson Sampling to balance exploration and exploitation by gradually favouring exploitation as sampling evolves; AUTOMUS uses Upper-Confidence Bound to negotiate this balance by considering how often specific clusters of the reaction space have been visited. To ensure a fair comparison, the first round of experiments was randomly generated and served as an identical starting point for all three optimisation campaigns.

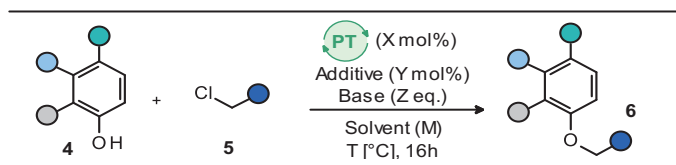
Starting from an initial yield of 21% identified in the shared random round, each optimiser substantially improved the results within the defined budget, though they reached different optima. For the investigated case study, the EDBO-EI platform identified optimised conditions yielding 87% of the target product after nine rounds. The AUTOMUS optimiser reached a 78% yield in just six iterations, while EDBO-TS achieved a maximum yield of 64%.

In all cases, by exploring less than 0.05% of the total reaction space, the BO framework successfully identified various high-yielding and sustainable copper-catalysed conditions that would have likely been missed *via* other optimisation methods. Notably, some discovered conditions were unreported in literature for Ullman couplings, highlighting the BO ability to find novel, unbiased solutions that escape human intuition. Furthermore, the identification of diverse optimal conditions on the model substrate allowed for a significantly higher success rate when the methodology was applied to a broader library of substrates.

2.2 Discovery of New Conditions for an Existing Nucleophilic Substitution Manufacturing Process

The second case study focused on a phase-transfer (PT) promoted nucleophilic substitution reaction (Scheme 2). This investigation was prompted by the need to identify new reaction conditions for the manufacturing process of an active ingredi-

ent, primarily due to regulatory constraints regarding solvent usage.



Scheme 2. PT-catalysed nucleophilic substitution of **4** into **5**.

Unlike the first case study, the primary objective of this work was to determine if a significantly larger chemical space (2,764,800 possible combinations) could be effectively explored using the same experimental budget of 80 reactions, conducted in eight rounds of ten parallel experiments. In fact, a current open question in BO is the determination of the optimal experimental budget for a given chemical space to maximize efficiency and effectiveness. To ensure representative sampling of the chemical landscape, 15 phase-transfer catalysts were selected based on clustering of their ‘C_count’ and ‘q-value’ descriptors (C_count = total number of carbons on the 4 chains; q-value = sum of reciprocals of the number of carbons on the 4 chains).^[11] Optimised parameters included the stoichiometry of the starting material, PT catalyst, base, and additive, as well as the identity of catalysts, bases, additives, and solvents. Molarity and temperature were also included. As in the previous case study, all categorical variables were handled *via* one-hot-encoding (OHE), which allows their representation as unique binary vectors of zeros (0) and ones (1).^[12] While OHE does not incorporate inherent chemical information, it demonstrates only a marginal decrease in performance compared to DFT-based encoding, while eliminating the need for prior computational efforts.^[9]

Using the EDBO optimizer with Thompson Sampling, the campaign identified several high-yielding reaction conditions. These discovered conditions were diverse in terms of the catalysts, bases, additives, and solvents utilised, allowing the team to select the most appropriate options based on cost and scalability. The chosen conditions were further evaluated using DoE to assess robustness and model the reaction space near the optimum. Following a successful scale-up, this approach delivered a viable, sustainable alternative to the original manufacturing process.

2.3 Considerations on Bayesian Optimisation

While the advantages of BO seem compelling, a balanced scientific evaluation must consider its strengths and inherent limitations in comparison to other optimisation techniques. A primary benefit of BO is its extraordinary efficiency in navigating vast, multi-dimensional reaction spaces that would be practically impossible to explore using conventional methods. Unlike traditional DoE, which is primarily suited for the fine-tuning of continuous variables, BO simultaneously investigates categorical and numerical factors. Because the user can define how many experiments to run in parallel, BO does not require any specialised laboratory equipment (which instead significantly helps streamlining HTE workflows). BO can also optimise multiple (sometimes competing) objectives concurrently, allowing the inclusion of the cost and sustainability metrics alongside chemical yield from the start of the development process.

Despite these benefits, BO ‘black-box’ nature remains a challenge. At the conclusion of an optimisation campaign, the scientist often does not completely understand why specific conditions are optimal or how different variables interact (unlike DoE, which

provides a clearer mathematical model and ‘process envelope’ of reaction trends). Furthermore, the technique still faces significant open questions, such as the optimal experimental budget for a specific reaction space size and identifying whether a better acquisition function exists for a given case study.

Consequently, BO should not be viewed as a replacement for other optimisation tools but rather as a powerful precursor in a modern research workflow. In this hybrid approach, BO is utilised in the early stages to rapidly narrow down a massive reaction space to a few high-performing regions. Once these regions are identified, the chosen conditions can be further studied with a DoE aimed at assessing robustness as well as defining a model of the reaction space near the optimum. This integrated approach identifies optimal conditions with significantly fewer experiments than OFAT, HTE, or DoE alone for spaces of comparable magnitude.

The efficiency gains offered by BO can be further amplified when combined with laboratory automation, which removes the experimental throughput bottleneck that otherwise limits the number of iterative cycles achievable within a given timeframe.

3. Lab Automation as an Enabler for Accelerated Process Development

In process chemistry, automation addresses a well-known bottleneck: the time-intensive, repetitive execution of experiments that limits the throughput of optimisation campaigns. Considerable work has been done in this area over the past years, especially on flow chemistry, and several reports of self-optimising platforms have been published.^[13] Flow chemistry, when constrained to the liquid phase, constitutes a natural entry point to automation; handling solids introduces substantial additional challenges related to clogging and representative sampling. Within the Enabling Processing Technologies team, we have pursued continuous efforts since 2023 to make automation a standard practice for flow chemistry experiments, running DoE-based optimisation campaigns as a standard way of working and, more recently, incorporating BO.

Our standard instrumental setup comprises modules for operations (actuators) and data recording (sensors): pumps, thermostats, pressure controllers, samplers, and similar units. A distributed control system (DCS) orchestrates these unit operations through a graphical user interface (GUI), handling multiple communication protocols (4–20 mA, RS-232, TCP/IP, OPC-UA, *etc.*) to integrate diverse equipment.

The standard workflow (Fig. 2) for an automated flow chemistry experiment starts by defining input parameters that describe the steady state to be reached. These parameters are then translated from human-interpretable values (residence time, equivalents, *etc.*) to setpoints (flow rates, signals, *etc.*). Those are sent to the equipment (pumps, thermostats, pressure regulators, *etc.*) composing the installation and the experiment is started.

Once the experiment has been executed, upon sampling it can be analysed either *via* offline or online analysis. The re-

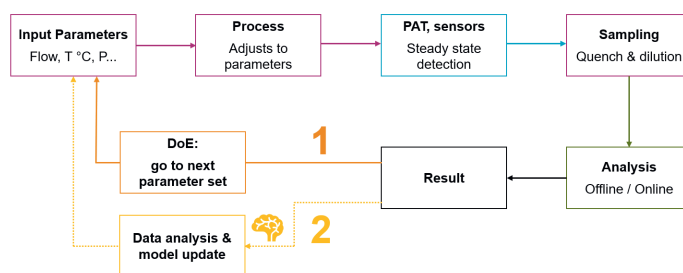


Fig. 2. Standard workflow for automated flow experiments (DoE or closed loop).

sult triggers the execution of the next experiment. In a classical DoE optimisation campaign, the next experiment in the list is executed. However, the introduction of online analytical measurements opens the door for autonomous / closed loop experimentation.

An analysis of Syngenta internal data showed that for a standard 30-experiment optimisation campaign, automation reduces hands-on time by approximately 85% (Fig. 3). The residual time required consisted of the preparation of feed solutions and definition of process space boundaries and constraints for the optimisation.

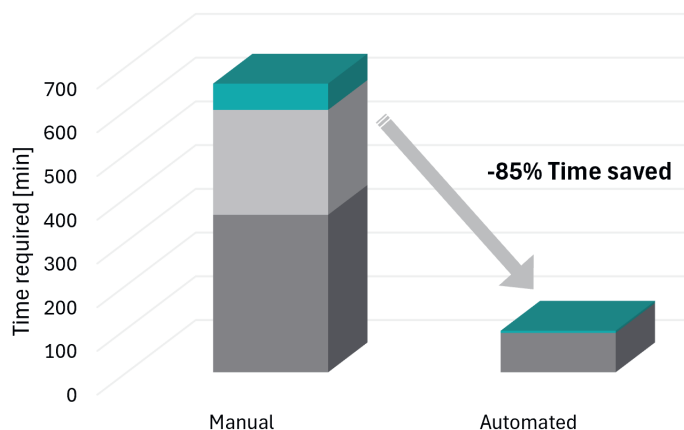


Fig. 3. Time requirement comparison between manual and automated experimentation for $n=30$ experiments (estimation based on Syngenta internal data).

In the sections below, we will showcase two examples of how automation in flow chemistry serves as an enabler for accelerated process development.

3.1 Automated DoE in Flow

In 2023, we started by implementing in our laboratories an automated flow chemistry experimentation workflow. In these initial steps, we decided to leave analytical data analysis outside of the workflow. Our standard setup consisted of pumps, pressure sensors, temperature sensors, and a sampler equipped with a valve to either collect or discard the reaction mixture (Fig. 4). A sample was collected upon reaching steady state, assumed to occur after stabilisation of all process setpoints followed by three residence times to the collection valve — a widely accepted criterion for flow systems.^[14] At the outlet of the reactor, a pump performed quench, dilution, and internal standard (ISTD) addition of the collected sample to ensure representative and precise analytical results.

This automated set up was tested in several DoE optimisation campaigns. The experimentalist pre-filled an Excel template with the series of experiments to be performed, as well as a representation of the setup. This Excel file is read by a proprietary Python script that translates each experiment from human-readable inputs into machine-based setpoints (Fig. 5). An additional software layer enables fully automated execution *via* programmatic interaction with the DCS. An internally developed Python module drives the complete DoE workflow: a list of input vectors describing the planned experiments is prepared; the module sends each coordinate to the corresponding equipment in the appropriate format; after reaching steady state, a sample is drawn, the setup is purged, and parameters are adjusted for the next condition. This iterative process continues until the full experimental program is complete. Analysis is then performed offline using the appropriate analytical method (HPLC, GC, NMR).



Fig. 4. First-generation standard setup comprising pumps (highlighted in orange), pressure and temperature sensors (highlighted in blue), pressure regulators, and a sampler (highlighted in red). Equipment is controlled by a DCS (Hitec Zang GmbH), itself driven by a Python script.

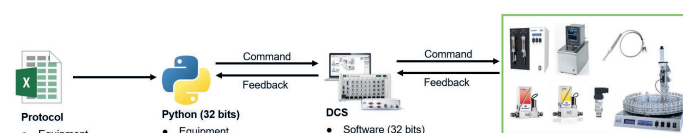
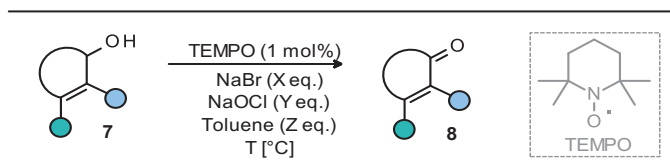


Fig. 5. The different layers involved in the first-generation automation workflow: from human (table) to Python to DCS to equipment.

This workflow was tested to optimise the biphasic oxidation of an allylic alcohol **7** to the corresponding ketone **8** in flow (Scheme 3).



Scheme 3. Biphasic allylic alcohol **7** oxidation to ketone **8** optimised in flow with 4 parameters: temperature and equivalencies (NaBr, NaOCl, and toluene).

The effects of the four parameters were evaluated through a full-factorial DoE with two or three levels, depending on the parameter (Table 1).

This resulted in 26 experiments performed autonomously up to sample collection. The samples were analysed by ¹H NMR, and

Table 1. Parameter space probed during the DoE for the transformation of alcohol **7** into ketone **8**.

Parameter	Unit	Min	Max	Levels
NaBr	equiv	0.1	0.3	2
NaOCl	equiv	1	1.2	2
Toluene	equiv	0	2	2
Temperature	°C	30	60	3

a model describing with high confidence the response of the system over the explored domain was obtained (Fig. 6). While this example demonstrated the usefulness of our automated workflow, the main remaining bottleneck was the manual and time-intensive offline ^1H NMR analysis.

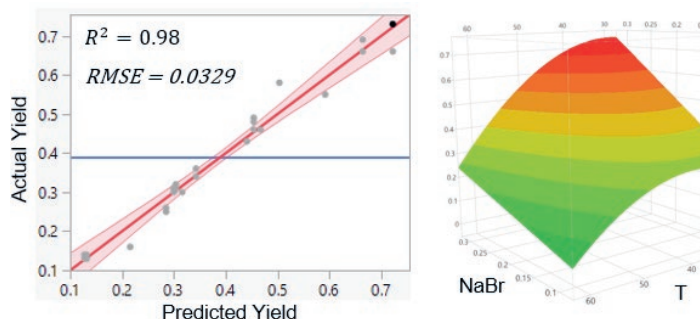
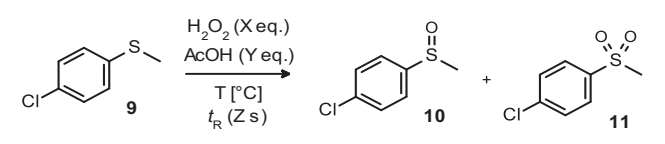


Fig. 6. (Left) Actual vs. Predicted response for yield of ketone **8**. (Right) Response surface of yield vs. temperature and NaBr equivalency.

In a subsequent iteration of this workflow, the oxidation of a sulfide **9** to its respective sulfoxide **10** was optimised (Scheme 4).



Scheme 4. Oxidation of sulfide **9** to sulfoxide **10**. Sulfone **11** is the over-oxidation by-product to be minimised. The ISTD used was benzonitrile.

Aiming at removing the analytical bottleneck of the previous example, the associated DoE was executed with a different experimental setup (Fig. 7), including a sampling valve linked to an UHPLC system to perform analytics. After addition of a quench medium and an ISTD, the reaction mixture flowed through the sampling valve. Upon reaching steady state, the valve is switched to injection and a trigger signal is sent to the UHPLC system to start an analysis, before the system switches to the next set of conditions.

Also in this case, a full-factorial DoE aiming at optimising H_2O_2 equivalents, AcOH equivalents, temperature, and residence time was performed in 26 experiments. The trials were randomised, except for temperature, to avoid heating/cooling cycles. The full experimental program (Fig. 8), including liquid chromatography (LC) analysis, was completed in less than 3 hours in an

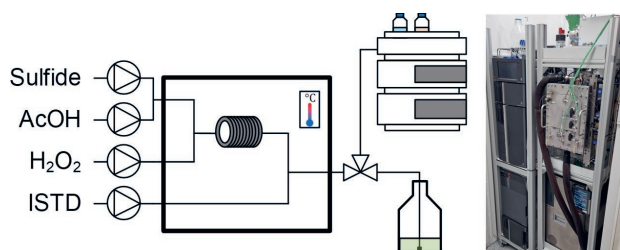


Fig. 7. Setup used for the optimisation of the oxidation of sulfide **9** to its corresponding sulfoxide **10**.

unattended fashion. A 94% yield of sulfoxide **10** could be obtained with the following reaction conditions: 1.2 equivalents of H_2O_2 , 5 equivalents of AcOH at 80°C , with 20 seconds of residence time.

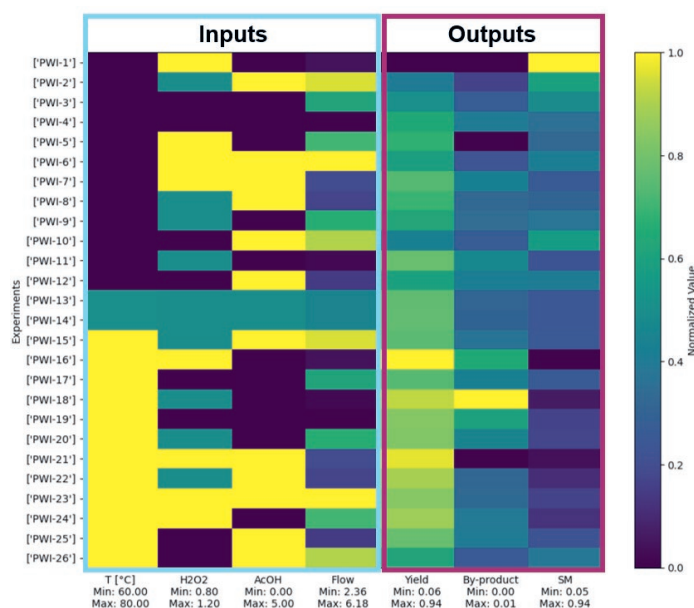


Fig. 8. Colormap of DoE conditions (inputs) tested for optimising the oxidation of sulfide **9** to sulfoxide **10**, and the corresponding results (outputs).

This automated workflow can be extended beyond DoE to a fully closed-loop optimisation, where experimental outcomes directly inform the selection of subsequent conditions without human intervention. This requires sufficiently reliable analytical data to ensure an optimiser can select the next experimental conditions correctly.

3.2 Moving to Closed-Loop Optimisation: Reaching the Desired Data Quality

The analytical data collected during process development plays a central role: it allows the building of an understanding of the system and supports robust scale up of a process from the laboratory to manufacturing. However, a model is only as good as the data it is built from. Automating the analysis of reaction mixtures is challenging, as unexpected perturbations can arise during sample preparation (*i.e.* inhomogeneity due to solids or immiscible liquids) or during analysis itself (*i.e.* overlapping peaks, detector response outside the calibrated range). Among the analytical methods used in process development, LC is one of the most widely adopted as it is very precise, provided it operates within its validated range.

Several reports of self-optimising flow platforms using automated LC analysis have been published.^[13f-h,15,16] While these are impressive demonstrations, the analytical data-processing steps were tailored to the specific reactions being studied. In an industrial setting, a highly automated workflow that still requires extensive pre-work on analytical data acquisition and careful manual validation of generated data is less desirable. Some of these challenges have been addressed with the release of the MOCCA (Multivariate Online Contextual Chromatographic Analysis) Python library,^[17] which provides automated baseline correction, peak deconvolution, and compound tracking for liquid chromatography-photodiode array detector (LC-DAD) data. However, one limitation still remained: the absence of flexibility with re-

spect to sample concentration to ensure that the detector response falls within its linear range.

In agrochemical process development, productivity is a key performance metric that imposes constraints on concentration. In a typical flow chemistry optimisation campaign, the concentration of the analyte of interest may range from 1 to 250 mg/mL, spanning more than two orders of magnitude. Standard LC-DAD analysis at a single wavelength often fails to provide accurate quantification across this entire range: at high concentrations the detector saturates, while at low concentrations the signal falls below reliable detection limits. Without addressing this, closed-loop optimisation is restricted to qualitative guidance, and quantitative modelling – whether kinetic, DoE, or multi-objective BO – becomes unreliable. To address this without requiring further sample processing, we implemented a multilinear calibration protocol that utilises the full DAD dataset instead of a single wavelength. When using an ISTD, this results in a calibration matrix that enables quantification of compounds at any wavelength where they exhibit absorption. This approach can expand the exploitable concentration range by two orders of magnitude: at high concentrations, wavelengths with low extinction coefficients remain within the linear range, while at low concentrations, wavelengths with high extinction coefficients are exploited (Fig. 9).

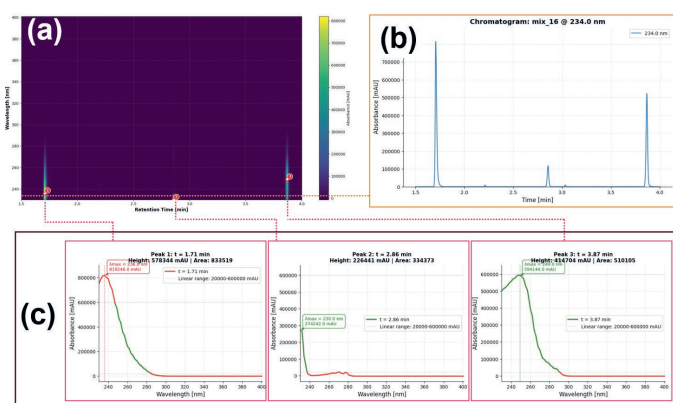


Fig. 9. (a) DAD signal matrix from a UHPLC sample. (b) Chromatogram extracted at 234 nm. (c) UV-Vis spectra of compounds associated with peaks at 1.71, 2.86, and 3.87 min retention time. The spectra fall within the linear range of the detector when the line is green. The substance at 1.71 min cannot be quantified reliably at 234 nm.

The first step of the workflow consists of generating the calibration matrix. The calibration matrix can be obtained by measuring only three samples, which contain a substance **X** and the internal standard ISTD at known concentrations (Fig. 10a). This matrix contains the relative response factors (RRF) for a substance **X** versus the ISTD at every possible pair of wavelengths ($\lambda_x, \lambda_{ISTD}$) across the recorded spectrum. The matrix is then used to analyse reaction mixtures with an unknown concentration of the analyte **X** after addition of the internal standard. To achieve this goal, we developed a Python script which uses the previously generated calibration matrix to output a results matrix. The results matrix provides a quantification estimate for substance **X** for each valid wavelength pair (Fig. 10b). Rather than relying on a single data point, the system generates a concentration estimate for every valid wavelength pair identified in the matrix. Finally, these multiple estimates undergo statistical averaging, allowing significant improvement of the precision and robustness of the final concentration value by mitigating the impact of baseline noise or spectral perturbations.

This automated framework enables reliable, unattended quantification over a concentration range spanning more than two

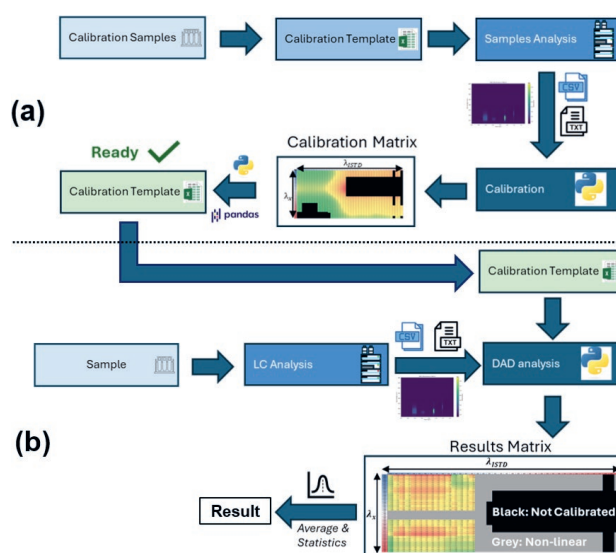


Fig. 10. (a) Multilinear calibration workflow leading to the generation of calibration matrices vs. ISTD. (b) Multilinear analysis workflow relying on the multilinear calibration matrices.

orders of magnitude, providing the high-quality analytical data which is necessary for a fully autonomous, closed-loop reaction optimisation. This approach is currently being deployed internally to help make autonomous process optimisation part of our standard flow chemistry working practices.

Access to higher-quality analytical data enables improved modelling regardless of the framework – kinetic models, DoE, or BO alike. It also lays the foundation for advanced real-time process monitoring, as described in the following section.^[18]

4. Process Analytical Technology (PAT): Transforming Complex Spectral Data into Real-Time Process Control Models

Process Analytical Technology (PAT) is a framework for designing, analysing, and controlling chemical processes through the real-time measurement of critical quality and performance attributes. It relies on a variety of analytical techniques (*i.e.* infrared spectroscopy (IR)), allowing for continuous process understanding and adaptive control. PAT has greatly improved over the past decades through advances in flexibility and robustness of instrumentation, as well as the development and availability of analysis algorithms and tools.^[19,20] However, especially in a process development environment, two main challenges still persist.

First, traditional PAT applications suffer from critical data fragmentation: *in situ* sensors generate real-time data in proprietary software, offline analytical measurements reside in laboratory information management systems (LIMS), and process parameters (*i.e.* temperature, pressure, flow rates, agitation) are usually captured in a separate data acquisition platform. This compartmentalisation hinders holistic process understanding, as each data modality remains isolated, and it is often manually processed by different teams using incompatible tools. As a result, potentially valuable analyses are often not pursued, since the effort required to merge, clean, and analyse these datasets becomes prohibitive.

To address this challenge, Syngenta's digital transformation efforts focus on connecting these previously siloed data streams, through the development of proprietary analytical data management systems, laboratory-based process orchestration systems, and tailored electronic lab notebooks. This integrated data foundation will enable multiple applications, including PAT, by reduc-

ing the effort required to harmonise data and accelerating end-to-end analysis and process understanding.

Second, the timeline required to develop a PAT method often exceeds the length of a process optimisation campaign. This challenge and the associated solutions will be summarised in the next sections.

4.1 Streamlining PAT Method Development

Robust qualitative and quantitative PAT method development can be a lengthy process and the availability of sufficiently trained personal to analyse data manually can become a bottleneck especially during early process development when conditions change fast. Within the PAT team at Syngenta, we focus on improvements to address this method development challenge.

Traditionally, the creation of calibration models linking in-process measurements to final product quality attributes relies on Partial Least Squares (PLS) regression. This method can predict a property of interest starting from many related measurements by combining them into a few weighted factors. To achieve acceptable predictive accuracy, a PLS regression requires many calibration samples (typically 30–100 samples spanning the expected range of product variation), as well as iterative model refinement. Consequently, the timeline for the development of a single PLS model often extends weeks to months, involving designed experiments to generate calibration samples, off-line reference analyses, spectral acquisition, and iterative model optimisation. This makes PLS regression poorly suited for chemical process development, where reaction conditions can change fast; also the transferability of the generated model from laboratory to plant scale might be limited.

An alternative is traditional peak integration, which however becomes unreliable when structurally similar compounds produce overlapping signals. While Indirect Hard Modelling (IHM) techniques can handle peak overlap, they require substantial expertise and time investment for model development.

Multivariate Curve Resolution (MCR) offers substantial time-efficiency and robustness advantages over these approaches in the case of spectroscopic reaction monitoring. MCR techniques have been known for a long time,^[21,22] as they offer an automated way to deconvolute pure component spectra and concentration profiles from the spectrum of an unknown mixture of chemicals. The MCR problem can be formulated as Nonnegative Matrix Factorization (NMF): $X \approx WH$, where X is the matrix of the mixture spectra, W represents the pure spectra and H contains concentrations. Despite the power of the method, MCR has not been widely adopted in the PAT community. Two main factors likely contribute to this: the limited availability of versatile, cost-effective software, and a lack of confidence in the obtained results due to the inherent ambiguity of the mathematical problem (*i.e.* many combinations of pure spectra and concentrations could fit the experimental data equally well). Furthermore, the effectiveness of NMF-based MCR analysis depends heavily on both algorithmic choice and experimental design.

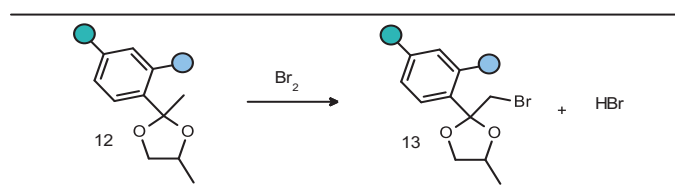
In a recently submitted paper^[23] we provided a comprehensive comparison of three NMF algorithms, including the classical and wide-spread Alternating Least Squares (ALS), the Fast Projected Gradient Method (FPGM) and NMF minimum volume regularisation (MinVol), across different datasets. Our study found that the latter two consistently outperform the classical ALS algorithm, especially in datasets having poor signal to noise ratios (*i.e.* detection of reaction intermediates with minimal accumulation). To facilitate the broader adoption of the MCR methodology within the PAT community, we first derived practical guidance on how to use offline concentration measurements as constraints, in order to maximise the model accuracy, robustness and interpretability. In addition, we implemented the three algorithms in a new open-source Python package, *McrNmf*,^[24]

enabling their straightforward integration into automated data processing workflows.

4.2 Bridging Laboratory PAT Methods and Industrial-Scale Monitoring

The real power of the NMF algorithm described above emerges once it is embedded into larger, automated data pipelines, in which raw spectral data is automatically collected, pre-processed, analysed and finally converted into a process monitoring signal without requiring manual intervention at each step. Building such workflows in Python, rather than relying on commercial software, gives us the flexibility to rapidly chain together modular algorithms and tailor them to the specific needs of a project. This is illustrated in the following example.

The synthesis of a fungicide active ingredient includes the bromination of compound **12** with bromine to yield the mono-brominated target compound **13** (Scheme 5). Undesired dibrominated compounds can form in the case of poor stoichiometric control, leading to out-of-specification batches.



Scheme 5. Bromination of **12** leading to the mono-brominated compound **13** with dibrominated species as by-products.

While critical, reliably detecting the reaction endpoint in real time had for a long time been challenging. Standard inline measurement methods were either insufficiently selective or incompatible with the process equipment. A non-contact Raman probe (993 nm) was the only viable option, but this produced data with a poor signal-to-noise ratio (Fig. 11). The nature of the transformation and the inability to build a meaningful calibration design space ruled out classical calibration approaches such as PLS. A simpler multivariate method (Principal Component Analysis, PCA) in combination with spectral pre-treatment was also attempted but lacked the sensitivity needed to reliably detect the quality specification limit.

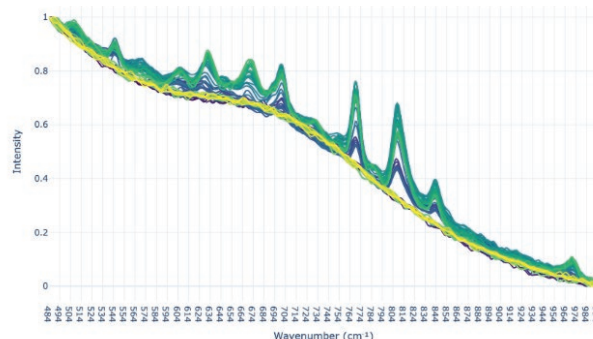


Fig. 11. Raman spectral raw data over the course of the reaction.

We therefore developed a custom Python pipeline to address this challenge (Fig. 12). First, the raw spectral data was pre-processed for correcting the noisy baseline. Classical baseline correction algorithms (*i.e.* Asymmetrically Reweighted Penalized Least Squares, ARPLS, which is known to work well with Raman

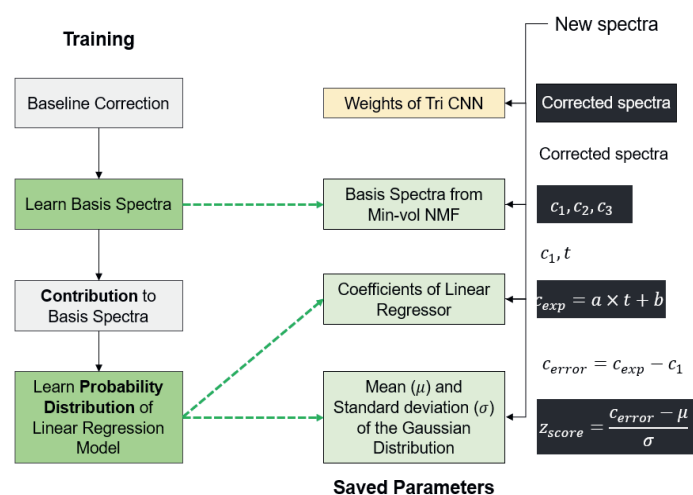


Fig. 12. Training and monitoring data pipeline for end of reaction monitoring.

data)^[25] also cut off low intensity signals. In contrast, a recently published neural network-based algorithm (triCNN)^[26] worked well, preserving weak but chemically meaningful signals. Second, the MinVol NMF algorithm described in the previous section was applied to extract the dominant spectral features from the noisy mixture data. This revealed a component whose signal appears specifically at the end of the reaction and serves as a natural monitoring indicator, even though its complete chemical identity remains unknown.

When a reaction is proceeding normally, the extracted signal follows a predictable linear trend throughout the bromine dosage. Deviations from this expected behaviour (quantified as a z-score) will serve as a real-time monitoring indicator, cutting off the dosage when the specification limits are reached.

This pipeline successfully transformed a complex, noisy dataset into a simple yet sensitive real-time monitoring model which could not have been achieved at present times with standard commercial software. We recognise that maintaining custom code comes with its own challenges, and we therefore view Python-based workflows not as a replacement for commercial tools, but as a flexible extension of the analytical toolbox, best suited for rapid prototyping and for enabling advanced analyses where standard solutions fall short. Large language models (LLMs) can help the user to more easily build their own data pipelines through vibe-coding. However, we do not recommend this approach for applications where high confidence and reproducibility are needed. Instead, building subject specific custom AI agents could be a good way forward.

5. Conclusions

This perspective illustrates how targeted digital tools address concrete bottlenecks across process development stages, from early-stage reaction optimisation to real-time monitoring at manufacturing scale. Bayesian optimisation efficiently navigates large reaction spaces with minimal experimental effort. Laboratory automation, combined with multilinear calibration, reduces hands-on time by approximately 85% and lays the foundation for autonomous closed-loop optimisation. Advances in NMF algorithms and modular Python pipelines lower the barrier to real-time PAT monitoring, even in challenging industrial settings. Together, these tools form a connected pipeline: BO identifies promising conditions, automation enables the generation of large volumes of high-quality data – both to feed iterative optimisation and to build robust models – and PAT translates that data into real-time monitoring at scale. This continuity – from discovery to manufacturing – is the core of our ongoing digital transformation.

Received: May 14, 2026

- [1] a) S. Desimpel, M. Dorbec, K. M. Van Geem, C. V. Stevens, *Chem. Soc. Rev.* **2026**, 55, 2731, <https://doi.org/10.1039/D5CS00962F>; b) Y. Wu, A. Walsh, A. M. Ganose, *Digital Discovery* **2024**, 3, 1086, <https://doi.org/10.1039/D3DD00234A>; c) J. Guo, B. Rankovic, P. Schwaller, *CHIMIA* **2023**, 77, 1, <https://doi.org/10.2533/chimia.2023.31>.
- [2] V. Czitrom, *Am. Stat.* **1999**, 53, 126, <https://doi.org/10.1080/00031305.1999.10474445>.
- [3] B. J. Wall, M. T. Koeritz, L. M. Stanley, B. VanVeller, *ACS Catal.* **2025**, 15, 8885, <https://doi.org/10.1021/acscatal.5c01626>.
- [4] S. Felten, M. Shevlin, M. H. Emmert, 'Enabling Tools and Techniques for Organic Synthesis: A Practical Guide to Experimentation, Automation, and Computation', John Wiley & Sons, Inc., New Jersey, **2023**, Ch.6.
- [5] E. Braconi, *Nat. Rev. Methods Primers* **2023**, 74, <https://doi.org/10.1038/s43586-023-00266-3>.
- [6] a) D. P. Gutierrez, L. M. Folkmann, H. Tribukait, L. M. Roch, *CHIMIA* **2023**, 77, 12, <https://doi.org/10.2533/chimia.2023.7>; b) C. J. Taylor, A. Pomberger, K. C. Felton, R. Grainger, M. Barecka, T. W. Chamberlain, R. A. Bourne, C. N. Johnson, A. A. Lapkin, *Chem. Rev.* **2023**, 123, 3089, <https://doi.org/10.1021/acs.chemrev.2c00798>.
- [7] E. Braconi, E. Godineau, *ACS Sustainable Chem. Eng.* **2023**, 11, 10545, <https://doi.org/10.1021/acssuschemeng.3c02455>.
- [8] Q. Yang, Y. Zhao, D. Ma, *Org. Process Res. Dev.* **2022**, 26, 1690, <https://doi.org/10.1021/acs.oprd.2c00050>.
- [9] B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams, A. Doyle, *Nature* **2021**, 590, 89, <https://doi.org/10.1038/s41586-021-03213-y>.
- [10] <https://www.kebotix.com/> (accessed March 2026).
- [11] <https://reagents.acscipr.org/reagent-guides/n-alkylation-reagent-guide/list-of-reagents/phase-transfer-catalysis/>, accessed March 20, 2026.
- [12] A. Pomberger, A. A. Pedrina McCarthy, A. Khan, S. Sung, J. Taylor, M. J. Gaunt, L. Colwell, D. Walz, A. A. Lapkin, *React. Chem. Eng.* **2022**, 7, 1368, <https://doi.org/10.1039/D2RE00008C>.
- [13] a) A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne, A. A. Lapkin, *Chem. Eng. J.* **2018**, 352, 277, <https://doi.org/10.1016/j.cej.2018.07.031>; b) A. Slattery, Z. Wen, P. Tenblad, J. Sanjosé-Orduna, D. Pintossi, T. den Hartog, T. Noël, *Science* **2024**, 383, 6681, <https://doi.org/10.1126/science.adj1817>; c) F. L. Wagner, P. Sagmeister, T. G. Tampone, V. Manee, D. Yerkozhanov, F. G. Buono, J. D. Williams, C. O. Kappe, *ACS Sustainable Chem. Eng.* **2024**, 12, 10002, <https://doi.org/10.1021/acssuschemeng.4c03253>; d) S. Knoll, C. E. Jusner, P. Sagmeister, J. D. Williams, C. A. Hone, M. Horn, C. O. Kappe, *React. Chem. Eng.* **2022**, 7, 2375, <https://doi.org/10.1039/D2RE00208F>; e) P. Sagmeister, F. F. Ort, C. E. Jusner, D. Hebrault, T. Tampone, F. G. Buono, J. D. Williams, C. O. Kappe, *Adv. Sci.* **2022**, 9, 2105547, <https://doi.org/10.1002/advs.202105547>; f) E. Rial-Rodríguez, F. L. Wagner, T. Fuchß, A. Sommer, J. Krieger, H.-M. Eggenweiler, C. O. Kappe, G. Laudadio, *Chem* **2026**, 102952, <https://doi.org/10.1016/j.chempr.2026.102952>; g) D. Cortés-Borda, E. Wimmer, B. Gouilleux, E. Barré, N. Oger, L. Goulamaly, L. Peault, B. Charrier, C. Truchet, P. Giraudeau, M. Rodríguez-Zubiri, E. Le Grogne, F.-X. Felpin, *J. Org. Chem.* **2018**, 83, 14286, <https://doi.org/10.1021/acs.joc.8b01821>; h) A. D. Clayton, E. O. Pyzer-Knapp, M. Purdie, M. F. Jones, A. Barthelme, J. Pavey, N. Kapur, T. W. Chamberlain, A. J. Blacker, R. A. Bourne, *Angew. Chem. Int. Ed.* **2023**, 62, e20214511, <https://doi.org/10.1002/anie.20214511>; i) A. Senthil Vel, Y. Horbenko, N. El Sabbagh, B. Groh, D. Cortés-Borda, J. Farjon, P. Giraudeau, J.-N. Dumez, J.-P. Krieger, F.-X. Felpin, *React. Chem. Eng.* **2026**, 11, 1274, <https://doi.org/10.1039/d5re00504c>.
- [14] M. B. Plutschack, B. Pieber, K. Gilmore, P. H. Seeberger, *Chem. Rev.* **2017**, 117, 11796, <https://doi.org/10.1021/acs.chemrev.7b00183>.
- [15] M. I. Jeraal, S. Sung, A. A. Lapkin, *Chem. Methods* **2021**, 1, 71, <https://doi.org/10.1002/cmt.202000044>.
- [16] K. Simon, P. Sagmeister, R. Munday, K. Leslie, C. A. Hone, C. O. Kappe, *Catal. Sci. Technol.* **2022**, 12, 1799, <https://doi.org/10.1039/d2cy00059h>.
- [17] C. P. Haas, M. Lübbsmeyer, E. H. Jin, M. A. McDonald, B. A. Koscher, N. Guimond, L. Di Rocco, H. Kayser, S. Leweke, S. Nidenführ, R. Nicholls, E. Greeves, D. M. Barber, J. Hillenbrand, G. Volpin, K. F. Jensen, *ACS Cent. Sci.* **2023**, 9, 307, <https://doi.org/10.1021/acscentsci.2c01042>.
- [18] 'Process Analytical Technology', Eds. K. A. Bakeev, J. C. Menezes, **2010**, p. 521.
- [19] L. L. Simon, H. Pataki, G. Marosi, F. Meemken, K. Hungerbühler, A. Baiker, S. Tummala, B. Glennon, M. Kuentz, G. Steele, H. J. M. Kramer, J. W. Rydzak, Z. Chen, J. Morris, F. Kjell, R. Singh, R. Gani, K. V. Gernaey, M. Louhi-Kultanen, J. O'Reilly, N. Sandler, O. Antikainen, J. Yliruusi, P. Froberg, J. Ulrich, R. D. Braatz, T. Leyssens, M. von Stosch, R. Oliveira, R. B. H. Tan, H. Wu, M. Khan, D. O'Grady, A. Pandey, R. Westra, E. Delle-Casse, D. Pape, D. Angelosante, Y. Maret, O. Steiger, M. Lenner, K. Abbou-Oucherif, Z. K. Nagy, J. D. Litster, V. K. Kamaraju, M.-S. Chiu, *Org. Process Res. Dev.* **2015**, 19, 3, <https://doi.org/10.1021/op500261y>.
- [20] S. J. Mazivila, J. L. M. Santos, *TrAC, Trends Anal. Chem.* **2022**, 157, 116698, <https://doi.org/10.1016/j.trac.2022.116698>.

- [21] A. de Juan, R. Tauler, *Anal. Chim. Acta* **2021**, *1145*, 59, <https://doi.org/10.1016/j.aca.2020.10.051>.
- [22] W. H. Lawton, E. A. Sylvestre, *Technometrics* **1971**, *13*, 617, <https://doi.org/10.1080/00401706.1971.10488823>.
- [23] S. A. Vasudevan, R. Kohlbrenner, N. Ritschel, T. Vent-Schmidt, *ChemRxiv* **2025**, <https://doi.org/10.26434/chemrxiv-2025-qs5g4>.
- [24] <https://github.com/siddarthVasudevan/mcmmf>, accessed May 13, 2026.
- [25] S.-J. Baek, A. Park, Y.-J. Ahn, J. Choo, *Analyst* **2015**, *140*, 250, <https://doi.org/10.1039/c4an01061b>.
- [26] T. Chen, Y. Son, C. Dong, S.-J. Baek, *Analyst*, **2025**, *150*, 2653, <https://doi.org/10.1039/d5an00253b>.

License and Terms



This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (<https://chimia.ch/chimia/about>).

The definitive version of this article is the electronic one that can be found at <https://doi.org/10.2533/chimia.2026.583>