Received: 10 September 2024 | Revised: 6 December 2024 | Accepted: 9 December 2024

DOI: 10.1002/cjce.25601

RESEARCH ARTICLE

THE CANADIAN JOURNAL OF CHEMICAL ENGINEERING

Experimental methods in chemical engineering–Validation of steady-state simulation

Caroline Brucel | **Émilie Thibault** | **Gregory S. Patience** | **Paul Stuart**

Chemical Engineering, Polytechnique Montréal, Montréal, Québec, Canada

Correspondence
Paul Stuart, Chemical Engineering, Polytechnique Montréal, C.P. 6079, Succ. "CV", Montréal, H3C 3A7 QC, Canada.
Email: paul.stuart@polymtl.ca

Abstract

Steady-state simulation (Aspen, PRO/II, WinGEMS, CADSIM Plus) guides equipment selection, operating conditions, and optimization to design chemical processes like Kraft pulping, specialty chemicals, and petrochemical complexes. Ensuring that the simulation characterizes the yields, heat transfer loads, purity, utilities demand, and profitability requires data that represents the physicochemical and transport properties of each stream and unit operation. Here, we present strategies to validate steady-state simulations against plant data and expectations from operators. To build and validate simulations requires real-time data, but errors contaminate measurements and dynamic conditions—start-up, shut-downs, process upsets—compromise fidelity. A pre-treatment step removes incongruous data to build the simulation on process conditions representative of steady-state. Working through the process with experts (informal validation) and comparing simulation results with plant data (formal validation) reduces gross error with an objective to achieve a simulation accuracy to within one standard deviation of measurement variability. A bibliometric review highlights the limited focus on steady-state simulation validation in the field of process engineering. Most articles mention accuracy but neglect to describe how it is evaluated. Despite this scarcity, validation remains a critical factor in various domains of chemical engineering research. Interviews with professionals offer a practical perspective on the applications of simulation in an industrial context like process monitoring, equipment performance analysis, operator training, and decision-making. Finally, a case study demonstrates how to implement data treatment and validation for Kraft mill brownstock washing department: Applying multiple validation techniques increases the value and confidence in the simulation.

KEYWORDS

industrial process simulation, practical validation, simulation validation, steady-state simulation

1 | INTRODUCTION

Process simulators are chemical engineering tools to design and operate plants to within tolerances that maximize profitability while meeting environmental and safety objectives. However, their effectiveness lies solely in their ability to accurately replicate the reality of the process. Some simulation software can characterize time-varying processes, for the design and optimization of expected unsteady process behaviours include batch and transient operations. Specialty chemicals and pharmaceuticals produced in batches and processes like Kraft pulping can include batch-continuous modes of production. Batch operating can be difficult to model as it displays both continuous and discrete dynamics. Simulations need to replicate the sequential operating logic and control in addition to representing the successive integrated operations occurring. Their simulation at the plantwide level is not an industry standard due to this complexity, but reducing their downtime (production bottleneck) and optimizing their scheduling improves their performance and reduce their operating costs.[1] Process dynamics, where there is unsteady-state(transient) behaviour in continuous operations, happen during special situations like start-ups and shutdowns, process disturbances, and planned operation changes (moving from one grade of product to another). Restoring and maintaining a process at desired operating conditions while adhering to safety, economic, environmental, and quality requirements is the objective of process control. Process control systems include a wide range of activities, from the measurement and actuation of control valves on a short time scale to analysis of process data to update controller set points and planning that considers economic projections over months of operation.[2] Dynamic simulation can be employed to evaluate alternative control strategies to better respond to those unsteady-state conditions. Control systems suggested for processes with prevalent unstable dynamics and dead time were tested in simulations improve performance measures compared to usual control techniques.[3–5] However, that their creation requires more intensive effort, process data, and equipment specifications makes them impractical and expensive to model on a plantwide scale.[6]

Steady-state simulators are cost-effective design tools and resources that operators rely on to make daily decisions and develop long-term scenarios to de-bottleneck process steps to maximize productivity. Validation refers to achieving a defined expectable level of accuracy based on the uncertainty of the process measurements. For instance, when the measurement variability of feedstock flow meters is in the ±3% range, the expectation of the simulation results can only be expected to be within a margin of 3%. However, it is more likely that this accuracy may exceed 50% of this value (e.g., ±5%). An orifice flowmeter's measurements are based on three factors: pressure drop, fluid density, and instrument geometry, with the volumetric flowrate depending on the product of the first two variables. Uncertainty propagation in measurements contributes to the uncertainty in the predicted simulation. The uncertainty of any function f (like production rate) is the square root of the sum of the squares of the product of the partial differential with respect to the factor, x_i, and its uncertainty, Δ_i:

$$\Delta_f^2 = \left(\frac{\partial f}{\partial x_1}\Delta_1\right)^2 + \left(\frac{\partial f}{\partial x_2}\Delta_2\right)^2 + \left(\frac{\partial f}{\partial x_3}\Delta_3\right)^2 + \ldots + \left(\frac{\partial f}{\partial x_n}\Delta_n\right)^2$$

For functions that are a product of factors raised to any power, a_i, ($f = x_1^{a1}\, x_2^{a2} \ldots\, x_n^{a3}$) this equation simplifies to the following:

$$\frac{\Delta_f}{f} = \sqrt{\sum_{i=1}^{n}\left(\frac{a_i}{x_i}\Delta_i\right)^2}$$

When the ratio of Δ_j/x_j is three times greater than the other factors, then the contribution of the other factors is no better than 10% of this single factor, and thus is ignored. Patience[7] states that "as a general guideline for functions that are products of factors, as in the case of an orifice flowmeter, a factor f_2 is negligible when:

$$\frac{a_2}{x_2}\Delta_2 < \frac{1}{3}\frac{a_1}{x_1}\Delta_1$$

and for functions that are sums of factors, it is negligible when":

$$a_2\Delta_2 \leq \frac{1}{3}a_1\Delta_1$$

This uncertainty affecting measurement quality is in addition to data scarcity: even the most modern mills do not measure every variable. Engineers are required to make assumptions to close mass and energy balances to simulate the process. The large number of degrees of freedom around the process also contribute to uncertainty. Validating a simulation requires having access to knowledge or real-time data, and the techniques to compare it to the process it represents.

Distributed control systems have logged data electronically since the 1960s, but the number of variables and the quantities of data available have increased exponentially. Since the early 2000s, data management systems have centralized these results, making them easier and faster to store, access, and manipulate. The access to huge amounts of real-time process data has improved the way simulations are built. The kinds of real-time data gathered and the means of gathering and processing them impact simulation construction and validation.

This tutorial-review on simulation validation is part of a series of articles dedicated to experimental methods in chemical engineering.[8] Simulation users, whether in an industrial or a research context, benefit from expanding and diversifying their knowledge of validation techniques and form a better understanding of their applicability to industrial process simulation. First, the theory section reviews validation concepts and what aspects of a model are validated. The evolution of industrial process simulations and their validation are examined, focusing on aspects that are still present, or that have had an impact on advancing the current techniques. Next, the state of process simulations and their validation today, with the changes brought by modern data management present in all industrial processes, are discussed. In the third section, a survey of articles published about industrial process simulation validation and the applications of steady-state process simulation validation differentiate the major themes researched in the industrial sector. Opinions gathered from interviews with industrial modellers and mill engineers are also included. Finally, a case study of validating a pulp mill washing department simulation is presented, including the conducted validation and result analysis.

2 | THEORY

A process model mathematically characterizes the mass and energy balances around each unit operation of a chemical plant. These models are expressed as systems of non-linear equations. A simulation executes the model with numerical values representing the conditions of the process in the equations.[9] Engineers apply steady-state simulation in varied circumstances: In the case of a complex industrial setting, the whole process is evaluated to operate the plant or to consider alternative design. Validation confirms the extent to which a simulation accurately represents the process. Data and process information are fundamental elements to build a simulation, especially during the validation step. With the availability of distributed control systems and work-stations at plant sites, manual measurements of operating conditions have been gradually displaced. However, operators continue to sample streams manually for analysis in the laboratory. The frequency varies a couple of hours to daily sampling, which introduces uncertainty when considering either steady state or transient conditions. Accounting for variability is an underlying criterion to validate steady-state simulation.

2.1 | Simulation validation concepts

The concept of verification, validation, and accreditation (VV&A) is applicable in many scientific fields, including chemical process engineering.[10] Sargent[11] proposed a simulation validation framework that comprises verifying the validity of the relationships between the conceptual model, the computerized simulation, and the process (referred to as the system-Figure 1). These three elements are linked by model development: the conceptual model is based on the real-world process, the simulation is the implementation of the conceptual process, and experimentation is conducted on the simulation to compare it to the process.

The aspects of simulation validity comprise (1) conceptual model validity, (2) computerized model verification, and (3) operational validity. Considering these three criteria and their connections improve the simulation validation outcomes. (1) Conceptual model validation established suitable theories and assumptions constituting the conceptual model. (2) Computerized simulation verification confirms that the program was implemented correctly. (3) Operational validity evaluates that the simulation outputs meet the accuracy targets for the simulation's intended purpose over the domain of its intended applicability–that the simulation will be able to represent the process behaviours that it was planned to represent, including various operating conditions.

Data validity has been defined as ensuring that the data necessary for simulation building, evaluation, and testing and conducting experiments to solve the problem are adequate and correct.[11] It impacts all other validation and verification activities and is required to build the conceptual model. Sufficient, appropriate, and accurate data about the process are needed to formulate theories and logical relationships, both to build the conceptual model and to test its assumptions. Operational validation relies on coherent information on process outputs

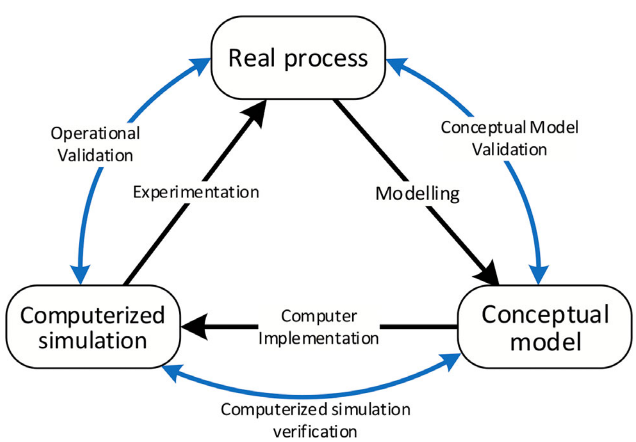

FIGURE 1 Relations between modelling and the real process.

(behavioural data)–yield, energy demand, productivity, remediation load, and waste treatment: the outputs of the simulation must be compared to those of the real process.

Verification is defined as 'the process of ensuring that the model design (conceptual model) has been transformed into a computer model with sufficient accuracy'[12] or as 'the process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model'.[13] Conceptual model validation examines the basic unit operations and streams before conducting any simulation.[14] These concepts define what and when to validate a simulation, but not how. Engineers apply face validation, graphic validation, and sensitivity analysis for steady-state processes.[15]

It is possible to increase the confidence that a simulation represents the process by validating it. Further validation increases confidence in the simulation but comes at an exponential increase in cost and effort. A simulation must be validated for its intended purpose.[16]

In the case of modern process engineering, conceptual model validity and operational validity are the most relevant.[9] Engineers select validation techniques to assess the simulation at different stages of the simulation building methodology, from fundamental concepts and principles of the process, verification of the layout, and finally validation of the simulation results. The effectiveness of operational validation depends on the quality and quantity of process data. Data validity needs to be integrated into the simulation building and validation process because insufficient data and faulty measurements render it impossible to build a reliable simulation.

2.2 | Simulation of industrial processes: Historical perspective

Simulation has been a topic of research in many domains and has evolved along with technological advances in computers. Computer simulation for chemical engineering emerged in the 1950s with the development of single process-unit models. In the following decade, modular simulation appeared, which consisted of individual process-unit models (modules), pre-programmed in the software, that could be linked together into networks.[17] In parallel, industrial processes also became more complex, the instrumentation to control these processes improved, and systems to gather and store measurements were developed.

2.2.1 | Data availability

Data directly and automatically measured, such as flow, temperature, and pressure has been largely gathered from industrial processes, but the monitoring of more specific process variables often relied on manual sampling and laboratory measurements. Before the 1960s, the absence of widespread computerization and data storage meant that few measured values were maintained in a historian.[18] The state of industrial data gathering and storage has evolved since the introduction of more accessible and powerful computer systems.[19]

Before today's easy access to low-cost data storage, plant personnel chose which variables to conserve. Now, data historians collect, compress, store, validate, and preserve instrument and sensor measurements.[20] Earlier data collection and storage systems were part of the control system; they provided only a few days' worth of evenly spaced samples, and only daily or weekly averages were kept long-term. Complementary techniques to archive more data were introduced in the 1980s[20]: exception reporting, which reduces the stream of information coming from controllers, data loggers, and distributed control system (DCS) to the data historian; and compression, which then reduces the data storage requirements in the data historian. Data exception reporting reduces the number of data points sent to the historian by filtering the data recorded at the source by exception deviation limits set to a percentage of span of the particular variable and minimum and maximum times; a new value is sent to the historian if, beyond a certain minimum time, the change in the variable's value exceeds the set limit or if the maximum time is exceeded. This ensures both that excessive numbers of values of a rapidly oscillating variable are not stored (by means of the minimum time) and that more stable periods are still recorded regularly, and that the sensor is still working properly (by means of the maximum time). Exception recording alone reduces the quantity of information sent to the historian by over 95% (Figure 2). However, at plant scale and over time, this still represents a large accumulation of data: plants contain thousands of sensors measuring analogue real time points continuously. Compression algorithms minimize the amount of storage space while retaining the required accuracy and fast access to historical data. These work similarly to exception reporting, but taking advantage of considering the dataset as a whole time series rather than as it is measured, removing intermediary data points if they do not represent a significant change from the overall trend (Figure 2A). The boxcar algorithm is based on compression deviation (Figure 2B). A line is drawn through the first two points of a sequence, and two bands are drawn above and below it. Each new point inside the deviations extends this 'box' until a point outside is found (and the preceding point is stored), and a new box is then formed using the latest point and the previously stored one. A backslope correction verifies that

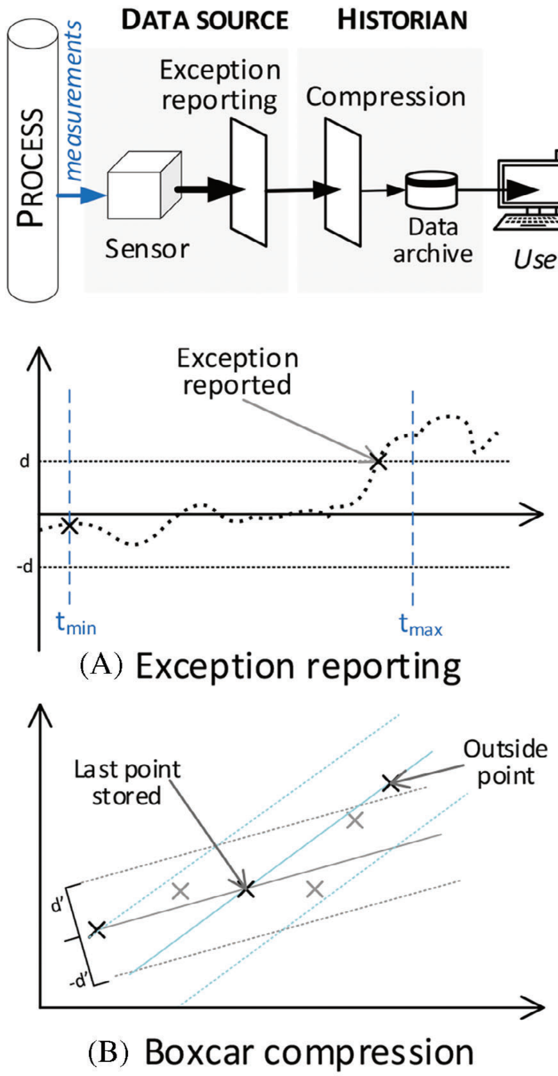

FIGURE 2 Historical real-time process data, from the process to the user and the effects of (A) exception reporting and (B) boxcar compression on the process measurements.

only points representative of the trend are kept. The archived data that have been treated by exception reporting at the data source and by compression in the data historian can now be extracted from the archive.

Although the archiving process limits the amount of data available, not all data are representative of the process because of instrument accuracy and errors present in the measurement signal. Errors in process data are inevitable and categorized as either random or gross errors.[21–23] Random errors are usually of small magnitude and cannot be predicted but are characterized by statistics; on a graphical representation of the measured values of a variable over time, these errors give the signal the appearance of noise. Gross errors are usually of a larger magnitude than random errors.

Both errors degrade data quality, but random errors are characterized by a normal statistical distribution, and its impact is less than that of gross errors. Gross errors are systematic and, if not detected, impact the quality and interpretation of the process data like sensor miscalibration, sensor wear, or fouling; the latter degrades measurement quality over time.[24] The accumulation of both kind of errors degrade the data's representativeness of the process, affecting the performance of plant control systems and the simulations built using these data.[23]

Simulating existing processes relies on their measurements, but the presence of gross error can strongly bias process data.

Data reconciliation (DR) procedures are implemented to correct gross error in measured data adjust their values to respect mass and energy conservation laws and other constraints of a process model. DR consists mainly of a constrained minimization problem, with the objective of minimizing the difference between measured and corrected values. An important step of DR is the identification of gross errors, usually using principle component analysis, statistical tests evaluating measurement distribution or neural networks methods.[25]

DR has been applied successfully to petrochemical plants and refineries[22,26] as well as in smaller processes, like an industrial pyrolysis reactor, to improve data quality and optimize their control.[27] The reactor was modelled using all 36 variables measured in the process and 11 equations. The presence of gross errors was confirmed by a global test and identified using the serial elimination technique. An error-in-variable method reconciled the model and estimated of a missing heat transfer coefficient at the same time. The DR procedure confirmed some of the process assumptions while also correcting a couple of flowrate measurements.

Data reconciliation's limitations include the need for a certain amount of measurement redundancy to be applied. Its application to large industrial processes requires building complex process models, adapted solving algorithms, and handling multiple sources of gross errors.

Both exception reporting and compression reduce the amount (and precision) of the process data recorded, but limited storage capacities made them indispensable to be able to record significant historical data. Mills prefer retaining data that support operations: the number of measurements preserved from the quantity that is gathered and poorly represents the process when the sampling frequency is too low to represent process events. For simulation purposes, periods of stable or pseudo steady-state operation can be difficult to discern from these data, which damages subsequent simulation projects.

Before the advent of larger plant data storage capacities, detailed real-time data was unavailable to modellers, who had to contend with the resources at their disposal or incur additional costs related to data collection, such as field sampling campaigns. This data availability had consequences for the simulation and validation methods, making modellers rely more on the knowledge of process experts and less on process measurements to compare the simulation and the real process.

Data treatment techniques detect and correct these errors, but the ones implemented industrially focus on control and monitoring. Treatment of historical data, especially of bigger datasets, is limited by time constraints. Projects, such as simulations, require adapted extraction and treatment of process data to meet their needs.

2.2.2 | Steady-state process simulation methodology

With the advances in technology since the 1960s, steady-state process models have increasingly been built using modular simulation programs, which contain sector-specific library for process units and chemical component packages (petrochemical, mineral processing, pulp and paper, etc.). These programs reduce the amount of time and effort needed for the engineer to program the process and input parameters, like physical and thermodynamic properties of its components.[9] Many articles describe software packages like Aspen Plus, including recommended modelling strategies and case studies.[28,29]

In an industrial context where every project must be economically justified, each simulation building step must serve a purpose. There are many published simulation modelling methodologies,[30–33] and most are not exclusive to chemical process engineering. Their complexity varies depending on the scope of the project, but most agree on the general steps of defining the project objective, preparing the model, and gathering process information before building the simulation. Ülgen et al.[33] present an extensive general simulation methodology, applicable to the industrial chemical engineering domain (Figure 3). It includes major steps broken down into sub-steps that include technical implementation and the roles of the participating actors: the choice of the type of modelling, the level of detail to include, and the involvement of the modeller(s), process experts, operators, and management are discussed in each step of this methodology.

In the first step of this simulation methodology, the 'Definition of the Problem', the modellers decide whether simulation is the right tool to solve problem surrounding the process considered. The study objectives,

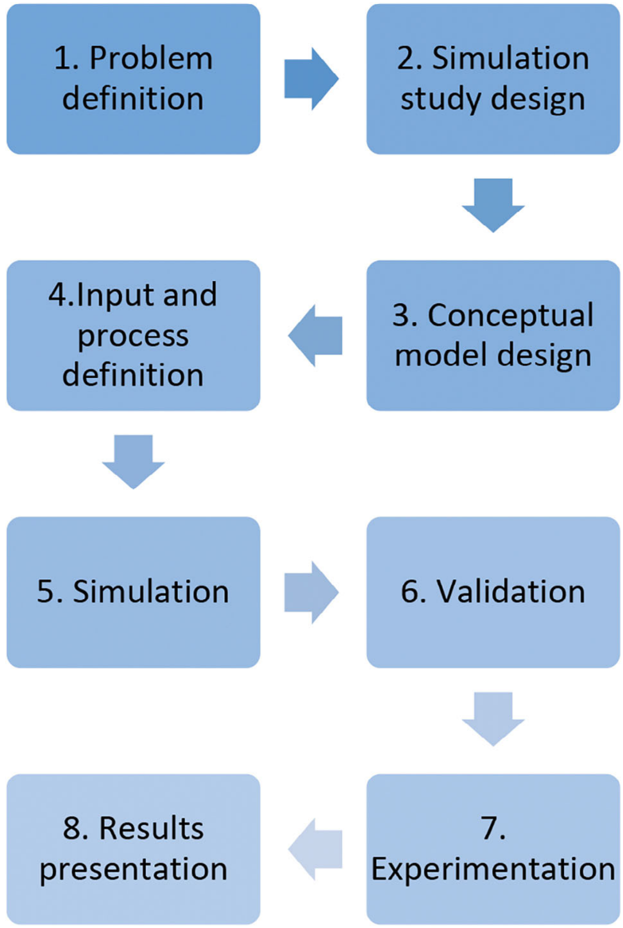

FIGURE 3 Simulation methodology steps.

the issues that need to be addressed, and the boundaries and level of detail are defined. Is a whole plant simulated? A department? Is one model needed, or several? Are all components considered or only measured ones? Does all the equipment present need to be included or only major process units? This step also involves gathering resources, conducting a feasibility study (evaluating the cost–benefit of the simulation), and planning the simulation project; the level of detail of the simulation and the availability of data guide the estimation of tasks durations and budget expenses.

The second step, 'Design of the simulation study', expands on the previous step: the use case of the simulation determines what to include in the model (e.g., solving design or operation problems, working as a training tool). Major assumptions about the process, such as what units and variables to include, are chosen. This list needs to be kept up to date as a reference throughout the simulation building process and be included in the simulation documentation because it explains the overall logic followed by the model.

Usually, a simulation study consists of a base case and additional scenarios to test alternatives. The number of scenarios is either determined at this stage or reevaluated during testing.

The selection of simulation tool selection considers criteria like the complexity of the process, the software available, its costs, the users' experience with it, and the simulation interface.

The planning of data collection includes evaluating the available data, when and how they were collected and what is missing, and how critical these data are for the study. Is it possible to replace data with expert estimates? Data treatment needs must be evaluated to estimate time requirements for the data validation carried out during the fourth step. Measurement uncertainties and their propagation can be evaluated during this step: type of sensor and instrumentation are collected from the in-plant information management system. Alternatively, general instrument accuracies are available in the literature.[34]

Evaluating the human requirements of the simulation study involves identifying experts with knowledge in the specific process being studied, such as process managers, engineers, and other modellers. The data collection phase necessitates the involvement of a process data management system administrator, and coordination between the modellers and the data collector to establish a suitable procedure. Different levels of management and users have different needs to address to guarantee the success of the project: they influence the simulation's presentation, and the information included in its deliverables. The number, milestone dates, level of detail and format of reports, user manual, and training—if necessary—must be identified.

The third step, the design of the conceptual model, encompasses the modelling strategy: deciding how the process will be represented using the simulation tool chosen to meet the simulation study objectives.

Batch or semi-batch operations in a continuous process may require a modelling decision about how it will be represented with the limits of the software. Approximating the process in a software meant for continuous simulation may result in simplification about its operation (scheduling).[28]

The process equipment like tanks, reactors, or washers usually have modules available in the simulation software. If several options exist, or if an exact representation does not, the modeller must make an educated guess. The simulation objectives dictate the level of detail needed, but it is usually easier to start with a simpler model and add details later in the modelling. Using a black box to represent part of the process, on the other hand, must be justified, for example by the study scope, objectives, or the lack of data. More complex operations may need special attention, additional consultation with experts, and data collection to properly simulate them. The modelling strategy should consider these constraints to produce an accurate conceptual model.

The simulation result's numerical outputs are made more comprehensible using graphs or animations. These graphical interfaces should be planned for at this stage.

The last task of this step is the verification of the conceptual model. The users and experts need to approve the modelling assumptions made, the operating logic of the process, and the selected simulation input data. The resulting flowsheet is the one implemented in the simulation software.

The fourth step, the formulation of inputs, assumptions, and process definitions, delves deeper into the actual process to be simulated. This step involves completing the conceptual model by identifying the process operating philosophy (its physical constraints and dynamic events like batch operation), collecting all the data necessary to build the simulation and listing specific assumptions to address the absence of data.

How the process is operated is important to translate into the model: process flow and the logic that describes it, operating choices for specific process units, scheduling breaks, batch sequences, and so forth, all indicate components' travel in the process. If the study deals with large-scale process modifications, physical constraints like equipment placement and distances are pertinent to add.

More in-depth data collection is performed at this stage: process measurements, laboratory data, and other pertinent process information are extracted from data management systems, including raw material and product specifications and equipment operating conditions. Modellers and process experts analyze these data to ascertain their usability, that is, identifying errors or faulty sensors and validating correct values for input variables or parameters that change between alternative scenarios. These raw process data are then treated to remove unwanted values that are not representative of the regimes to simulate. The resulting input values should be validated against process specifications or expert opinions.

A list of model assumptions—including general ones described in the second step, those regarding component behaviour, and unit-specific ones—to translate into the software should be defined to fill in the simulation. These assumptions are based on process specifications, expert knowledge, or literature.

The fifth step is building the simulation. Its procedure relies mainly on the conceptual model, selected process information, and common simulation guidelines and techniques.[29] Commercially available software have their own procedure to guide the building of a new simulation, starting usually with the definition of a component list, from those predefined or options to create new ones, and the selection of methods of property estimation.[35,36] The simulation itself is usually built sequentially, adding process units and connecting them one at a time, dealing with secondary streams and units after having modelled the main process stream, and lastly addressing the tear and recycle streams.

The sixth step, validation, depends directly on the objectives of the simulation study and involves making the simulation sufficiently accurate for its intended purpose.[16] Although validation is presented as a standalone step in methodologies, some of its activities are integrated with preceding steps. Verifying the conceptual model and data validation take place at the same time in the third and fourth steps. Verification of the conceptual model performed by checking that the flowsheet proposed for the simulation corresponds to reality and that the assumptions representing the real process are the right ones for the simulation study. The data for the model inputs may be approved by process experts beforehand. The validation techniques for those activities and for operational validation, executed after the fifth step, should be selected in advance, based on the available resources and the desired accuracy.

In the seventh step, model experimentation explores the simulation capabilities, and the alternative scenarios planned. This step usually involves simulation users, who confirm changes that must be implemented and the conditions under which the simulations are expected to function. This is when the simulation study objectives are tackled, for instance, by performing optimization or testing process changes.

The eighth step is the documentation and presentation of results. Although documentation on the simulation has been completed throughout the building process, the final report summarizes the entire study. This report should include details and explanations of the decisions made during the simulation building process and present the results of experimentation. Details about the simulation life cycle, in the case of a long-term simulation that will be maintained throughout the life of the process, may also be included.

Most methodologies contain the same general steps: problem definition and simulation planning, information gathering, conceptual model creation, simulation, and validation. These steps refine the activities of the previous steps. Simulation methodology is a mature research topic; most differences in proposed methodologies are related to the size of the project and the process studied. Another marked difference is the extent of the validation step: only some of the relevant methodologies consider—or even mention—the iterative nature of the process of building a simulation.

Validation in itself is seldom described in publications. This is partly explained by its dependence on the type of simulation and the objectives of the simulation study. Process measurements contain errors but there is rarely mention of how the data obtained is treated or that data validation as part of process data collection is an integral part of building a simulation.

2.2.3 | Review of simulation validation techniques

The methods comparing simulations to their processes are the same in most fields; researchers have described these techniques with their benefits and limitations. Sargent[15] and Balci[37] have defined validation techniques (Table 1).

Authors classify validation techniques by how they validate,[38] by what they validate,[39] and by what information is available.[40] Sargent distinguishes validation techniques as subjective or objective: objective techniques involve a statistical test or procedure, in contrast to subjective techniques, which rely on expert opinions.[15]

Face validation is a validation technique that is mostly employed in the preliminary stages of simulation building. Potential simulation users and process experts are asked to evaluate the simulation logic, conceptual model, or outputs for representativeness. This has the advantage of involving users in building the simulation, increasing their perception of its validity and their acceptance of the simulation.[41] Another validation technique relying on process experts is the structured walk-through, which targets the conceptual model and its assumptions.[14] Validation techniques where the decisive factor is expert opinion is prone to personal bias.[42]

Direct comparison techniques, like graphic validation, comparison to existing data, and historical data validation, rely on evaluation of the difference between the simulation results and the real process data. Graphical representation of the simulation results is preponderant because it places them in context with the process data, either by displaying those values directly or using measures characterizing the variables being compared (mean, maximum, minimum, etc.).[11]

Sensitivity analysis is the study of how the uncertainty in the model output (numerical or otherwise) is linked to different sources of uncertainty in the model input, using this information to validate simulation behaviour[43] and to compare to existing data such as outputs from other models or from similar existing processes.[14,44] In the context of simulation validation, sensitivity analysis is a method of evaluating the impact

TABLE 1 Common simulation validation techniques.

Technique	Definition
Face validation	Asking individuals knowledgeable about the process whether the model's assumptions or the simulation's behaviour are accurate
Graphic validation	Outputs from the simulation and other references (e.g., industrial data, laboratory experimentation results) are compared on the same graph, which can include statistical information like averages and box plots
Comparison with existing data	Comparing the simulation output with data from other models or similar existing processes
Extreme condition validation	Inputting unlikely combinations of inputs and parameters into the simulation, which should respond in a plausible manner. For example, if the input of a steady-state simulation is null, the output should also be zero. The simulation should only accept realistic values within normal operating range as inputs.
Historical data validation	If historical data from the process exist, part of the data is employed to build the simulation, and the rest, to verify whether it behaves like the process. Using other historical data as input to the simulation should give outputs like those of the real process.
Sensitivity analysis	The study of how the uncertainty in the model output is linked to different sources of uncertainty in the model input.
Traces	These involve following the behaviour of different types of variables throughout the simulation to determine whether the process logic is followed with enough accuracy.

of the uncertainty in a model's inputs on its outputs by changing the values of specific inputs and parameters to evaluate the calculated outputs. The same relationships should be observable in the simulation and the real process. Sensitive parameters identified should be sufficiently accurate to represent the process for the simulation's intended purpose.[15] Different types of sensitivity analysis are applied to process simulation, including one-at-a-time and multiple parameter analysis. The first implies changing one parameter at a time and seeing how much the output changes, thus evaluating the parameter's effects on the process outputs. The second approach evaluates interactions between multiple parameters, but requires planning to avoid being overloaded by the number of combinations to test.[43]

The choice of validation techniques depends on the type of simulation and the information or budget available. A simulation of a process with few measurements available may rely mostly on qualitative validation techniques, using process knowledge, whereas a simulation of a well-monitored process is validated with either qualitative or quantitative techniques that take process data as a credible detailed description of the real process. These techniques have been classified in different ways throughout literature and represent different points of view regarding validation.

When simulating existing processes, we classify validation techniques according to type of process information:

- Informal validation is qualitative: the simulation results are assessed with either direct numerical comparison (graphical techniques) or testing against general process knowledge. This approach relies on experts to agree or disagree with the model. Graphical comparison is a preferred technique in many publications as a means of communicating results where an order-of-magnitude comparison is simply presented.
- Formal validation is quantitative and implies a mathematical approach, using multiple simulated values, statistics, and experiments to evaluate the simulation. The disadvantage of these techniques is their need for additional data and time, but they are more rigorous than informal validation because they rely on specific values to decide whether a simulation is valid. Most formal techniques fall under operational validation, which validates the behaviour, or results, of the simulation.

Simulation software advances have greatly accelerated the process of building a simulation by providing already-programmed mass and energy balances, calculations, and component properties and facilitating flowsheeting. However, the choices of unit models, streams, thermodynamic packages, and parameters still remain the engineer's responsibility. Each step of a methodology is designed to narrow down and justify these decisions. Simulation methodologies are usually presented as a sequence of steps that result in a finished product, but in practice, the process is iterative.

The validation step in particular requires modifying the simulation or its inputs chosen in previous steps when the model does not represent the process accurately enough. The various techniques proposed all have advantages and disadvantages, and it is preferred to employ several techniques with different criteria to validate a simulation to cover its various aspects.

Literature about simulation validation techniques remain broad enough to be applicable to different fields, but some techniques are more applicable than others to steady-state simulation. Few publications are specifically dedicated to simulation validation in chemical process engineering; most publications about simulation only mention it briefly, without focusing on it.

2.3 | Simulation of industrial processes: Current perspective

In recent decades, major advances in simulation have been related to software and data management technologies. Computational and storage capacities have increased exponentially, while their costs have decreased. Large amounts of data are collected every day throughout industrial processes. Most of it is for process control, with some being accessed to monitor quality and carry out process troubleshooting. Tools have had to be developed to fully realize their potential.[45]

Steady-state simulation building has improved progressively due to automating the process design steps.[46] On-line applications like digital twins have gained in popularity,[47] and simulation software meets higher performance standards: simulation validation principles have been applied to new and more complex simulations.

2.3.1 | Real-time data management systems (DMS) and industrial data use

The exponential popularity of information systems since the 90s has permitted more automation, which provides many process data collection opportunities. With enlarged storage capacities, the sampling frequency of data collection has increased, and data management has had to follow.[19]

Data historians have been supplanted by data management systems (DMS), which have a wider scope. This type of software collects and provides access to data from more data sources and storage locations, but its main function is to facilitate data-driven operational decisions. This means that a DMS enhances the classical functions of the data historian by providing integration of more data source types company-wide, making it easier to access data from remote locations and incorporating data contextualization, processing, and visualization capabilities. Process data are therefore less constrained to mills, enabling other stakeholders to access them more easily. Business analysts and data scientists create predictive models to draw insight from operational data; suppliers remotely monitor their own equipment and provide diagnostics; and remote engineers manage asset models and analytics to prevent downtime. This increase in collaboration opportunities has emerged thanks to Industry 4.0's focus on connectivity, automation, and data usage, moving from local on-premise data management to supporting information sharing to taking full advantage of it.[48] An example of a process data management system is Aveva's PI System, which has been implemented in many chemical, mining, and forest products industries.[49]

Although the amount of data has become larger, gross and random errors are still present, which means that data require treatment. A DMS provides access to data processing as it is central to the usage of process data today and necessary to manipulate the huge amounts of data to produce meaningful information.[50] Methods like multivariate statistical projection compress massive amounts of data into smaller numbers of variables to facilitate process monitoring and interpretation.[51] Other data processing techniques are implemented, including online filters and treatments that improve real-time process control and monitoring like machine learning-based ones.[52]

More punctual data processing may be needed to meet the requirements of projects like process simulation. This kind of signal processing approach is expanded into its integral steps.[53] Data cleaning remove erroneous signals; moving averages and polynomial filters employ a weighted sum of previous measurements, whereas Fourier and wavelet transforms filtering, target random error frequencies in the signal.[54] These techniques produce a smoother signal that better represents the process to differentiate steady state from unsteady or transition periods. This detection step identifies periods contained in datasets for decision-making activities like process simulation. Statistical techniques, like a mathematical evidence test, or direct approaches like linear regression and Student's *t*-test determine whether a variable is stable during the time period tested.[55] Engineers may apply an operating regime detection step to subdivide this steady-state data into different operating conditions corresponding to different control setpoints, for example, to produce different product grades or to operate through different seasons. This is done by clustering techniques in conjunction with process knowledge to distinguish groups of plant data and identify the factors that differentiate them into operating regimes.

Off-line manual approaches to data cleaning present different challenges from more automated approaches like implementing both quantitative and qualitative techniques. Whereas the first techniques rely on statistical methods to identify abnormal data, the second applies constraints or rules to detect errors but requires process knowledge to implement.[56]

Industrial data analysis tools, like the EXPLORE software by Canmet ENERGY,[53,57] apply advanced statistical analysis of process data, but also practical pretreatment of real-time data extracted directly from data management systems, or like the JMP software,[58] which includes data visualization and exploratory data analysis functions, statistical predictive modelling, and data mining. This software possesses capabilities for practical data cleaning in addition to more control- and operation monitoring-oriented applications.

DMS are built for specific plant needs, which rarely include gathering data for simulations, but more advanced systems facilitate data access and sharing outside the plant environment. This coincides with the availability of more advanced on-line and off-line data processing software that have the ability to remove the errors that are inevitably present.

2.3.2 | Validation techniques using real-time process data

Although validation principles have not changed since the introduction of more advanced DMS, process simulation has made numerous advances, such as more complex process representations,[59] hybrid models combining first-principles and data-driven modelling,[60] and automation of some simulation building steps.[47] Validation still requires human intervention; success of a simulation remain linked to its validation and acceptability by its users. The technological means, like quantity of information and computing power, have increased exponentially, expanding the 'reach' of existing validation techniques; their correct application still depends on the modeller.

Simulation verification and validation (V&V) research experts identified and reviewed multiple issues at the start of the century that are still present today. Among these, terminology issues and differences in definitions of common concepts and acronyms hinder their communication and application. This problem needs to be addressed explicitly during each simulation study. Advances in simulation frameworks enhance V&V capabilities, and the limitations on access to information necessary for effective validation are being corrected by the development of more advanced data processing and process simulation software, which facilitates exchange of data from different sources.[61] Technological advances have increased processing capabilities for both data processing and simulation of complex processes, making both faster: executing runs of a simulation in a shorter period makes more extensive validation processes more practically feasible.

The concept of data validity, 'ensuring that the data necessary for model building, model evaluation and testing, and conducting the model experiments to solve the problem are adequate and correct',[44] has recently been given more attention. Using data representative of normal operating conditions provides a better basis for all formal validation techniques. For example, access to more real-time data provides opportunities to discern different pseudo steady-state operating conditions that are present in the process and opportunities to more effectively apply formal validation techniques like sensitivity analysis.

Design of experiments (DOE) is a possible approach that is employed as part of formal validation. DOE itself consists in the planning of testing that minimize the number of experiments needed when considering many variables, their ranges of values, and analysis of the those experimental results[62]; it has been successfully implemented in industrial settings to improve design or optimize processes.[63] DOE was first developed to facilitate physical process experimentation, but computer simulation has enabled cheaper reproductible testing without the costs and limitations of physical processes, like uncontrolled factors that cause random errors.[64] DOE relies on statistical methods: factorial DOE may be employed to create an experimental plan that incorporates the additive effect of modifying process inputs and the effect of interactions among those inputs on the simulation output. Fractional factorial designs further reduce the number of experiments needed, especially in cases where the process has many variables. These experiments make it possible to validate a more complex simulation because they provide a wider and more relevant range of simulation results to compare with real process data.[65]

DMS provides data for more immediate applications (process control and quality assessment), but also access historical real-time process data. That has improved the simulation and validation of existing processes. Simulations have become more complex, yet validation techniques have remained the same because they involve verifying fundamental aspects of simulations like their underlying logic or their outputs. However, using more advanced simulation and data processing tools has made the application of formal techniques easier.

Data validity means that the data are adequate and correct for the simulation project, and the validity of a simulation for its intended use is dependent on the degree to which it represents the process. Even the current techniques and technologies of data processing and process simulation cannot replace process knowledge and the role of expert judgement in those activities.

There are few publications on simulation that take advantage of the current amounts of data available from industrial processes. This is possibly due to confidentiality reasons. DMSs are operated in industrial settings where researchers are excluded from the main target audience for information sharing. Generally, data and any models derived from it remain for internal applications.

2.4 | Critical analysis

There are many validation techniques, and the simulation modeller must have the appropriate knowledge to choose the right one(s) for the right objective. From the three aspects of validity, conceptual model validity and operational validity are the most relevant because already-programmed simulation software is almost exclusively employed in process engineering, lessening the need for verification of computer-based models. Conceptual model validity requires both process information and a clear statement of the simulation's purposes to verify the process representation.

Categorization of validation techniques as either informal or formal identifies their validation criteria: whether the evaluation of the representativeness of the simulation is qualitative or quantitative. This usually means that the former is easier to implement, whereas formal validation techniques require more planning, data gathering, and treatment. Simplification of the validation step in a simulation-building methodology usually means that formal validation is overlooked in favour of using verification and informal techniques only. This approach is ineffective due to the following reasons:

1. The combination of process automation and the absence of longer-term key process-knowledge roles in older industries contribute to a knowledge gap in the process industries.
2. Subjectivity and human error among experts influence the quality of qualitative validation results.

A combination of both categories of techniques is more effective and part of a practical validation strategy because they verify different aspects of the simulation and complement each other. Repeating face validation with different experts strengthens informal validation and complement formal validation techniques when process data are lacking. Informal validation techniques are effective in the earlier stages of simulation building, ensuring conceptual model validation and verification of flowsheets, whereas formal validation techniques are more effective at the operation validation stage.

In the process engineering research domain, little fundamental research has been done on the practical validation of simulations of industrial processes. The rest of the simulation-building process is well understood and documented. Methodologies are usually presented as a chronological sequence of steps, but they are iterative in practice, and some activities are executed concurrently. In the context of design and optimization, more experimental activities can be added to provide or correct process information like model reparameterization, which plays a role in improving models based on experimental data like statistical models and kinetic reaction models. For example, modelling of ethylene hydrogenation reaction's rate was improved by implementing a sequential experimental design coupled with reparameterization. This reduced the correlation between two of its parameters and the volume of the confidence region of the estimated parameters.[66]

Computer simulation programs developed since the 1950s have simplified the process of building simulations. However, the choices of process unit models, streams, parameters, and inputs still remain the engineer's responsibility: the success of the simulations produced relies on the modeller's process knowledge.

Steady-state simulations are most commonly employed either to evaluate multi-unit processes or to estimate parameters for specific unit operations, as part of equipment sizing and design projects for example. Simulation projects are budgeted according to their scope, with the working simulation as the final product. Without an understanding of the importance of validation and a clear and planned strategy, this step is often overlooked or oversimplified during simulation development due to:

1. Additional costs needed to implement formal validation;
2. Historically, a lack of real process data for additional comparison, or (currently) the opposite—too much raw data—that need additional treatment to be suitable for comparison;
3. Modellers' lack of knowledge or familiarity with validation techniques.

Formal validation techniques rely on data and statistical methods to evaluate the representativeness of simulations, which make them more difficult to implement than informal techniques. Process design or optimization projects need their own data collection campaigns to obtain specific operating data.

Currently, the state of process knowledge, employed for informal validation, is industry-dependent but the technological advances and the falling costs of computing power and storage facilitates the application of formal validation techniques The availability of so much process data poses a new problem: choosing the correct data, and the right amount of it to employ in simulation building and validation. Access to real-time data has made formal validation easier (and cheaper), but it is only effective if the data exploited for comparison is a good representation of steady-state process operation and if the validation techniques chosen are appropriate. Given the large amounts of data available and the uncertainties introduced by raw data, data validation becomes integral to the overall simulation validation.

To extract correct information from raw data, it must be processed: data treatment techniques predominantly for punctual projects in industrial settings are still mostly manual. Even with extensive research on automated data treatment, techniques based on filters and even machine learning are not widely implemented at the mill level. This has started to change thanks to the implementation of more extensive DMS, which include some data processing capabilities.

3 | APPLICATIONS

3.1 | Bibliometric review

From 1992 to 2023, Web of Science indexed 505 articles related to process simulation validation.[67] *Industrial & Engineering Chemistry Research* published almost 10% of the articles (48), and other journals that have published many include *Computers & Chemical Engineering* (31), *Chemical Engineering Science* (29), *Chemical Engineering Journal* (22), and *Applied Energy* (21). Since 2020, the most cited articles discussed optimization of heat exchange networks,[68] spray-evaporation of desalination systems,[69] and fault analysis of reformer units in hydrogen plants.[70] Since 1992, *The Canadian Journal of Chemical Engineering* published seven articles and the latest two deal with optimizing a natural gas dehydration facility[71] and real time monitoring with mid-infrared spectroscopy.[72]

A bibliometric analysis identified five clusters of research: kinetics (with the most keywords), optimization, and mass transfer/CO_2 capture (Figure 4). The absence of keywords associated with specific fields confirms that simulation, and by extension its validation, is a tool that is generic enough to chemical engineering that it is not disproportionally associated with any one sector.

Typical themes include petrochemical process modelling-production of ethylene from ethanol by dehydration in an industrial-scale plant, for example,[73] where the validation section includes a representation of the 10 industrial plant operating conditions at steady state that were considered for the simulation. Three were employed to build the simulation, and the other seven to validate it. Flowrate and chemical composition were compared with industrial data and temperature profile, ethanol conversion rate, and the ethylene molar fraction inside the reactor were compared to pilot plant measurements and data from another model.

Some articles[73–78] state that they validated the model with plant or experimental data without any elaboration. In other cases, the validation data is only described by its origin (e.g., theoretical, literature, other industrial process data), and the validation result presented is a graph comparing the simulation results to the industrial process data: plant data from the industrial fluidized catalytic cracking unit[79] and simulation of gasification and gas purification units are examples.[80] Schmidt and Strube[81] presented a detailed model validation methodology to support the implementation of liquid–liquid extraction in the manufacturing of new high-value biological products.

Articles vary in the portion of their text reserved for simulation validation; simple mention of validation in the body of the article and graphical comparison of results and literature values are common, while more details about the validation process or result data characterization is rarer. Publications from this bibliometric review could be separated into three categories:

- Minimal validation description: for example, 'the results were validated using data from this source', or a statement that simulation outputs were compared to results from another study or model, without further quantitative description of either.
- Graphical comparison: comparison of simulation results with sparse process data.
- Formal validation, including process data description, explicit validation procedure, and results.

A fourth category, which includes only articles specifically on the topic of validation, remains a very small minority, especially in the field of chemical engineering. Papers exclusively on the subject of simulation validation are more often published in a dedicated research area.[10,82,83]

Regardless of the objectives or intended application of the simulation, this survey aligns well with the preliminary bibliometric review analysis results and the background critical analysis to conclude that: (a) most articles do not consider validation their principal topic, and (b) advances in data management are reflected in the 'depth' of the validation approach applied. Explicit validations are more prevalent in recent articles, coinciding with increased accessibility to industrial data.

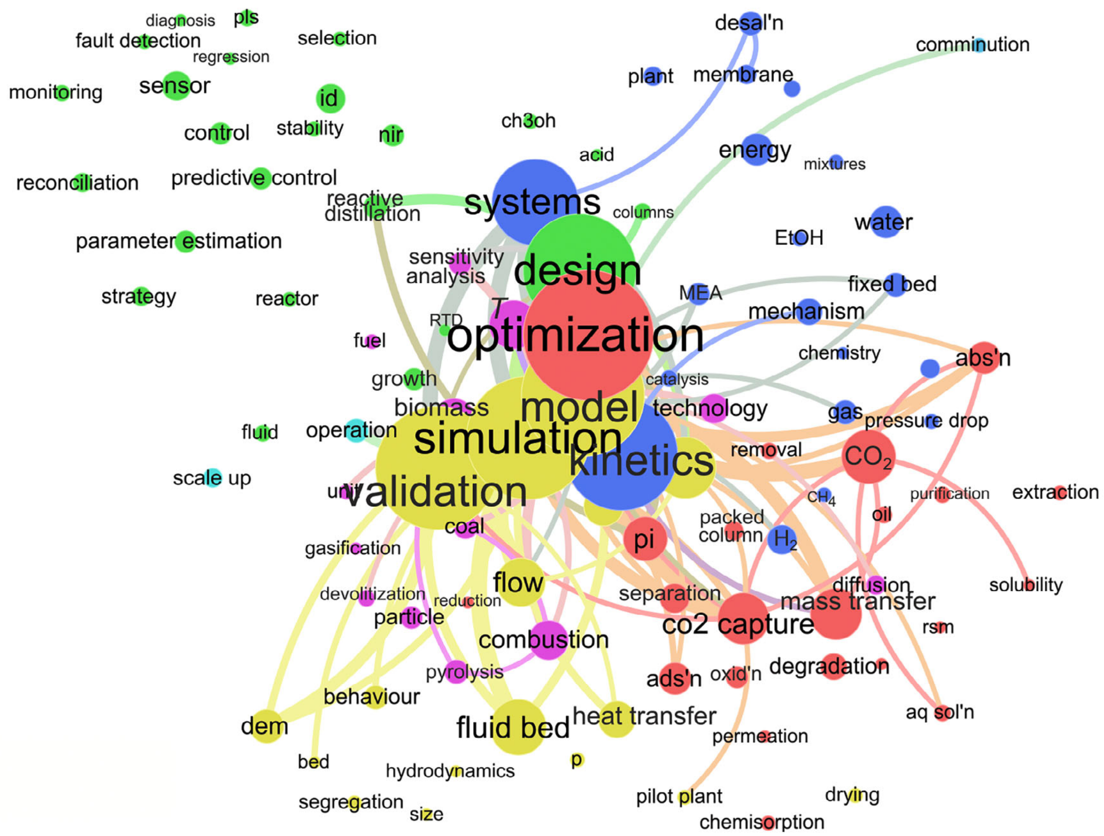

FIGURE 4 Web of Science bibliometric map with 'process simulation' and 'validation' in the Topic search criteria (1992 and 2023).[52,53] VOSViewer assigned the 100 top cited keywords into 5 clusters. The size of the nodes correlates with the number of articles that mention the keyword: The largest node represents 53 articles (optimization), while the smallest nodes represent 5 articles. We resized the model, simulation, and validation nodes to 50 from 127, 126, and 91 to minimize overlap with nodes in the vicinity. The red cluster, with the most keywords, is centred around optimization with CO_2/CO_2 capture as minor topics. The green cluster has the second most keywords and deals with design, control, parameter estimation, and reconciliation. The blue cluster encompasses systems and applications like kinetics, catalysis, and membranes. The yellow cluster comprises the search criteria: Simulation and validation. The magenta cluster focuses on combustion, pyrolysis, and biomass. abs'n, absorption; ads'n, adsorption; aq sol'n, aqueous solution; dem, discrete element method; desal'n, desalination; devol'n, devolitization; MEA, monoethanolamine; NIR, near infrared; P, pressure; PI, process intensification; PLS, partial least squares; RSM, respose surface methodology; RTD, residence time distribution; T, temperature.

3.2 | Industrial process simulation applications

Interviews with industry experts in the pulp and paper industry have resulted in descriptions of the practical aspects of simulation and its usage in industrial settings.[53] Perspectives on simulations and their validation include their application for process monitoring, optimization projects, and training.

Process simulation covers a large spectrum of applications and serves as a basis for many decisions. For example, simulations provide soft sensors for process control when instrumentation is not available. By analyzing the components and temperature data provided by the simulation, it is possible to employ the predicted values as additional control points in the absence of real instruments. If a measuring instrument breaks, the simulated value can be substituted temporarily. Simulation also guides equipment performance analysis and assessment of process yield and is the basis for a wide range of applications.

Both steady-state and dynamic simulation are employed to understand processes. Dynamic simulations replicate the live process. The current process must be correctly understood, including how it reacts to changes, before conceptualizing major operating modifications. Developing a dynamic simulation is considered much harder than developing a steady-state one; therefore, mill personnel and consultants tend to favour steady-state simulation. Mill staff employ it for energy and water reduction, for specific projects (to maximize production at minimum cost, or to increase production), to assess the impact of major capital expenditure projects, and evaluate the success of a or the impact of breaks. Consultants develop simulations for process design, for example, to justify a project, and to balance flow sheets for new designs or for retrofits and upgrades. Mill staff also employ process simulation for operational training, commissioning the plant, testing new configurations, analyzing process yield, monitoring equipment performance, debottlenecking, troubleshooting, process optimization and improvements, identifying what would happen if production increased, and auditing mills. Steady-state mill simulations are very much oriented towards normal operations.

For companies, simulations are produced by the corporate technology group, and subject matter experts employ it in their mill audits. Each mill has its own simulation file that represents the whole mill, and the corporate side oversees updating the simulation. These simulations require 15%–20% of the PI tags and result in significant costs, as well as representing important opportunities, and hence they are maintained by the corporate power group. They are mostly tools for yearly planning, for improvement, and for justification for capital projects, upgrades, new equipment, changes, or benchmarking. Mill personnel have access to the results, but they do not run the simulations.

Unfortunately, because most mills have divested themselves of their research departments, whole-mill simulations are not commonly maintained. Most mill personnel are indeed busy putting out fires, so they will build a dynamic simulation only around a unit operation like a turbine or a boiler, or simulate specific areas where the impacts of process dynamics are larger. These units are usually product grade-independent, and hence there is no need to distinguish and simulate the different grades. Indeed, on the power side, grades make a marginal contribution. However, one mill group did simulate the paper machine and considered their different product grades.

Mill staff assume that their processes run in steady-state most of the time, and that therefore a steady-state simulation is sufficient for their applications. Moreover, data quality is also an issue here; picking a period in historical data to find some fairly steady-state data and averaging them to create a steady-state simulation is much safer than attempting to replicate process dynamics if the measurements are not trusted. For all the mills, seasonality was almost never considered in the simulation. Because most mills surveyed were in the south, external temperature had less impact on their process than it would have in a harsher climate. Only one of the mills simulated winter conditions as a worst-case scenario.

In general, process simulations are done using Excel, WinGEMS (Valmet), IDEAS (Andritz), and CADSIM Plus (Aurel Systems). Software packages, like CADSIM Plus, have additional functions like least-squares fit of the data for their dynamic simulation. They run thousands of trials every minute, changing the measurements until they get the best match/fit to the measured variable. They put weights on the measurements coming out of sensors according to the level of trust they have. In the end, they write the corrected information back into the DCS and the historian. They are currently working on adding a real-time optimization layer to their software to see, for instance, what could be done with the current production to save money. To efficiently optimize process operations, a solver is not enough; the process model and the measurements must be accurate.

When a mill did have a process simulation, it was always validated by comparing the simulated values to either the process measurements or the quantity (weight/tonnages) of the final product. An alternative and widely employed method of validation was to seek input from process experts and specialists in the mill to ascertain if the simulated values were coherent. The absence of feedback and expertise from mill personnel renders the creation of an accurate simulation impossible.

Digital twins are tools that have been gaining popularity. They are considered bona fide simulations running live alongside the process; real-time process simulators, cloning the process, support real-time applications.

Process simulations benefit from measurement redundancy, but in sectors like the pulp and paper industry, process experts have made assumptions for places where there were no sensors. Therefore, building a simulation highlights the specific area where it would be critical to add sensors for better decision-making.

Steady-state simulations are the preferred type of simulation due to their practical application in mills. Industrial process simulation has the potential to greatly impact decision-making by enhancing the value of available data. However, there are significant challenges associated with its implementation. Mill-wide process simulations are uncommon and are typically developed externally by consultants or dedicated corporate modellers. This is because such projects are time-consuming and would disrupt the daily responsibilities of mill personnel. As any modifications made to the real process must also be implemented in the model, ongoing investments are necessary to keep the simulation accurate.

4 | CASE STUDY: THE PRACTICAL VALIDATION OF A WASHING DEPARTMENT IN A KRAFT PULP MILL

This case study's simulation was created as a decision tool to evaluate changes in steady-state process operation for an existing. The simulation must represent the mass and energy balances around the process to a sufficient degree for this purpose. A short description of the department in this case study is followed by the simulation, validation methodology, and validation results. Verification of the simulation flowsheet by mill experts, informal validation and a formal validation were performed.

4.1 | Process description

The system modelled is the brownstock washing (BSW) department of a Canadian dissolving pulp Kraft mill (Figure 5). Its function is to remove black liquor (containing spent cooking chemicals and solubilized wood components like lignin) from the wood pulp. This system is composed of five main sections. The pulp is first fed into the knotter section, where uncooked chips and knots are removed. The washer section, two parallel lines of drum washers operating in countercurrent, separates the pulp from the black liquor by diluting it with wash liquor, extracting and injecting it in the previous washer in the line. Pulp is then mixed before entering a screening section made up of a three screens cascade system, in which the last one's rejects are sent to the cleaner section where small particles, like sand, are removed. The pulp accepted by the first screen is sent to one last drum washer, whose role is to thicken the washed pulp and provide wash liquor to the rest of the department. The pulp is then sent to a buffer tank that feeds the bleaching department.

The process data from the BSW department includes 24 flow and consistency variables, whose locations are identified on its flowsheet (Figure 5). Values were extracted from the mill data management system over 1 year of operation as 10-min averages to reduce their volume and facilitate processing.

4.2 | Department simulation

The process simulation was built using the methodology outlined in the theory (Figure 3). During the problem definition step, a steady-state simulation was chosen because it offered a representation with an appropriate level of detail (process units) and was deemed sufficiently flexible for daily operation decision-making. The simulation as built considered flow and pulp consistencies variables as they are the most abundant throughout the mill.

Mode (the most common operation of the process) and best performance (a stable period of a few days where the pulp production average was over its 90th percentile) operating regimes were selected to be modelled during the design of the simulation study step. CADSIM Plus, a process simulation software from Aurel Systems[84] that includes unit modules and components in the pulp and paper sector, was chosen as the modelling tool.

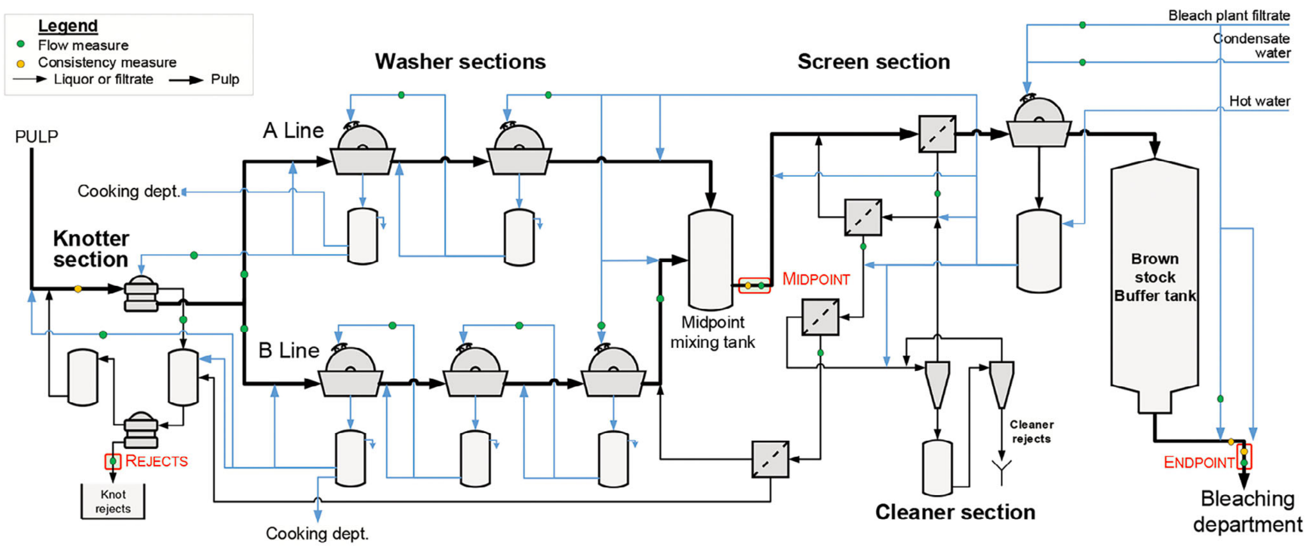

FIGURE 5 Flowsheet of the brownstock washing (BSW) department simulation.

Tanks were represented using generic modules while knotters, washers, screens, and cleaners were all available as separating unit modules in the software. As an example, the screens model selected requires a passage ratio parameter P_r, the ratio of consistencies of the pulp through a screen aperture and pulp upstream of that aperture. Reject rates describe the partitioning of inlet pulp fibres, on a volumetric (R_v) or mass (R_w) basis, which are the ratios of reject and feed flow rates. The passage ratio relates those reject rates and complete the mass balances around screen process units, considering one feed and two outlets:

$$R_w = R_v^{P_r}$$

Of the 24 measurements identified in the department, 20 were flow and 4 were pulp consistencies. A magnetic flow meter is one of the most commonly used meters in the pulp and paper industry because of its relatively low cost and no significant pressure drop. It can be used on liquids and slurries and its accuracy is 1±0.5%.[85]

Data treatment of those measurements was conducted during the input and process definition step. The process data was cleaned using the EXPLORE software from Canmet Energy.[86] Most errors, shutdowns, and unsteady time periods were removed using a pre-established manual data treatment process (Figure 6). The values for the two operating regimes to simulate were identified from the treated data, and the software produced a statistical summary (mean, standard deviation, quartiles, etc.) of the treated variables.

4.3 | Validation activities

Verification that the simulation flowsheet is a correct representation of the real process was conducted during the conceptual model design step. In the case of the BSW department simulation, it consisted in the verification that the operating components (major equipment) were all present, located in their right place, and correctly connected. This activity rectified process stream connections, like the ones in the knotter system, whose documentation was not up to date. Operators were able to correctly define that part of the flowsheet.

An informal validation, employing the face validation technique, was conducted once the initial simulation was operational, using the treated values of the mode and the best performance operating regimes as input. It was conducted on a one-on-one basis, asking several operators and the department supervisor to validate that the simulation's principal outputs of pulp were realistic. In addition to their validation of the outputs, experts specified the usual range of values that they observed or the target values of the process's most monitored variables. This insight delimited real process operating ranges to compare with subsequent simulation outputs.

Sensitivity analysis, as a formal validation technique, was finally employed to determine the impact of changes in the inputs and parameters on the outputs of the simulation. First, the impact of the individual variables is evaluated over their normal operating range (based on their treated data) by one-at-a-time sensitivity analysis. Second,

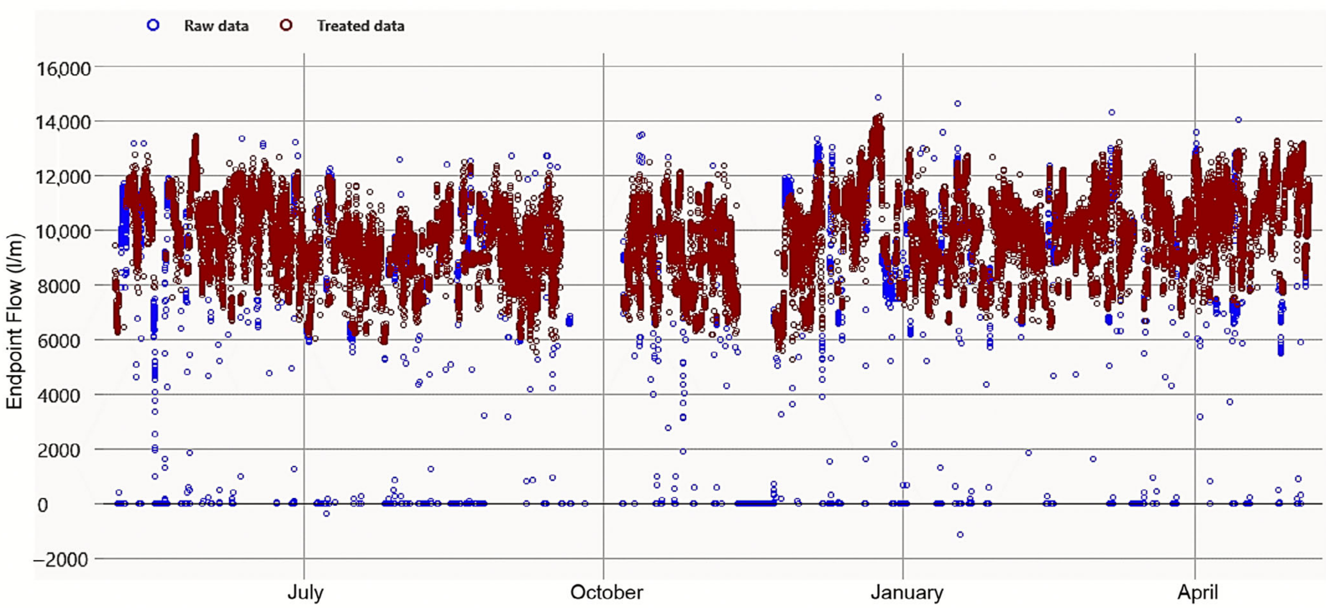

FIGURE 6 Data treatment results of the endpoint flow variable (mode operating regime).

a multi-variable sensitivity analysis was conducted to evaluate the impact of the full range of pulp conditions entering the BSW to imitate real operation where both the quantity and consistency of pulp vary. The effects of parameter variations were considered by comparing the simulation results to the treated data and ranges of values from mill experts for unmeasured variables (like washer pulp consistency outputs) to verify whether the variation was like the real process's. Any simulation results outside of expected ranges led to appropriate modifications.

4.4 | Validation results

Five variables indicated in red in BSW flowsheet (Figure 6) are presented here to follow the effects of the validation activities. The midpoint flow and consistency were the values measured after the midpoint mix tank, whereas the endpoint flow and consistency were measured after the brown stock buffer tank; the knotter rejects flow was also included because it represents the biggest loss of pulp in this part of the mill.

The distribution of the treated data, the initial simulation results, and the informally and formally validated results of the mode operating regime are compared. All values were normalized using their respective mode, the value that the simulation targets, for a better visualization (Figure 7), The evolution of validation's impact on the simulation results is represented in comparison to the treated data's minimum, maximum, and average $\pm$ standard deviation for each variable.

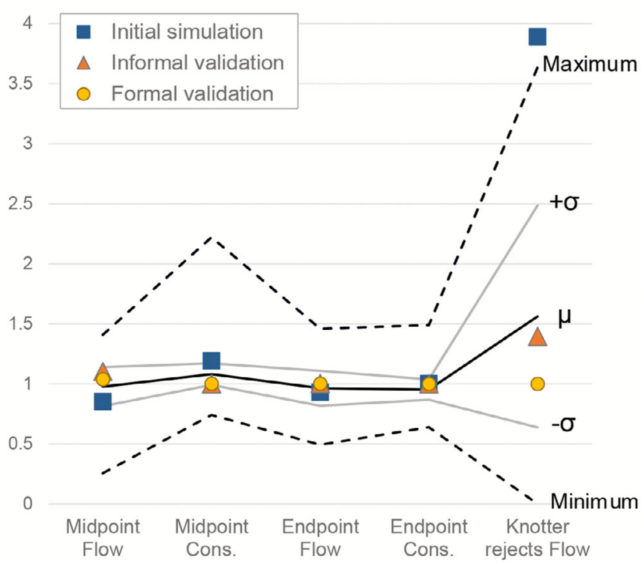

FIGURE 7 Comparison of the (treated) measured variable, their quartiles, and their simulated result counterparts for the mode regime, values normalized.

The endpoint consistency variable has the narrowest treated data distribution, and its simulated values are the most accurate of the five variables. For all variables, the formal validated simulation results are closer to the treated data values, and in most cases, the informal validation also provided reasonably close results. The knotter rejects flows are the values that are the furthest from their mode before formal validation.

The knotter rejects process data has a wider range than the other variables and it was also sampled differently: the rejects are weighed in a container once a week before being sent to waste treatment, making their values less reliable than 10-min measurement averages. Investigating the initial simulated value of that variable revealed an error in the knotter parameters that was corrected during informal validation. Even with the reject flow measurement uncertainty, the simulation was validated with all the variables' simulated values within the confidence interval of one standard deviation of their treated values.

Informal validation is particularly serviceable when process data is lacking, as was the case in this department: considering two variables (total flow and pulp), there were only 24 measurements from the 95 streams in the BSW department. The subjectivity of face validation is mitigated by consulting different experts separately. Although four out of the five variables were deemed to be already sufficiently good with just informal validation, formal validation is both a tool to improve the simulation's weak points and a way to confirm the corrected values of the informally validated simulation.

5 | CONCLUSIONS

Simulations have served in the design of new projects as well as the retrofitting and troubleshooting of existing ones and as tools evaluate process emissions, costs, and productivity. Validation ensures the accuracy of a simulation by comparing it to reality by the intermediary of data and expert knowledge.

Advances in data access, Quantity, and the means to interpret thanks to Big Data, the Internet of Things, advanced filters, and artificial intelligence have improved the validation of the simulation's results (operational validation). However, not everything to be simulated is measured, and engineers must verify conceptual models and operating hypotheses before the simulation is built; expert knowledge is necessary to address those challenges. The loss of experienced veterans in certain well-established industries, pressures for better economic and environmental performances, and the rise of new equipment and bio-based processes emphasizes the need for validation that employ multiple techniques and utilize all process information.

Validation is often not a primary concern in publications that centre around simulation. This lack of emphasis casts doubts on the credibility of their findings, despite overall robust research methodology. The absence of knowledge regarding validation principles and techniques contributes to this oversight.

By systematically incorporating validation procedures and demonstrating their results, both industry professionals and researchers have the opportunity to enhance the accuracy and acceptability of their simulations.

AUTHOR CONTRIBUTIONS

Caroline Brucel: Writing – review and editing; writing – original draft; visualization; conceptualization; data curation; formal analysis; methodology; investigation. **Émilie Thibault:** Writing – review and editing; resources; writing – original draft; formal analysis; investigation. **Gregory S. Patience:** Resources; writing – review and editing; formal analysis; supervision; validation. **Paul Stuart:** Resources; writing – review and editing; project administration; methodology; funding acquisition.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/cjce.25601.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Gregory S. Patience https://orcid.org/0000-0001-6593-7986

REFERENCES

[1] F. D. Bähner, O. A. Prado-Rubio, J. K. Huusom, *Chem. Eng. Res. Des.* **2020**, *159*, 66.
[2] D. E. Seborg, T. F. Edgar, D. A. Mellichamp, F. J. Doyle III., *Process Dynamics and Control*, John Wiley & Sons, Hoboken, NJ **2016**.
[3] P. Aryan, G. L. Raja, R. Vilanova, *International Journal of Systems Science* **2024**, *1*, 2874.
[4] D. Mukherjee, G. L. Raja, P. Kundu, A. Ghosh, *Asian Journal of Control* **2023**, *25*, 2165.
[5] A. S. Rao, M. Chidambaram, *PID Control in the Third Millennium: Lessons Learned and New Approaches*, Springer, London **2012**, p. 75.
[6] W. Luyben, *Plantwide Dynamic Simulators in Chemical Processing and Control*, CRC Press, Boca Raton, FL **2002**.
[7] G. S. Patience, in *Experimental Methods and Instrumentation for Chemical Engineers* (Ed: G. S. Patience), Elsevier, Amsterdam, The Netherlands **2017**.
[8] G. S. Patience, *Can. J. Chem. Eng.* **2018**, *96*, 2312.
[9] J. De Tommaso, F. Rossi, N. Moradi, C. Pirola, G. Patience, F. Galli, *Can. J. Chem. Eng.* **2020**, *98*, 2301.
[10] O. Balci, in *Winter Simul. Conf. Proc* (Eds: D. J. Medeiros, E. F. Watson, J. S. Carson, M. S. Manivannan), IEEE, Washington DC **1998**, p. 41.
[11] R. G. Sargent, in *Winter Simul. Conf. Proc* (Eds: B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, E. Yücesan), IEEE, Baltimore, MD **2010**, p. 166.
[12] P. K. Davis, Generalizing Concepts and Methods of Verification, Validation, and Accreditation (VV&A) for Military Simulations. **1992** https://www.rand.org/pubs/reports/R4249.html (accessed: December 2023).
[13] B. H. Thacker, S. W. Doebling, F. M. Hemez, M. C. Anderson, J. E. Pepin, E. A. Rodriguez, Concepts of Model Verification and Validation. **2004** https://www.osti.gov/servlets/purl/835920 (accessed: November 2023).
[14] A. M. Law, in *Winter Simul. Conf. Proc* (Eds: M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, R. G. Ingalls), IEEE, Austin, TX **2009**, p. 1402.
[15] R. G. Sargent, in *Winter Simul. Conf. Proc* (Eds: A. Thesen, H. Grant, W. D. Kelton), IEEE, Atlanta, GA **1987**, p. 33.
[16] M. S. Martis, *Electron. J. Bus. Res. Methods* **2006**, *4*, 39.
[17] R. Motard, M. Shacham, E. Rosen, *AIChE J.* **1975**, *21*, 417.
[18] T. J. Williams, *Proceedings-IEEE* **1966**, *54*, 1751.
[19] B. Balasko, J. Abonyi, *Hung. J. Ind. Chem.* **2007**, *35*, 75.
[20] D. Barr, *Proc. Int. Conf. Life Management of Power Plants*, IET: The Institution of Engineering and Technology, Edinburgh, UK **1994**.
[21] A. Galan, C. De Prada, G. Gutierrez, D. Sarabia, R. Gonzalez, *J. Process Control* **2021**, *100*, 41.
[22] Y. Miao, H. Su, G. Rong, J. Chu, *Meas. Control* **2009**, *42*, 209.
[23] S. Narasimhan, C. Jordache, *Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data*, Gulf Publishing Company, Houston, TX **1999**.
[24] M. J. Bagajewicz, *Process Plant Instrumentation: Design and Upgrade*, CRC press, Boca Raton, FL **2000**.
[25] C. M. Crowe, *J. Process Control* **1996**, *6*, 89.
[26] A. Behroozsarand, S. Afshari, *Comput. Chem. Eng.* **2020**, *137*, 106803.
[27] G. Weiss, J. Romagnoli, K. Islam, *Comput. Chem. Eng.* **1996**, *20*, 1441.
[28] D. C. Y. Foo, R. Elyas, in *Chemical Engineering Process Simulation, 2* (Ed: D. C. Y. Foo), Elsevier, Amsterdam **2022**.
[29] J. Haydary, *Chemical Process Design and Simulation: Aspen Plus and Aspen Hysys Applications*, John Wiley & Sons, Hoboken, NJ **2019**.
[30] J. Banks, *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, John Wiley & Sons, Hoboken, NJ **1998**.
[31] A. M. Law, M. G. McComas, in *Winter Simul. Conf. Proc* (Eds: B. L. Nelson, W. D. Kelton, G. M. Clark), IEEE Computer Society, Los Alamitos, CA **1991**, p. 21.
[32] K. J. Musselman, in *Winter Simul. Conf. Proc* (Eds: J. J. Swain, D. Goldsman, R. C. Crain, J. R. Wilson), IEEE, Arlington, VA **1992**, p. 115.
[33] O. Ülgen, J. J. Black, B. Johnsonbaugh, R. Klunge, in *MSUG* (Ed: E. Williams), University of Michigan, Dearborn, MI **2006**.
[34] J. Pomroy, *Chem. Eng.* **1996**, *103*, 5.
[35] J. Haydary, in *Chemical Process Design and Simulation: Aspen plus and Aspen Hysys Applications* (Ed: J. Haydary), John Wiley & Sons, Hoboken, NJ **2018**, p. 15.
[36] R. Elyas, *Chemical Engineering Process Simulation*, Elsevier, Amsterdam, The Netherlands **2023**, p. 57.

[37] O. Balci, in *Winter Simul. Conf. Proc* (Eds: E. A. MacNair, K. J. Musselman, P. Heidelberger), IEEE, Washington, D.C **1989**, p. 62.
[38] R. B. Whitner, O. Balci, in *Winter Simul. Conf. Proc* (Eds: E. A. MacNair, K. J. Musselman, P. Heidelberger), IEEE, Washington, DC **1989**, p. 559.
[39] M. Olsen, M. Raunak, in *Winter Simul. Conf. Proc* (Eds: R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, M. E. Kuhl), IEEE Press, Washington, DC **2013**, p. 1569.
[40] J. P. C. Kleijnen, in *CentER Discussion*, Operations Research, Tilburg, The Netherlands, **1998**.
[41] J. Banks, in *Discrete Event System Simulation* (Eds: W. J. Fabrycky, J. H. Mize), Pearson Education, Upper Saddle River, NJ **2005**.
[42] D. G. Mayer, D. G. Butler, *Ecol. Modell.* **1993**, *68*, 21.
[43] J. Norton, *Environ. Model. Softw.* **2015**, *69*, 166.
[44] R. G. Sargent, in *Winter Simul. Conf. Proc* (Eds: L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, M. D. Rossetti), IEEE, Huntington Beach, CA **2015**, p. 1729.
[45] L. Chiang, B. Lu, I. Castillo, *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8*, 63.
[46] A. A. Linninger, S. Chowdhry, V. Bahl, H. Krendl, H. Pinger, *Comput. Chem. Eng.* **2000**, *24*, 591.
[47] S. Sierla, L. Sorsamäki, M. Azangoo, A. Villberg, E. Hytönen, V. Vyatkin, *Appl. Sci* **2020**, *10*, 6959.
[48] J. Lang, S. Crook, *Enabling the Industry 4.0 Ecosystem with Edge-to-Cloud Data Management*. **2021** https://www.aveva.com/content/dam/aveva/documents/reports/IDC-Technology-Spotlight-Enabling-the-Industry-4.0-Ecosystem-with-Edge-to-Cloud-Data-Management.pdf (accessed: January 2024).
[49] M. Balzan, *Expanding its Digital Information Infrastructure Helps the Pulp, Paper and Forest Products Industry Take Advantage of New Market Dynamics*, AVEVA Group Limited, Cambridge, UK **2023**.
[50] J. A. Romagnoli, M. C. Sanchez, in *Data Processing and Reconciliation for Chemical Process Operations* (Eds: G. Stephanopoulos, J. Perkins), Academic Press, Inc., Orlando, FL **1999**.
[51] T. Kourti, J. F. Macgregor, *ASQ WCQI*, American Society for Quality, Milwaukee, WI **2004**, p. 449.
[52] S. Xu, B. Lu, M. Baldea, T. Edgar, W. Wojsznis, T. Blevins, M. Nixon, *Rev. Chem. Eng.* **2015**, *31*, 453.
[53] É. Thibault, M. Chioua, M. McKay, M. Korbel, G. S. Patience, P. R. Stuart, *Can. J. Chem. Eng.* **2023**, *101*, 6055.
[54] É. Thibault, F. L. Désilets, B. Poulin, M. Chioua, P. Stuart, *Comput. Chem. Eng.* **2023**, *178*, 108380.
[55] É. Thibault, J. D. Kelly, F. Lebreux Desilets, M. Chioua, B. Poulin, P. Stuart, *Processes* **2023**, *11*, 2376.
[56] X. Chu, I. F. Ilyas, S. Krishnan, J. Wang, in *Proc. ACM SIGMOD Int. Conf. Manag. Data* (Eds: F. Özcan, G. Koutrika, S. Madden), Association for Computing Machinery, San Francisco, CA **2016**, p. 2201.
[57] M. Amazouz, Improving Process Operation Using the Power of Advanced Data Analysis. **2015** https://natural-resources.canada.ca/maps-tools-publications/publications/energy-publications/publications/explore-software-improving-process-operation-using-power-advanced-data-analysis/17883 (accessed: May 2024).
[58] J. Sall, JMP *(17.1)*, SAS, Cary **2023**.
[59] K. Palmer, M. Realff, *Chem. Eng. Res. Des.* **2002**, *80*, 773.
[60] V. Venkatasubramanian, *AIChE J.* **2019**, *65*, 466.
[61] D. K. Pace, *Johns Hopkins APL Tech. Dig.* **2004**, *25*, 163.
[62] J. K. Telford, *Johns Hopkins APL Tech. Dig.* **2007**, *27*, 224.
[63] R. Davis, P. John, in *Statistical Approaches with Emphasis on Design of Experiments Applied to Chemical Processes* (Ed: V. Silva), Intech Open, London **2018**.
[64] S. S. Garud, I. A. Karimi, M. Kraft, *Comput. Chem. Eng.* **2017**, *106*, 71.
[65] G. Franceschini, S. Macchietto, *Chem. Eng. Sci.* **2008**, *63*, 4846.
[66] A. K. Agarwal, M. L. Brisk, *Ind. Eng. Chem. Process Des. Dev.* **1985**, *24*, 203.
[67] Clarivate Analytics Web of Science™ Core Collection. apps.webofknowledge.com (accessed: May 2024).
[68] L. F. Santos, C. B. Costa, J. A. Caballero, M. A. Ravagnani, *Appl. Energy* **2020**, *262*, 114441.
[69] P. Guo, T. Li, P. Li, Y. Zhai, J. Li, *Desalination* **2020**, *473*, 114195.
[70] A. Kumar, A. Bhattacharya, J. Flores-Cerrillo, *Comput. Chem. Eng.* **2020**, *136*, 106756.
[71] H. Anisi, S. Shahhosseini, A. Fallah, *Can. J. Chem. Eng.* **2022**, *100*, 476.
[72] K. C. S. Rodrigues, I. I. K. Veloso, M. P. A. Ribeiro, A. J. G. Cruz, A. C. Badino, *Can. J. Chem. Eng.* **2021**, *99*, 401.
[73] J. G. S. S. Maia, R. B. Demuner, A. R. Secchi, P. A. Melo, R. W. do Carmo, G. S. Gusmão, *Ind. Eng. Chem. Res.* **2018**, *57*, 6401.
[74] A. A. Adetayo, J. D. Litster, I. T. Cameron, *Comput. Chem. Eng.* **1995**, *19*, 383.
[75] W.-B. Su, J.-R. Chang, *Ind. Eng. Chem. Res.* **2000**, *39*, 4140.
[76] A. Martinho, H. A. Matos, R. Gani, B. Sarup, W. Youngreen, *Food Bioprod. Process.* **2008**, *86*, 87.
[77] F. Yan, X. Zhang, C. Yang, B. Hu, W. Qian, Z. Song, *Can. J. Chem. Eng.* **2023**, *101*, 4506.
[78] R. Gholami, A. Alvarez-Majmutov, M. Ali, J. Chen, *Can. J. Chem. Eng.* **2021**, *99*, 222.
[79] G. Li, H. Sui, C. Xu, X. Li, *Asia-Pac. J. Chem. Eng.* **2013**, *8*, 507.
[80] M. Pérez-Fortes, A. Bojarski, S. Ferrer-Nadal, G. M. Kopanos, J. M. Nougués, E. Velo, L. Puigjaner, in *18th Eur. Symp. Comput.-Aided Chem. Eng* (Eds: B. Braunschweig, X. Joulia), Elsevier, Netherlands **2008**, p. 793.
[81] A. Schmidt, J. Strube, *Processes* **2019**, *7*, 298.
[82] O. Balci, R. E. Nance, J. D. Arthur, W. F. Ormsby, in *Winter Simul. Conf. Proc* (Eds: E. Yücesan, C.-H. Chen, J. L. Snowdon, J. M. Charnes), IEEE, San Diego, CA **2002**, p. 653.
[83] R. G. Sargent, O. Balci, in *Winter Simul. Conf. Proc* (Eds: W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. W. Evans, E. Page), IEEE, Las Vegas, NV **2017**, p. 292.
[84] *CADSIM Plus (3.4)*, Aurel Systems, Burnaby, Canada **2019**.
[85] P. Hart, M. Brown, *Brownstock Washing: Fundamentals and Practices*, TAPPI Press, Peachtree Corners, GA **2017**.
[86] *EXPLORE (2.0)*, CanmetENERGY, Varennes, Canada **2022**.

How to cite this article: C. Brucel, É. Thibault, G. S. Patience, P. Stuart, *Can. J. Chem. Eng.* **2025**, 1. https://doi.org/10.1002/cjce.25601