Check for updates

Received: 11 January 2024 | Revised: 2 April 2024 | Accepted: 14 May 2024

DOI: 10.1002/cjce.25374

MINI REVIEW

THE CANADIAN JOURNAL OF CHEMICAL ENGINEERING

Experimental methods in chemical engineering: Monte Carlo

Ergys Pahija[1] | **Soonho Hwangbo**[2,3] | **Thomas Saulnier-Bellemare**[4] | **Gregory S. Patience**[4]

[1]Department of Chemical and Biotechnological Engineering, Université de Sherbrooke, Sherbrooke, Québec, Canada

[2]Department of Chemical Engineering, Gyeongsang National University, Jinju, South Korea

[3]Department of Materials Engineering and Convergence Technology (BK21), Gyeongsang National University, Jinju, South Korea

[4]Department of Chemical Engineering, Polytechnique de Montréal, Montréal, Québec, Canada

Correspondence
Gregory S. Patience, Department of Chemical Engineering, Polytechnique de Montréal, Montréal, H3C 3A7 QC, Canada.
Email: gregory-s.patience@polymtl.ca

Abstract

Monte Carlo (MC) methods employ a statistical approach to evaluate complex mathematical models that lack analytical solutions and assess their uncertainties. To this end, techniques such as Markov chain Monte Carlo (MCMC), bootstrap, and sequential MC methods repeat the same operations over a specified range of conditions. Consequently, both the frequentist and Bayesian statistical approaches are computationally intensive, depending on the problem formulation. Improving sampling techniques and identifying sources of error reduce the computational demand but do not guarantee that the solution reaches the global optimum. Moreover, efficient algorithms and advances in hardware continue to decrease computation time. MC methods are applicable to a plethora of problems ranging from medicine to computational chemistry, economics, and industrial safety, making them integral to the ongoing industrial digitalization by evaluating the quality of applied models. In chemical engineering, MC simulations are used in four clusters of research: design, systems, and optimization; molecular simulation, including CO_2 and carbon capture; adsorption and molecular dynamics; and thermodynamics. There is limited cross-referencing between the design cluster and the other three, which presents an interesting area for future research. This mini-review presents two applications within chemical engineering: emissions and energy forecasting.

KEYWORDS

chemical engineering, computational chemistry, digitalization, Monte Carlo, process design

1 | INTRODUCTION

Mathematical models characterize natural phenomena, human behaviour, and project economics to forecast trends or expected viability over a wide range of operating conditions while minimizing computational cost. The reliability of models in engineering and science depends on the quality of the input data and their uncertainties to derive parameters and define appropriate model assumptions. Monte Carlo (MC) methods quantify such uncertainties and increase the confidence of the model forecasts. They repeat calculations at various input conditions to

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *The Canadian Journal of Chemical Engineering* published by Wiley Periodicals LLC on behalf of Canadian Society for Chemical Engineering.

generate a statistical dataset, which quantifies the uncertainty of solutions.[1]

Besides conducting statistical analysis of complex systems, MC replaces computationally demanding applications such as those in computational chemistry while maximizing the fidelity of the model responses. MC also resolves complex integrals (quadrature), models buried in simulation packages (blackbox),[2] and it estimates model parameters for safety, economics, and industrial processes. In chemical engineering, models to represent physical phenomena are applied to design unit operations, develop process simulators, derive reaction kinetics, predict thermodynamic properties, and project plant economics. Robust plant models calculate return on investment (ROI), profitability (internal rate of return—IRR, net present value—NPV, payback time—t_{pb}) as a function of the factors that contribute to capital and operating expenditures. The estimate, $E_{t_{pb}}$, depends on a volume factor, Q, and variables that contribute to cost and profitability:

$$t_{pb} = f(\text{Q, capital investment, materials \& labour, depreciation, interest, ...}) \quad (1)$$

Market fluctuations contribute to the uncertainty in project feasibility as they are directly related to cost of materials (capital), labour, financing charges, and the final sales price of the product. MC simulations evaluate the risks associated with these fluctuations even for periods of global crises like high inflation after the pandemic and wars.[3]

Before exploring the theory of MC methods in chemical engineering, it is valuable to understand the historical context that allowed the development of these methods in scientific applications. In 1687, Jacob Bernoulli published a treatise related to the law of large numbers that is now the basis of MC. His brother, Johann Bernoulli, was the first to quantify probability. A century later, Pierre-Simon de Laplace evaluated π through an MC experiment (i.e., dropping needles)[2]; then, physicist Enrico Fermi applied sampling methods to his neutron studies, predicting experimental results accurately.[4] Nicholas Metropolis, Stanislaw Ulam, and John von Neumann applied statistical methods on neutron diffusion with the Electronic Numerical Integrator and Computer (ENIAC), among the first electronic computers.[2] In 1949, Metropolis and Ulam published a paper that gives the name to the renowned method.[5] The name comes from Ulam's interest in the game of chance, and it refers to the Monte Carlo district in Monaco.[2] Following Metropolis' publication, MC spread to the field of nuclear physics. It was only from the eighties that researchers applied MC in other fields like weather forecasting, finance, transportation, and medicine.

These historical events catalyzed the development of MC methods that will be presented in this mini-review. This article is part of a series on experimental methods in chemical engineering[6] including computational fluid dynamics–discrete element method (CFD-DEM),[7] artificial neural network (ANN),[8] density functional theory (DFT),[9] and process simulation.[10] We apply MC to assess uncertainty and sensitivity, and identify relationships between factors in experiments and systems rather than measuring the data with an instrument. In this context, model variables can change over time, whereas parameters are constants derived from experimental data. In some circumstances, model parameters need an update when the model behaviour changes, which is also standard in dynamic problems.

2 | GENERAL THEORY

MC is a technique for numerical integration, optimization, and inverse problems in chemical engineering. MC simulation is a non-deterministic numerical integration approach that produces outputs for a large number of runs (exceeding thousands of runs). It facilitates high-dimensional integrations while minimizing computational cost,[11] compares computed values from large sets of randomly drawn numbers, estimates parameters, quantifies uncertainties, and identifies the most sensitive parameters of a system. In the case of numerical simulation and optimization, MC solves the deterministic problems supported by probability theory.

2.1 | Sampling techniques

Random sampling draws independent values from a population or from a probability distribution.[12–14] The time between decays from a radioactive source is truly random phenomena, but MC relies on random number generators based on deterministic computer algorithms available in software like Excel, Python, or MATLAB. In Excel, INT(RAND()*100) returns random numbers from

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

0 to 99 while NORMINV(RAND(),50,20) returns random real numbers with a mean of 50 and a standard deviation of 20 following a normal distribution. When compared to fully randomized sampling, Latin Hypercube achieves better coverage of the sampling space.[15] It divides the sampling space into intervals and draws one sample from each interval. The quasi-random method produces a more uniform distribution of the samples than pure random sampling. It divides the sampling space into intervals and then draws samples on a base 2 to form finer sequences, called Sobol sequence, but is limited by its high computational demand when dealing with multi-dimensional problems.

The sampling method depends on the application but must cover the entire sampling space while considering any eventual correlation within the space (Figure 1). Copula methods isolate variables in complex systems that have joint probabilities (or correlations).[16]

When sampling from a normal distribution or other function, the samples are unevenly distributed.[13] The expectation of samples drawn from a function, $f(\mathbf{x})$, with a probability distribution $p(\mathbf{x})$ is the integral:

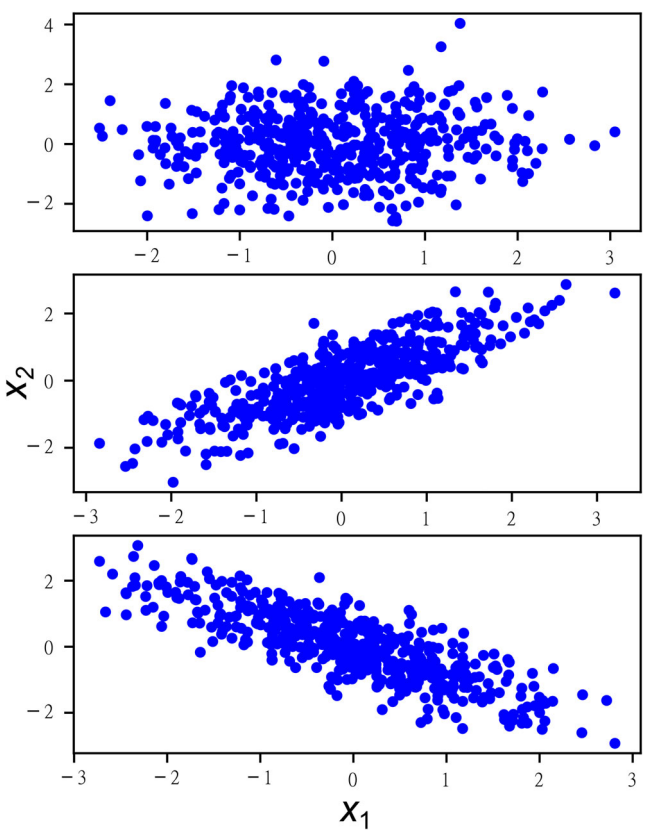

FIGURE 1 Examples of sample space correlation between factors x_1 and x_2 with the NumPy's generator (blue dots). Top: no correlation between variables, correlation index $\rho = 0$. Middle: a negative correlation, $\rho = -0.8$. Bottom: a positive correlation, $\rho = 0.8$.

$$E[f(\mathbf{x})] = \int p(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \tag{2}$$

The average, $\bar{f}$, of the function for a large number of samples approximate the true integral. Repeating the operation for a new set of samples returns a different $\bar{f}$, which is normally distributed around $E[f(\mathbf{x})]$ with variance $\mathrm{Var}[\mathbf{a}] = 1/N \mathrm{Var}[f(\mathbf{x})]$. Large samples (>50) approximate the mean and standard deviation to within about 10%–20% (for NORMINV(RAND(), $\bar{f}$, σ)). Other common sampling techniques include importance sampling (for overly complicated distributions), rejection sampling, slice sampling, Markov chain Monte Carlo (MCMC), and efficient Monte Carlo methods such as Hamiltonian Monte Carlo and the over-relaxation for Gibbs sampling.[13]

2.2 | Sensitivity analysis

Local sensitivity analysis validates a parameter estimation problem and assesses the impact of input variables on the output of a model. To perform a sensitivity analysis using MC simulation, a sampling space is defined based on a range of percentage changes of the value of the input variable for each analyzed variable. A large number of input conditions are generated for the model from the sample space, which produces a large number of outputs. We plot the outputs to identify local first-order partial derivatives. When the derivative of an output value is positive, the model output increases when incrementing the parameter and decreases when the derivative is negative. The magnitude of the change is proportional to the influence of the parameter on the output. Statistical and numerical methods adjust the local derivatives to global sensitivity indexes while accounting for the simulation noise.[17,18]

2.3 | Uncertainty analysis

MC quantifies the uncertainty of a measure or a model by defining a function, $f(\boldsymbol{x})$, and linking input variables, x, to a response, y. The definition of a sample space or a sampling distribution for each of the input variables involves drawing a large number of samples, usually from a probability distribution processed through the function $f(\boldsymbol{x})$. The standard deviation of the response is the uncertainty of the function. Consider, for example, the uncertainty of the length of 30 mm ruler. The function $y = x_2 - x_1$ gives the length, where x_1 is the measure at 0 mm and x_2 is the measure at 30 mm. Both measures have an error of 1 mm. Samples for x_1 and x_2 are drawn from normal probability distributions centred around the

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

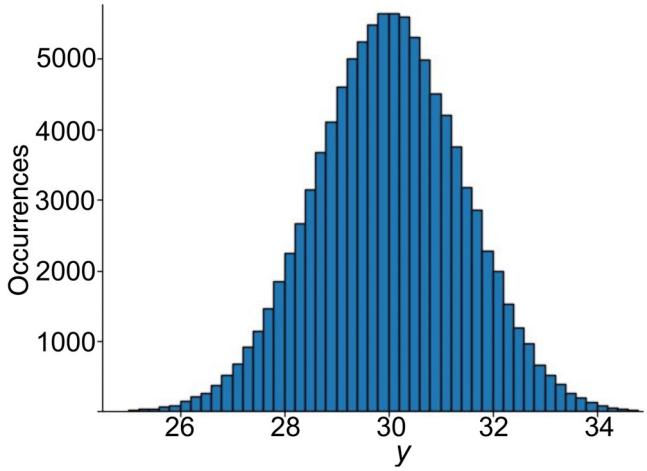

FIGURE 2 Monte Carlo (MC) distribution of length for the example of a 30 mm ruler.

means (0 or 30 mm) with a standard deviation of 1 mm. We calculate the length, y, 100,000 times with different variables (Code Listing 1). The mean resulting length is 30.0 mm with a standard deviation of 1.4 mm (Figure 2).

Code Listing 1: Python code example for the ruler example.

```
def myfun(x1,x2):
    f = x2-x1
    return f

mean = np.array([0,30]); std = ([1,1])
n_samples = 100000; n_variables = 2
x = np.zeros([n_samples,n_variables])
x1 = np.random.normal(mean[0],std[0],n_samples)
x2 = np.random.normal(mean[1],std[1],n_samples)
f = myfun(x1,x2)
```

The theory of error propagation is a common tool to estimate the uncertainties of measurements.[19] The error propagation considers the error introduced by measured variables, which translates to:

$$\Delta_f^2 = \left(\frac{\partial f}{\partial x_1}\Delta_1\right)^2 + \left(\frac{\partial f}{\partial x_2}\Delta_2\right)^2 + \cdots + \left(\frac{\partial f}{\partial x_n}\Delta_n\right)^2. \quad (3)$$

Considering the uncertainty of the position of the two endpoints is each 1 mm, the uncertainty, Δ_f, for the function $y = x_2 - x_1$ becomes:

$$\Delta_y = \sqrt{(\Delta_1)^2 + (\Delta_2)^2} = \sqrt{2} = 1.4\,\text{mm} \quad (4)$$

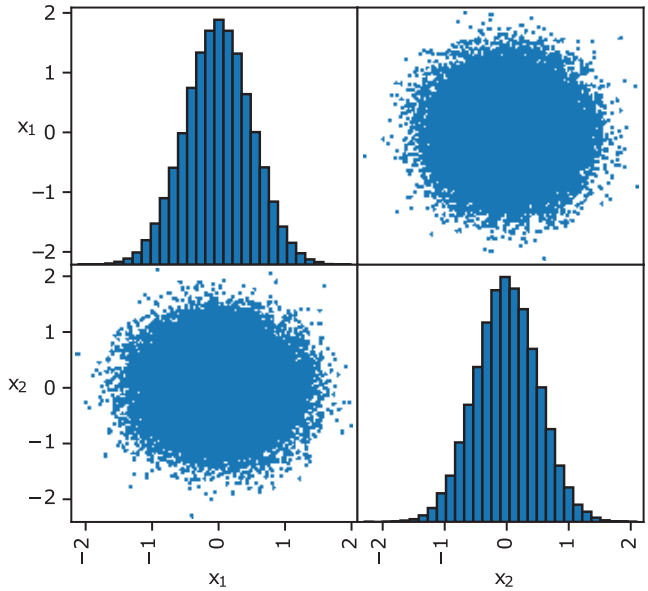

FIGURE 3 Samples collected for each parameter and representation of their distribution.

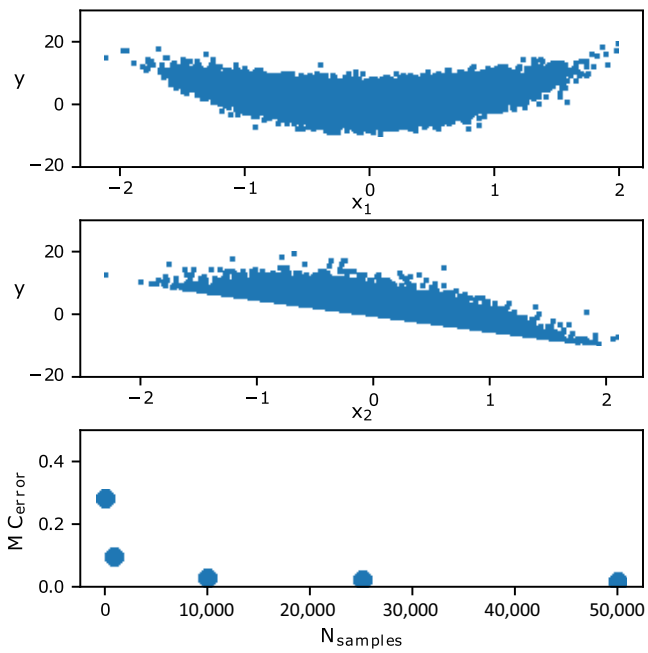

FIGURE 4 Monte Carlo (MC) simulation for the example. Plot for each variable against the function value (top and middle). MC error change with different number of trials (bottom).

In another example, the following steps estimate the uncertainty of $y = 4x_1^2 - 5x_2$:

1. Define the range and the distribution of the parameters (in this case assuming a normal distribution),
2. Sample from the input space (e.g., random sampling) (Figure 3),

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. Perform the MC simulation on n samples (Figure 4), and
4. Review the results.

The MC error represents the standard error of the simulation and it depends on the square root of the number of trials ($s/\sqrt{N}$).

Given a function $f(\boldsymbol{x})$, where $\boldsymbol{x}$ depends on d variables, the integral of the function can be approximated through MC. For a sampling number n tending to infinity, the estimates equals the integral (Equation (5)).

$$I = \int f(\boldsymbol{x})dx = \lim_{n\to\infty}\frac{1}{n}\sum_{1}^{n} f(x_n) \tag{5}$$

The MC integration error drops with increasing n by $\sigma(f)/\sqrt{(n)}$, where the standard deviation $\sigma(f)$ can be obtained from the sample variance (Equation (6)), where E represents the estimate.

$$\sigma^2(f) \approx s^2(f) = \frac{1}{n-1}\sum_{1}^{n}(f(x_n)-E)^2 \tag{6}$$

2.4 | Parameter estimation problem

MC simulation estimates the sensitivity of a parameter vector, θ, on the system, f, and a state variable x (Equations (7) and (8)).[20]

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = f(\boldsymbol{x},\boldsymbol{\theta},t) \tag{7}$$

$$\boldsymbol{y} = g(\boldsymbol{x},\boldsymbol{\theta},t) \tag{8}$$

2.4.1 | Frequentist approach

In the frequentist approach, the model parameters $\boldsymbol{\theta}$ are treated as true and fixed values while their estimators $\widehat{\boldsymbol{\theta}}$ are considered as random variables occupying a sampling space.[20] The output vector, $\boldsymbol{y}$, is:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\theta}) + \epsilon \tag{9}$$

where ϵ is the measurement error defined as Gaussian white noise $\epsilon \sim N(0,\sigma^2)$.

In a parameter estimation problem, the maximum likelihood estimation (MLE) maximizes the likelihood function between the model estimated with the random parameters $\boldsymbol{\theta}$, and the data (Equations (10) and (11)).

$$L(\boldsymbol{y},\boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{[\boldsymbol{y}-f(\boldsymbol{\theta})]^2}{2\sigma^2}\right) \tag{10}$$

$$\widehat{\boldsymbol{\theta}} \equiv \arg\max\ L(\boldsymbol{y},\boldsymbol{\theta}) \tag{11}$$

The least squares method (LSM) is a special case of the MLE problem, where the objective is the minimization of the squared difference between the data and the model results (Equations (12) and (13)). Traditional minimization algorithms solve this type of problem.

$$S(\boldsymbol{y},\boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\boldsymbol{y}-f(\boldsymbol{\theta}))^2}{2\sigma^2}\right) \tag{12}$$

$$\widehat{\boldsymbol{\theta}} \equiv \arg\min\ S(\boldsymbol{y},\boldsymbol{\theta}) \tag{13}$$

A covariance matrix shows the joint variability of two or more random variables, by a first- or second-order approximation (Equation (14)).[21]

$$\mathrm{Cov}\left(\widehat{\boldsymbol{\theta}}\right) = s^2(\boldsymbol{F}'\boldsymbol{F})^{-1} \text{ where } \boldsymbol{F} = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{14}$$

where $s^2 = S_{min}(\boldsymbol{y},\boldsymbol{\theta})/(n-p)$ is the unbiased estimation of σ^2 from residual parameters, n is the number of measurements, and p is the number of parameters. Following the linear regression theory, a general model is written as $\boldsymbol{Y} = \boldsymbol{X\theta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is an independent variable vector, $\boldsymbol{Y}$ is a dependant variable vector, $\boldsymbol{\theta}$ is a vector of unknown parameters, and $\boldsymbol{\epsilon}$ is the error term. The LSM estimate of $\boldsymbol{\theta}$ is given by the minimum of $S(\boldsymbol{\theta})$. The estimate in matrix notations becomes $\widehat{\boldsymbol{\theta}}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. The residual error is $\boldsymbol{\epsilon} = \boldsymbol{X}\widehat{\boldsymbol{\theta}}$. Thus, through LSM, the error of the estimated parameters $\widehat{\boldsymbol{\theta}}$ decreases as n increases, highlighting the need for a high number of samples during MC-assisted parameter estimations. The non-linear case follows a similar approach.

The confidence interval 1-α of predictions is function of the covariance matrix (Equation (15)). For large n, such as those used for MC simulations, the confidence interval for parameters assumes a t-distribution.[22]

$$\boldsymbol{Y}_{\mathbf{1}-\boldsymbol{\alpha}} = \boldsymbol{Y} \pm \sqrt{\mathrm{diag}(\mathrm{Cov}(\boldsymbol{Y}))}\cdot t(N-m,\alpha/2) \tag{15}$$

2.4.2 | Bayesian approach

The Bayesian approach considers parameters $\boldsymbol{\theta}$ as following probability statements.[20] The Metropolis and Metropolis–Hastings algorithms, as an example, estimate parameters to build a model explaining a set of data

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

points $\boldsymbol{x}$. They accept or reject random values of $\boldsymbol{\theta}^*$ generated by MCMC techniques using Bayes' theorem, which connects a set of independent observations $\boldsymbol{x}$ to a set of probability distribution parameters $\boldsymbol{\theta}$ by relating the posterior $p(\theta|x)$ with the likelihood $p(x|\theta)$, the prior $p(\theta)$, and the evidence $p(x)$ (Equation (16)).

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)} \tag{16}$$

$p(\theta)$ consists of a distribution function (normal, beta, binomial, ...) defined during the initialization of the problem. The likelihood $p(x|\theta)$ is usually evaluated through a transition kernel similar to those for MLE and LSM (Equation (17)).

$$\frac{1}{\sqrt{2\pi}} \exp\left(-0.5(\boldsymbol{x}-\boldsymbol{\theta})^2\right) \tag{17}$$

Finally, the probability of evidence $p(x)$ is expressed by the integral:

$$p(x) = \int p(\theta)p(x|\theta)d\theta \tag{18}$$

The integral is approximated by $\frac{1}{n}\sum p(X|\theta)$. Thus, the posterior $p(\theta|x)$ estimation is possible for any θ. MCMC technique generates a new random θ^* by using θ_{t-1} as an input for the probability distribution $p(\theta)$, from which we draw a random number from this distribution. The Metropolis algorithm accepts or rejects the new θ^* based on the comparison between the maximum acceptance probability 1 and the ratio r of $p(\theta^*|x)$ and $p(\theta_{t-1}|x)$. If $r > 1$, the new set of θ^* is accepted and is used for another iteration, while if $r < 1$, the θ^* is rejected and a new random θ^* has to be selected. The Metropolis–Hastings algorithm uses a second acceptance criterion for which the smallest value between the maximum acceptance probability 1 and the ratio r of $p(\theta^*|x)$ and $p(\theta_{t-1}|x)$ is compared to a randomly drawn number u from a uniform distribution between 0 and 1. θ^* is accepted if u is smaller than the other compared number. This means that the set of parameters has improved since the last iteration. The Metropolis and Metropolis–Hastings algorithms generate a distribution of parameters which can be used to build a model explaining a set of data points $\boldsymbol{x}$.[13]

2.4.3 | Bootstrap method

The bootstrap method evaluates parameters and their uncertainties from a set of experimental data.[23] The first step is to estimate parameters by either the MLE or LSM method. The procedure then randomly generates synthetic sets of data from the distribution of the existing data set. New parameters are evaluated for each new data set, for which the error is evaluated by comparing the parameter estimates from the original data set $\widehat{\boldsymbol{\theta}}$ to the parameter estimates from the synthetic MC set $\widehat{\boldsymbol{\theta}}^{MC}$.

The parameter estimates and their uncertainty provide the relevant statistical descriptors to build a symmetric correlation matrix which represents the relationship between parameters present in the problem. By increasing the number of sets of evaluated parameters $\widehat{\boldsymbol{\theta}}^{MC}$, the error on the mean and standard deviation decrease substantially. Other deviations from the traditional MC methods for parameter estimation were presented by Luengo et al.[24]

3 | APPLICATIONS

Most science categories have adopted Monte Carlo simulation massively: over 200 of the 250 categories reference it at least 30 times. Electrical electronic engineering leads with 37,000 of the 336,000 mentions in the bibliographic database, followed by physical chemistry and applied physics (each with 30,000 occurrences).[25] This section presents MC applications in chemical engineering, with specific application in processes and energy industry (including sustainable applications), oil & gas, stretching into computational chemistry, biology, energy, environmental science, and medicine.[26] Chemical engineering is ranked 34th with almost 6000 articles with MC as a keyword under the **Topic** search criteria. The earliest research dates back to the mid-1960s, with a couple of articles per year; 1998 was the first year with over 100 articles in chemical engineering, and it exceeded 300 articles a year in 2018. Each category has libraries like Speramint, GPyOpt, and BoTorch.[27–29] BoTorch couples the MC acquisition function with Bayesian optimization It exploits the sample average approximation while combining sampling with deterministic optimization and variance reduction techniques.[30]

To assess the major research topics in engineering, we queried the database (2013–2022) and generated a bibliometric map with VOSViewer (2900 articles, Figure 5). The map groups keywords that most often appear in the same articles into clusters with the same colour and links keywords that cross- reference these articles with lines. The size of the nodes (circles) correlates with the number of articles with that keyword. The red cluster has the most keywords with 43 and includes uncertainty-249 articles, optimization (opt'n)-232, systems-209, design-156 and

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

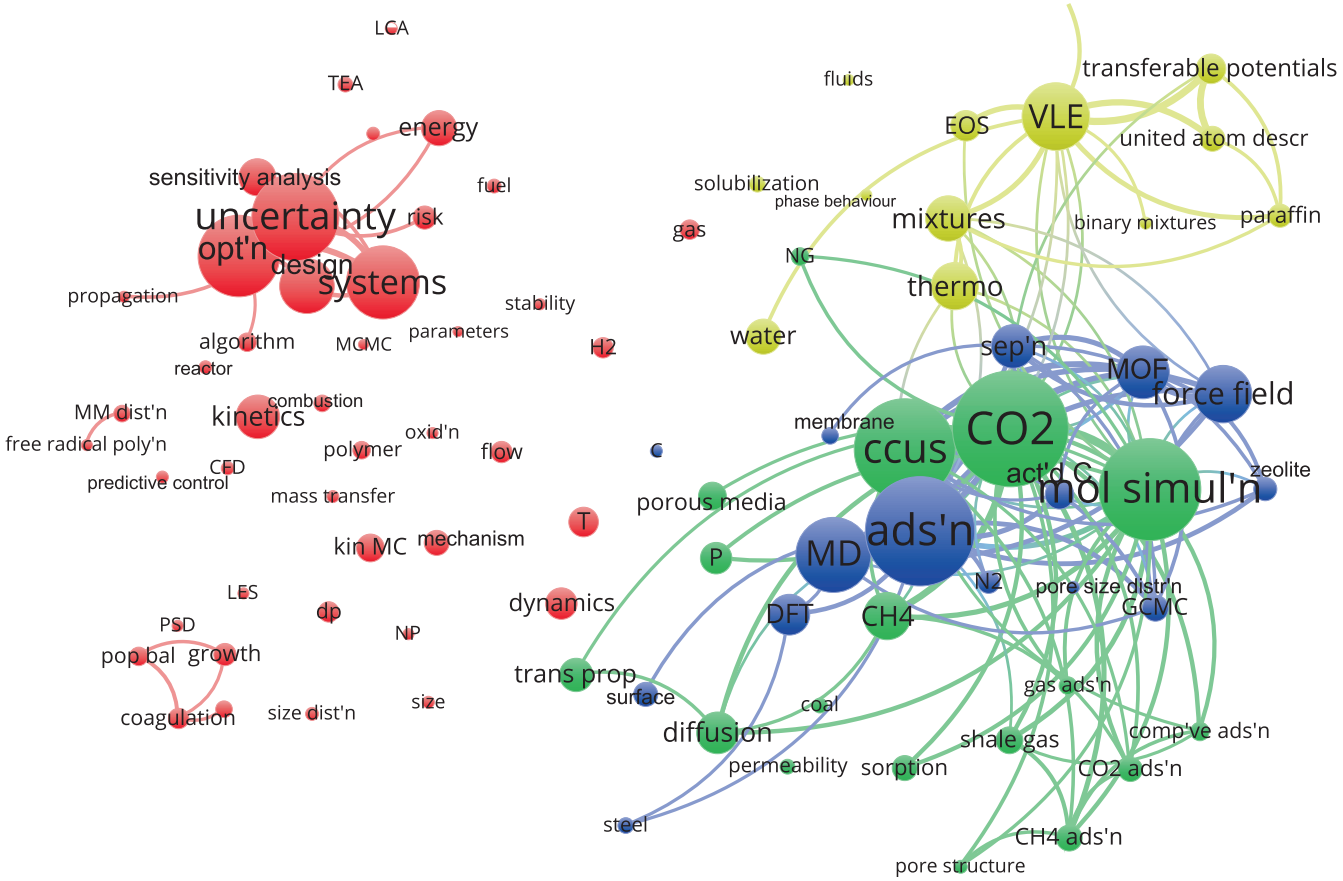

FIGURE 5 Bibliometric map of keywords created with VOSviewer software.[32] The data were derived from articles indexed in Web of Science[25] with Monte Carlo as search criteria in Topics within the Chemical Engineering Category. The areas of the circles are directly proportional to the number of keywords occurrences, while related research topics are grouped into coloured clusters. act'd C, activated carbon; ads'n, adsorption; aggreg'n, aggregration; C, carbon; ccus, carbon capture utilization sequestration; CFD, computational fluid dynamics; CH4 ads'n, methane adsorption; CO_2 ads'n, carbon dioxide adsorption; comp've ads'n, competitive adsorption; DFT, density functional theory; dp, particle diameter; EOS, equation of state; free radical poly'n, free radical polymerization; gas ads'n, gas adsorption; GCMC, gas chromatography mass spectrometry; kin MC, Monte Carlo kinetics; LCA, life cycle assessment; LES, large eddy simulation; MD, molecular dynamics; MOF, metal organic framework; mol simul'n, molecular simulation; MM dist'n, molar mass distribution; NG, natural gas; NP, nanoparticle; opt'n, optimization; oxid'n, oxidation; *P*, pressure; pop bal, population balance; PSD, particle size distribution; sens analysis, sensitivity analysis; sep'n, separation; *T*, temperature; TEA, technoeconomic analysis; thermo, thermodynamics; trans prop, transport properties; VLE, vapour-liquid equilibrium.

kinetics-124—applied chemical engineering. The other clusters have about the same number of keywords: green-18, blue-16, and yellow-13. CO_2 (330), molecular simulation (mol simul'n)-290, and carbon capture utilization sequestration(ccus)-286 dominate the green cluster, while adsorption (ads'n)-310, molecular dynamics (MD)-213, and force field-164, and vapour-liquid equilibrium (VLE)-192, thermodynamics (thermo)-136, and mixtures-130 are the main keywords in the blue and yellow clusters. Most of the links (cross-citations) are between the green, blue, and yellow clusters, with very few to the red cluster: of the 200 links, only 6 cross to the red cluster from the others. This indicates strong connection between physico-chemical applications but a weak link to the more traditional engineering topics. Future research should target relating process optimization and systems with chemistry (experimental or computational) and thermodynamic research. Indeed, sustainable process intensification exploits computational chemistry and experimental data will be a fundamental aspect in the development and adoption of carbon capture, utilization, and storage (CCUS) and will advance their technology readiness level towards pilot and industrial applications.[31]

Applied Energy is the chemical engineering journal that has published more articles than any other journal with 291 followed by *Industrial & Engineering Chemistry Research* (209), *Chemical Engineering Science* (161), *Fluid Phase Equilibria* (146), and *Chemical Engineering Journal* (120). *The Canadian Journal of Chemical Engineering* published 17 in the same time frame (2013–2022), but

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

published seven in 2023 with a heavy emphasis on polymers: 'Probabilistic Assessment of the Safety Profile of the Fischer-Tropsch Process with a Supercritical Solvent',[33] 'Use of Molecular Simulation for Evaluating Adsorption Equilibrium of Inhalation Anaesthetic Agents on Metal-Organic Frameworks',[34] 'Propagating Input Uncertainties Into Parameter Uncertainties and Model Prediction Uncertainties-A Review',[35] 'Early Bird Gets the Network: The Relative Importance of Reactivity Ratios, Ψ Parameter, and Crosslinker Level on Gel Formation in FRP',[36] 'Modelling Textural and Mass Transfer Properties for Gamma-Alumina Catalysts Using Randomly Generated Pore Networks',[37] 'Monte Carlo Simulation of Terpolymerization: Optimizing the Simulation and Post-Processing Times',[38] and 'Mathematical Modelling of Enzymatically Cross-Linked Polymer-Phenol Conjugates Using Deterministic and Stochastic Methods'.[39]

MC contributes to many chemical engineering problems like model identification to demonstrate how uncertainties from data and variables propagate in model predictions (red cluster).[40] When dealing with process analysis, MC evaluates model quality and stability by testing it on multiple simulations/predictions. Thus, there are two challenges associated with the application of MC: (1) computational cost and (2) problem formulation.[40] The development of surrogate models, efficient sampling strategies/algorithms, and parallel computing tackle the first challenge, while identifying the uncertainty sources for the sensitivity and uncertainty analysis deals with the problem formulation.[40]

3.1 | Chemical and bio-processes (red cluster)

The design of experiments (DOE) maximizes the research output by reducing the number of required experimental runs.[41] The Bayesian design of experiments applies an MC approximation and Gaussian process emulators to approximate the posterior distribution.[41] The response surface methodology is a common technique used for DOE and optimization (red cluster).[42,43] Response surface methodology combined with dynamic MC predicted how fast formic acid decomposed on a Cu surface as a function of T, P, and t (time).[42] Employing DOE and quantum MC simulations discovered candidate compounds as a corrosion inhibitor of steel.[43]

In analytical techniques applicable to chemical processes, MC can boost scanning electron microscope analysis by generating 3D images based on electron–sample interaction. In material science, a MC grain boundary optimization algorithm determines the lowest energy structure of SiO_2 and SiC.[44] Ion radiations cause thermal spikes and atomic relocation within amorphous structures (e.g., TiO_2). MC combined with molecular dynamics evaluated the displacement in such amorphous structures.[45]

In polymeric (red cluster) applications, a MC-simulation-based optimization and an adaptive simulation were tested to maximize the comonomer conversion ratio to achieving the target chemical composition distribution.[46] Kinetic MC tracks the arrangement of monomers and polymer chains and, when coupled with optimization algorithms, detects the most promising polymer microstructures.[47] Other applications of kinetic MC include chemical transport, structure and material property characterization, and catalysis hierarchical multiscale modelling.[48]

MC is often employed in bioprocesses. MC methods consider uncertainties for each component in biomass to better approximate pyrolysis compared to an only-cellulose approximation (red cluster—with combustion).[49] The traditional and sequential MC have been applied to estimate the parameters of a mechanistic model to monitor lactic acid bacteria fermentation,[50,51] whereas similar models can estimate parameters of complex biological population balances.[52] Further applications in biology and medicine applied active inference to understand biological intelligence or studying high energy beams in radiotherapy to assess the risk associated with radiation exposure.[53–56]

3.2 | Techno-economic analysis (TEA) and life-cycle assessment (LCA) (red cluster)

Chemical plant economics (CAPEX and OPEX) depends on feedstock and product prices, which vary with time and depend on location.[57,58] To estimate the sensitivity of feedstock price and operating parameters (temperature, pressure, and fuel options–electricity vs. methane) effect on economics, the cash flow is calculated based on random or random-normal samples of all prices and/or parameters.[57] The simulation determines the price intervals for each component to verify the profitability.[59] The economic evaluation of large plants operating with many compounds is constrained by computational resources (i.e., time, memory), thus requiring the selection of key prices (parameters) to generate scenarios and develop risk management strategies.[60] For example, researchers maximized selectivity and profitability of azelaic acid from vegetable oil rather than minimizing process steps. The MC simulation supported the decision to purchase higher-quality oil to reduce payback time and cut CAPEX.[60]

LCA commercial software incorporate uncertainties in their inventory data, where MC has been used to

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

calculate the bounds of the expected emissions and legal tolerance for populated areas.[61] Such methods have identified the most relevant parameters affecting the design, the economics, and the environmental impact.[62]

3.3 | Safety/risk (red cluster)

In mineral extraction applications, the risk of flyrocks induced by mine blasting can be assessed with a tree-based model and MC to calculate the distance distribution of debris.[63] Other MC algorithms are applied to position the blast ore boundaries.[64]

MC simulations guide laboratory research to contrast the producer's risk versus consumer risk: the uncertainty of HPLC assay of amoxicillin 500 mg tablets is an example. The decision-making problem demonstrated that the producer's risk is significantly higher than the consumer's risk.[65] In another HPLC example, MC-based model evaluated the solute polarity parameter depending on its molecular structure, whereas it optimizes HPLC procedures and methods.[66]

3.4 | Computational chemistry (blue cluster)

MC supports the evaluation of thermodynamic properties of atomistic systems to complement DFT (blue cluster).[9,67] Unlike molecular dynamics, MC is incapable of characterizing dynamic properties as the notion of time is absent.[9]

Characterizing adsorption is a major application of MC in chemical engineering (blue cluster).[68,69] High-performance computations incorporating molecular modelling with tools like grand canonical Monte Carlo (GCMC—blue cluster) characterize the adsorption process in crystalline adsorbents such as metal–organic frameworks and zeolites. Grounded in statistical mechanics principles, GCMC is a probabilistic tool to study equilibrium adsorption properties (e.g., isotherms) and evaluate structural characteristics (e.g., pore size distributions of the adsorbent, surface chemistry [yellow cluster]).[70] Energy storage in porous media is a common example of GCMC application. The GCMC method employs the grand canonical ensemble, allowing for dynamic adjustments in the number of molecules during simulations. Beyond energy exchange, this method finds primary application in systems where the particle count varies in response to external conditions.[71]

The computational drug design is another application where binding free energy estimation is essential. MC sampling methods combined with high-performance GPU simulations reduced the cost of molecular dynamics simulations (blue cluster), confirming the significance of combining sampling methods with regular simulation techniques to evaluate binding affinity estimates.[72]

The kinetic MC method (kin MC—red cluster) cut computational costs to simulate catalytic reactions. Furthermore, coupling MC with machine learning techniques identifies new catalytic compositions for chemical reactions.[73] The MC tree search algorithm combines the classical tree search algorithm with concepts of ML to synthesize new organic chemicals.[74] The MC tree search strategy accelerates the evaluation by three orders of magnitude.[75] To confirm the applicability of MC simulations, a comparison between auxiliary-field MC and other quantum chemistry methods, followed by geometry optimization, showed an agreement in equilibrium geometries with experimental values.[76]

3.5 | Industry 4.0 (red cluster)

Industrial digitalization is introducing advanced computational and monitoring technologies into traditional energy and production processes. The transition towards an optimal industrial reality consists of the selection and optimization of the most promising process technologies based on the risk assessment and the application of sustainable policies. MC methods to estimate the risk of a specific process or the probability that a specific process is uneconomic in the former case. This approach simulates process designs from a superstructure and selects the most efficient among all options.[77] For the latter case (optimization), some researchers reject MC to support decision-makers due to controversial simulation results.[78]

In renewable energy applications, sequential MC is computationally expensive compared to pseudo-sequential and non-sequential MC but simulates time series of variable energy sources and loads. Pseudo-sequential MC is computationally less demanding because of the sampling algorithm, while non-sequential MC is the most efficient but is incapable of simulating chronological aspects.[79] However, a novel pathway optimization method 'Global Calculator' model based on MCMC and genetic algorithm identifies sustainable pathways such as societal efforts to meet environmental goals (MCMC—red cluster). This type of information provides policymakers with valuable insights on which societal categories to address in future policies.[80]

4 | CASE STUDIES

This section presents a case study on predicting N_2O emissions and a scenario of forecasting renewable energy networks.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.1 | N_2O emissions

Wastewater treatment plants generate N_2O, a global warming contributor gas 300 times more powerful than CO_2.[81] Research constructs data driven models to characterize N_2O including dissolved oxygen (DO) $[g_{O_2}/m^3]$, ammonia (NH_4^+) $[g_N/m^3]$, nitrate (NO_3^-) $[g_N/m^3]$, N_2O concentration in the liquid phase (liquid-N_2O) $[g_N/m^3]$, temperature [°C], influent flowrate $[m^3/h]$, and air flowrate $[m^3/h]$ (Figure 6).[82,83] To forecast the dynamic emissions of N_2O requires an integrated framework including process modelling (like Aspen), process sensitivity analysis, signal processing, and fidelity.[84] Machine learning/deep learning approaches and MC simulation are capable of processing the data and then forecasting N_2O.[85,86] The main steps to generate data-driven models include:[87]

Data preparation: Gather data, merge data, data cleaning, and scaling.
Model training: The data has to be split (training vs. test) to train models. Grid search for the purpose of hyperparameters optimization also should be considered.
Model feasibility: Proposed models are evaluated versus statistical metrics such as coefficient of determination, root-mean-square error, mean absolute error, mean absolute percentage error, weighted absolute percentage error, Diebold-Mariano test, and information criteria.
Model analysis: MC-based global sensitivity analysis (i.e., generating random samples of inputs, performing MC simulations followed by returning Sobol sensitivity indices) results in a score of impacts on the developed model (Figure 6).

MC sampling offers flexibility and robustness in predicting N_2O emissions by accommodating uncertainties and variability in model inputs. It allows for quantification of risk, sensitivity analysis, and assessment of model performance, making it a valuable tool for decision-making in wastewater treatment plant management. Additionally, it can integrate expert knowledge and empirical data, enhancing the reliability of predictions.

4.2 | Energy forecasting

Many countries attempt to design renewable energy networks and reduce carbon emissions to create a sustainable world.[88] Decision-makers must examine various network scenarios forecasting future emissions.[89]

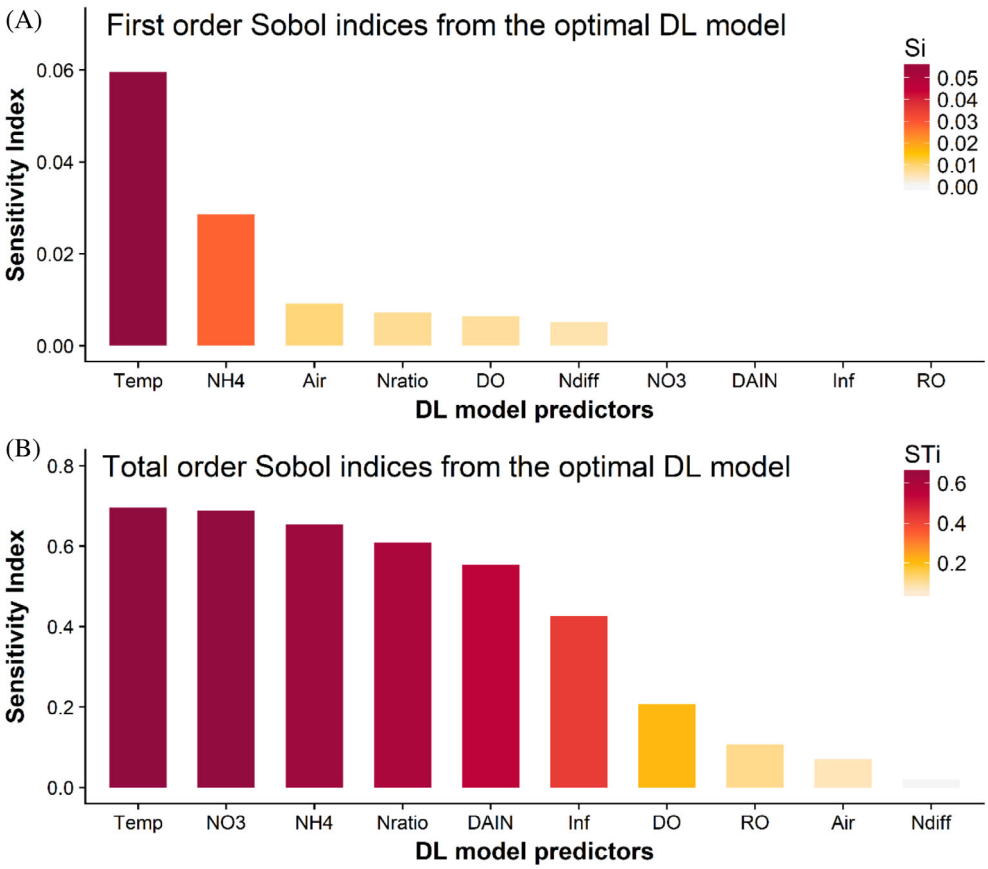

FIGURE 6 Results of global sensitivity analysis from deep learning-based model for biological systems: (A) the individual Sobol sensitivity indices and (B) the total Sobol sensitivity indices.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

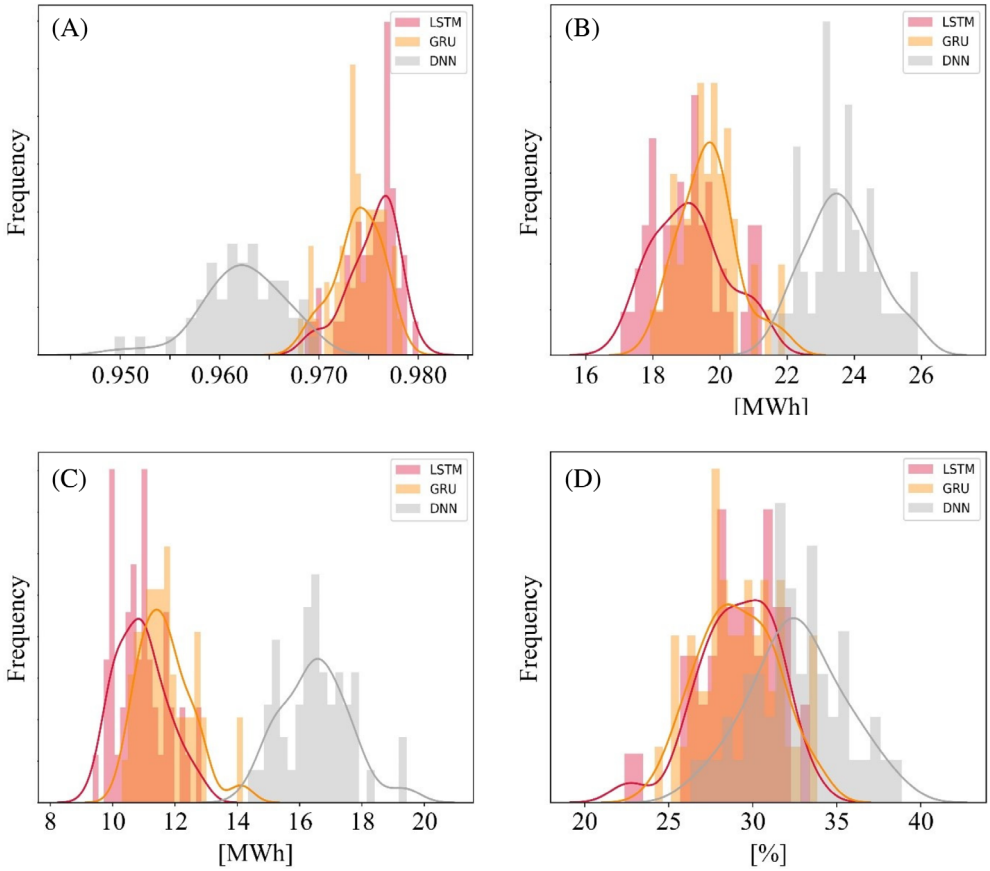

FIGURE 7 Distribution of performance metrics for renewable electricity supply forecasting models based on samples generated from deep learning-based sampling method (DNN, deep neural network; GRU, gated recurrent unit; LSTM, long short-term memory). Adapted with permission from Lee et al. 'Generative Model-Based Hybrid Forecasting Model for Renewable Electricity Supply Using Long Short-Term Memory Networks: A Case Study of South Korea's Energy Transition Policy'.[95]

The energy system models couple conventional sampling methods (e.g., random sampling, Latin Hypercube sampling with simple distributions).[90,91] In the case of renewable energy, fluctuations and unpredictable supply are inevitable. Advanced sampling methods based on stochastic machine learning characterize these uncertainties.[92] The variational auto-encoder (VAE) and generative adversarial networks (GANs) artificially generate a variety of samples that mimic real data.[93,94] The VAE trains a latent space between the encoder and decoder to produce new samples. GANs consist of the generative model and the discriminative model, generating high-quality samples by competing with each other. In doing so, they overcome the diversity of time-series data limitation and makes sense of scenario generation of renewables to contribute to energy policies. As regards one of the main results, a set of distributions of performance metrics for renewable forecasting models based on advanced sampling methods can be displayed (Figure 7).[95]

MC sampling is a general method for generating samples from a probability distribution to estimate quantities of interest, often used for uncertainty quantification. Advanced sampling based on stochastic machine learning, on the other hand, involves leveraging machine learning techniques to adaptively sample from complex and high-dimensional distributions. While both methods aim to sample from probability distributions, advanced sampling techniques typically employ sophisticated algorithms and may incorporate machine learning models for more efficient exploration of the parameter space compared to traditional MC methods.

5 | UNCERTAINTY

Mechanistic models rely on physico-chemical properties and phenomenological characterization to build an approximate model that describes complex systems (e.g., plastic deformation of biomass). Network models like MC techniques combined with molecular dynamics make the computation complex and time-consuming.[49] When increasing the number of simulations, the approximation towards the optimal solution improves. However, the global or local optimum evaluated through MC is not guaranteed to be the exact value. When the exact value is needed, other numerical methods represent a better option.

Despite the great utility of MC applications in uncertainty and sensitivity analysis, common inaccuracies and mistakes can be found in the literature[96]: Both uncertainty and sensitivity analysis should depend on a global exploration of the space of input factors.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

It is advisable to perform both uncertainty and sensitivity analysis. The uncertainty analysis informs of the robustness of the inference, ascertaining where volatility/uncertainty is coming from. Most models have many outputs, and these outputs can be used to answer a range of different questions. The relationship between the input factors and each different model output can vary significantly.[96] For this reason, it is essential to focus the sensitivity analysis on the question addressed by the model rather than just focusing on the model itself. Sensitivity and uncertainty analysis are themselves uncertain because there is considerable uncertainty in quantifying the uncertainty in input factors, thus not guaranteeing the absence of errors.

6 | CURRENT INDUSTRIAL APPLICATIONS AND FUTURE DIRECTIONS

MC methods find application to real industrial problems and guide industrial decision-makers in the selection of process operations.[97] These methods have a key role as monitoring approaches of dynamic systems, especially in resampling training data.[97] Data-driven models can make use of MC methods. For example, Bayesian neural networks (BNN) have the same topology as conventional ANN, but each parameter incorporates uncertainty.[97] Considering a trending topic such as CCUS that is transitioning to the industrial scale, MC plays a key role in multi-scenario studies, providing techno-economic, environmental, and supply chain estimations.[98] In another example that can be extended to chemical production, the uncertainties of cost demonstrated how Gen III nuclear plants today would bring significant economic losses in some countries.[99] Given the numerous applications presented in the previous sections, MC methods support economic, safety, and environmental decisions in the industry and these methods shine in fundamental applications in science and R&D, including chemical engineering. The areas of future development for MC in the industrial applications will concern economics and safety, and these methods will be integrated with fundamental scientific models. The fast growth of artificial intelligence applications will foster the integration of Bayesian approaches with deep learning algorithms and physical models to monitor, forecast, and simulate productions and economic scenarios.

7 | CONCLUSIONS

This mini review wants to emphasize the importance of MC methods in chemical engineering. MC supports engineers and scientists in analyzing complex problems and it proved its reliability in a wide range of chemical engineering applications. We presented examples of sampling methods including tools (e.g., Excel and Python), and uncertainty and sensitivity analysis to get started with MC simulations. The uncertainty and sensitivity analysis has a relevant role in engineering and defines the quality and reliability of mathematical models. Furthermore, we presented two case studies related to emissions and energy forecasting to demonstrate potential applications as well as the connection with trending artificial intelligence research. The application of MC methods relies on the development of an efficient algorithm and sampling methods whereas technological advances in computational capabilities will provide resources to run larger simulations when the necessity arises.

AUTHOR CONTRIBUTIONS

Ergys Pahija: Conceptualization; writing – original draft; writing – review and editing; methodology; visualization. **Soonho Hwangbo:** Conceptualization; methodology; writing – original draft; visualization. **Thomas Saulnier-Bellemare:** Writing – review and editing; visualization. **Gregory S. Patience:** Writing – review and editing; conceptualization; supervision; resources; visualization.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

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

REFERENCES

[1] A. Haghighat, *Monte Carlo Methods for Particle Transport*, CRC Press, Boca Raton, FL **2020**.
[2] W. L. Dunn, J. K. Shultis, *Exploring Monte Carlo Methods*, Elsevier, Amsterdam **2011**.
[3] A. Tobisova, A. Senova, R. Rozenberg, *Sustainability* **2022**, *14*, 8785.
[4] N. Metropolis, *Los Alamos Sci.* **1987**, *15*, 125.
[5] N. Metropolis, S. Ulam, *J. Am. Stat. Assoc.* **1949**, *44*, 335.
[6] G. S. Patience, *Can. J. Chem. Eng.* **2018**, *96*, 2312.
[7] A. Bérard, G. S. Patience, B. Blais, *Can. J. Chem. Eng.* **2020**, *98*, 424.
[8] J. Panerati, M. A. Schnellmann, C. Patience, G. Beltrame, G. S. Patience, *Can. J. Chem. Eng.* **2019**, *97*, 2372.
[9] H. Al-Mahayni, X. Wang, J.-P. Harvey, G. S. Patience, A. Seifitokaldani, *Can. J. Chem. Eng.* **2021**, *99*, 1885.
[10] J. D. Tommaso, F. Rossi, N. Moradi, C. Pirola, G. S. Patience, F. Galli, *Can. J. Chem. Eng.* **2020**, *98*, 2301.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[11] X. Wang, K.-T. Fang, *J. Complex.* **2003**, *19*, 101.
[12] Z. Wei, Y. Katoh, S. Ogasahara, Y. Yoshimoto, K. Kawai, Y. Ikeda, K. Eriguchi, K. Ohmori, S. Yoneda, presented at IEEE Int. Electron Dev. Meet. (IEDM), IEEE **2016**. pp. 4–8.
[13] B. Ghojogh, H. Nekoei, A. Ghojogh, F. Karray, M. Crowley, arXiv Preprint arXiv:2011.00901. **2020** https://arxiv.org/abs/2011.00901
[14] R. L. Harrison, presented at AIP Conf. Proc. Volume 1204, American Institute of Physics **2010**. pp. 17–21.
[15] Wiley StatsRef: Statistics Reference Online. https://www.wiley.com/learn/wileystatsref/pdf/WileyStatsRef-Sampler.pdf (accessed: June 2024).
[16] G. Masarotto, C. Varin, *Electron. J. Stat.* **2012**, *6*, 1517.
[17] R. Brun, M. Kühni, H. Siegrist, W. Gujer, P. Reichert, *Water Res.* **2002**, *36*, 4113.
[18] T. O. Drews, R. D. Braatz, R. C. Alkire, *J. Electrochem. Soc.* **2003**, *150*, C807.
[19] G. S. Patience, *Experimental Methods and Instrumentation for Chemical Engineers*, 2nd ed., Elsevier, Amsterdam **2017**.
[20] G. Sin, K. V. Gernaey, A. E. Lantz, *Biotechnol. Prog.* **2009**, *25*, 1043.
[21] M. Omlin, P. Reichert, *Ecol. Model.* **1999**, *115*, 45.
[22] J. Frutiger, C. Marcarie, J. Abildskov, G. Sin, *J. Chem. Eng. Data* **2016**, *61*, 602.
[23] P. Hall, J. Horowitz, *Annals Stat.* **1892**, *2013*, 41.
[24] D. Luengo, L. Martino, M. Bugallo, V. Elvira, S. Sarkka, *EURASIP J. Adv. Signal Process* **2020**, *2020*, 1.
[25] Clarivate Analytics, Web of Science™ Core Collection 2022. http://apps.webofknowledge.com (accessed: November 2022)
[26] A. Gasparini, K. R. Abrams, J. K. Barrett, R. W. Major, M. J. Sweeting, N. J. Brunskill, M. J. Crowther, *Stat. Neerlandica* **2020**, *74*, 5.
[27] D. G. Vlachos, *Adv. Chem. Eng.* **2005**, *30*, 1.
[28] R. Yondo, E. Andres, E. Valero, *Prog. Aerosp. Sci.* **2018**, *96*, 23.
[29] I. McCulloh, K. Kiernan, T. Kent, *Front. Big Data* **2020**, *3*, 37.
[30] M. Balandat, B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson, E. Bakshy, Adv. Neural Inf. Process. Syst. (NeurIPS). **2020**.
[31] E. Pahija, S. Golshan, B. Blais, D. C. Boffito, *Chem. Eng. Process. Intensif.* **2022**, *176*, 108958.
[32] N. J. V. Eck, L. Waltman, *Scientometrics* **2010**, *84*, 523.
[33] S. Herrera-Ovando, A. Jimenez-Gutierrez, N. K. Kazantzis, M. M. El-Halwagi, *Can. J. Chem. Eng.* **2024**, *102*, 548.
[34] M. Haeri-Nejad, M. Eic, *Can. J. Chem. Eng.* **2024**, *102*, 1646.
[35] K. Abdi, B. Celse, K. McAuley, *Can. J. Chem. Eng.* **2024**, *102*, 254.
[36] Y.-C. Lin, C. Liu, A. K. Tripathi, J. G. Tsavalas, *Can. J. Chem. Eng.* **2023**, *101*, 5382.
[37] S. Ferreira, J. J. Verstraete, E. Jolimaitre, D. Leinekugel-le-Cocq, C. Jallut, *Can. J. Chem. Eng.* **2023**, *101*, 1068.
[38] A. S. C. Rego, A. M. Amaral, A. L. T. Brandão, *Can. J. Chem. Eng.* **2023**, *101*, 5059.
[39] F. F. Karageorgos, A. C. Vasileiadou, C. Kiparissides, *Can. J. Chem. Eng.* **2023**, *101*, 4871.
[40] G. Sin, A. Espuña, *Front. Energy Research* **2020**, *8*, 1.
[41] A. M. Overstall, D. C. Woods, *Technometrics* **2017**, *59*, 458.
[42] M. Rafiee, H. Bashiri, *Catal. Commun.* **2020**, *137*, 105942.
[43] E. Naseri, M. Hajisafari, A. Kosari, M. Talari, S. Hosseinpour, A. Davoodi, *J. Mol. Liq.* **2018**, *269*, 193.
[44] M. Guziewski, A. D. Banadaki, S. Patala, S. P. Coleman, *Comput. Mater. Sci.* **2020**, *182*, 109771.
[45] A. Holm, S. G. Mayr, *Phys. Review B* **2021**, *103*, 174201.
[46] Y. Ma, X. Chen, L. T. Biegler, *Comput. Chem. Eng.* **2018**, *109*, 261.
[47] H. Gao, A. Waechter, I. A. Konstantinov, S. G. Arturo, L. J. Broadbelt, *Annals Stat.* **2018**, *108*, 268.
[48] M. Andersen, C. Panosetti, K. Reuter, *Front. Chem.* **2019**, *7*, 202.
[49] S. Hameed, A. Sharma, V. Pareek, H. Wu, Y. Yu, *Biomass Bioenergy* **2019**, *123*, 104.
[50] R. Spann, K. V. Gernaey, G. Sin, *Biochem. Eng. J.* **2019**, *151*, 107293.
[51] E. Pahija, R. Spann, G. Sin, *In Computer Aided Chemical Engineering*, Vol. 48, Elsevier, Milano **2020**, p. 1615.
[52] E. Pahija, P. Y. Lee, C.-W. Hui, *Process Integration Optim. For Sustain.* **2019**, *3*, 125.
[53] Z. Fountas, N. Sajid, P. A. Mediano, K. Friston, *Adv. Neural Information Processing Systems* **2020**, *33*, 11662.
[54] K. Kokurewicz, E. Brunetti, A. Curcio, D. Gamba, L. Garolfi, A. Gilardi, E. Senes, K. N. Sjobak, W. Farabolini, R. Corsini, D. A. Jaroszynski, *Commun. Phys.* **2021**, *4*, 1.
[55] K. Sheng, *Front. Medicine* **2020**, *14*, 431.
[56] V. Zumstein, J. Binder, S. Güsewell, P. Betschart, M. Pratsinis, G. Müllhaupt, L. Hechelhammer, H.-P. Schmid, D. Abt, *Eur. Urol. Focus* **2021**, *7*, 608.
[57] E. Martinez-Hernandez, L. F. Ramírez-Verduzco, M. A. Amezcua-Allieri, J. Aburto, *Chem. Eng. Research Design* **2019**, *146*, 60.
[58] P. Pérez-López, M. Montazeri, G. Feijoo, M. T. Moreira, M. J. Eckelman, *Sci. Total Environment* **2018**, *626*, 762.
[59] A. K. Srivastava, G. Kumar, P. Gupta, *Life Cycle Reliab. Saf. Eng.* **2020**, *9*, 77.
[60] G. A. D. L. Izeppi, J.-L. Dubois, A. Balle, A. Soutelo-Maria, *Ind. Crop. Prod.* **2020**, *150*, 112411.
[61] D. G. Kulas, M. C. Thies, D. R. Shonnard, *ACS Sustain. Chem. Eng.* **2021**, *9*, 5388.
[62] R. Mahmud, S. M. Moni, K. High, M. Carbajales-Dale, *J. Cleaner Prod.* **2021**, *317*, 128247.
[63] J. Ye, M. Koopialipoor, J. Zhou, D. J. Armaghani, X. He, *Natural Resour. Research* **2021**, *30*, 225.
[64] Z. Yu, X. Shi, X. Qiu, J. Zhou, X. Chen, Y. Gou, *Eng. Optim.* **2021**, *53*, 1467.
[65] L. Separovic, A. M. Saviano, F. R. Lourenço, *Measurement* **2018**, *119*, 41.
[66] A. M. Veselinović, J. B. Veselinović, G. M. Nikolić, A. P. Toropova, A. A. Toropov, *Struct. Chem.* **2016**, *27*, 821.
[67] R. E. Swai, *J. Petroleum Explor. Prod. Technol.* **2020**, *10*, 3515.
[68] A. Datar, M. Witman, L.-C. Lin, *AIChE J.* **2021**, *67*, e17447.
[69] A. P. Toropova, A. A. Toropov, A. Roncaglioni, E. Benfenati, *SAR QSAR Environ. Research* **2022**, *33*, 621.
[70] S. Saren, H. Chen, F. Miksik, T. Miyazaki, K. Thu, *Colloids Surfaces A: Physicochem. Eng. Aspects* **2023**, *684*, 133113.
[71] Y. Tao, G. Zhang, H. Xu, *Sustain. Mater. Technol.* **2022**, *32*, e00383.
[72] J. F. Gilabert, O. G. Carmona, A. Hogner, V. Guallar, *J. Chem. Inf. Model.* **2020**, *60*, 5529.
[73] E. Pahija, C. Panaritis, S. Gusarov, J. Shadbahr, F. Bensebaa, G. Patience, D. C. Boffito, *ACS Catal.* **2022**, *12*, 6887.
[74] M. Świechowski, K. Godlewski, B. Sawicki, J. Mańdziuk, *Artif. Intell. Review* **2023**, *56*, 2497.

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[75] S. Kajita, T. Kinjo, T. Nishi, *Commun. Phys.* **2020**, *3*, 1.
[76] M. Motta, S. Zhang, *J. Chem. Phys.* **2018**, *148*, 181101.
[77] R. Dickson, J. J. Liu, *Energy* **2021**, *221*, 119883.
[78] B. Crost, C. P. Traeger, *Econ. Lett.* **2013**, *120*, 552.
[79] A. Zakaria, F. B. Ismail, M. H. Lipu, M. A. Hannan, *Renewable Energy* **2020**, *145*, 1543.
[80] J. Garcia, O. Mwabonje, J. Woods, *Carbon Manag.* **2022**, *13*, 497.
[81] X. Chen, A. T. Mielczarek, K. Habicht, M. H. Andersen, D. Thornberg, G. Sin, *Environ. Sci. Technol.* **2019**, *53*, 12485.
[82] V. Venkatasubramanian, *AIChE J.* **2019**, *65*, 466.
[83] K. E. Mampaey, M. K. D. Kreuk, U. G. van Dongen, M. C. van Loosdrecht, E. I. Volcke, *Water Res.* **2016**, *88*, 575.
[84] E. Thibault, M. Chioua, M. M. McKay, M. Korbel, G. S. Patience, P. R. Stuart, *Can. J. Chem. Eng.* **2023**, *101*, 6055.
[85] S. Hwangbo, R. Al, X. Chen, G. Sin, *Environ. Sci. Technol.* **2021**, *55*, 2143.
[86] R. Al, C. R. Behera, A. Zubov, K. V. Gernaey, G. Sin, *Comput. Chem. Eng.* **2019**, *127*, 233.
[87] S. Hwangbo, R. Al, G. Sin, *Comput. Chem. Eng.* **2020**, *143*, 107071.
[88] W. Zappa, M. Junginger, M. V. D. Broek, *Appl. Energy* **2019**, *233*, 1027.
[89] H. Wang, Z. Lei, X. Zhang, B. Zhou, J. Peng, *Energy Convers. Manage.* **2019**, *198*, 111799.
[90] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, Hoboken, NJ **2008**.
[91] B. Iooss, P. Lemaître, *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and applications*, Springer, Boston, MA **2015**, p. 101.
[92] L. Godo-Pla, P. Emiliano, F. Valero, M. Poch, G. Sin, H. Monclús, *Process Saf. Environ. Prot.* **2019**, *125*, 317.
[93] Y. Zhang, Q. Ai, F. Xiao, R. Hao, T. Lu, *Int. J. Electr. Power Energy Syst.* **2020**, *114*, 105388.
[94] F. Harrou, A. Dairi, F. Kadri, Y. Sun, *Chaos, Solitons Fractals* **2020**, *139*, 110247.
[95] Y. Lee, B. Ha, S. Hwangbo, *Renewable Energy* **2022**, *200*, 69.
[96] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, Q. Wu, *Environ. Model. Softw.* **2019**, *114*, 29.
[97] M. Mowbray, M. Vallerio, C. Perez-Galvan, D. Zhang, A. D. R. Chanona, F. J. Navarro-Brull, *Reaction Chem. Eng.* **2022**, *7*, 1471.
[98] S. Nie, G. Cai, J. He, S. Wang, R. Bai, X. Chen, W. Wang, Z. Zhou, *Fuel* **2023**, *344*, 128047.
[99] B. Wealer, S. Bauer, C. Hirschhausen, C. Kemfert, L. Goke, *Renew. Sustain. Energy Rev.* **2021**, *143*, 110836.

How to cite this article: E. Pahija, S. Hwangbo, T. Saulnier-Bellemare, G. S. Patience, *Can. J. Chem. Eng.* **2024**, 1. https://doi.org/10.1002/cjce.25374

1939019x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cjce.25374 by Ecole Polytech De Montreal, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License