An optimal experimental design framework for fast kinetic model identification based on artificial neural networks

Enrico Sangoi [a], Marco Quaglio [a], Fabrizio Bezzo [b], Federico Galvanin [a],[*]

[a] *Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom*
[b] *CAPE-Lab (Computer-Aided Process Engineering Laboratory), Department of Industrial Engineering, University of Padova, 35131 Padova PD, Italy*

ARTICLE INFO

ABSTRACT

The development of mathematical models to describe reaction kinetics is crucial in process design, control, and optimisation. However, distinguishing between different candidate kinetic models presents a non-trivial challenge. Recent works on this topic introduced an approach that employs artificial neural networks (ANNs) to identify kinetic models. In this paper, the ANNs-based model identification approach is expanded by introducing an optimal experimental design procedure. The performance of the method is evaluated through a case study related to the identification of kinetics in a batch reaction system, where different combinations of experimental design variables and noise level on the measurements are compared to assess their impact on kinetic model identification. The proposed experimental design methodology effectively reduces the number of required experiments while enhancing the artificial neural network's ability to accurately identify the appropriate set of equations defining the kinetic model structure.

1. Introduction

In process systems engineering (PSE) the mathematical modelling plays a key role in understanding and characterising systems behaviour (Klatt and Marquardt, 2009). In particular when considering the chemical industry, the reactor is the heart of the process, where raw materials are converted into products. Accurately describing the phenomena taking place in reacting systems through mathematical models is essential for evaluating the advancement of chemical reactions within a reactor. This aspect holds significant importance in reactor design, control, and optimisation. The general description of the reaction unit requires several models, from the reactor model (e.g., CSTR, PFR, fluid-dynamics simulation) to the model of the systems dynamics (kinetic models). The identification of a kinetic model involves two aspects: i) the definition of the model structure, i.e. the mathematical formulation of reaction rate equations, and ii) an accurate estimation of the model parameters.

To identify a kinetic model, systematic model building approaches have been proposed in the literature (Asprey and Macchietto, 2000). However, the modelling process might pose several challenges to the modeller, both in terms of model structure identification and precise parameter estimation. For example, it might not be clear what reaction

steps are involved in the dynamics of the system (*model structure identification*) or the model could be affected by parameter identifiability issues (Asprey and Macchietto, 2000) that do not allow a precise estimation of the parameters (*parameter estimation*). On the other hand, instead of looking for a physics-based description of the systems dynamics, the modeller can tackle the problem using a data-driven approach, which is one of the foundations of machine learning (ML) technologies.

In the past years there has been an increasing interest in the application of ML techniques in many fields of science and engineering. Among several ML methods, especially artificial neural networks (ANNs) have been reported in a plethora of applications due to different key factors. Firstly, ANNs exhibit remarkable flexibility in approximating nonlinear continuous functions (Hornik et al., 1989). Secondly, there have been significant advancements in the development of efficient algorithms for training ANNs (Géron, 2019). Lastly, the cost of computational power has steadily decreased (Russell and Norvig, 2021), making it more accessible and contributing to the success of ANNs.

With respect to reaction kinetics there have been in past years different applications of ANNs. Neural networks can be used either as regressor or classifiers, but it is important to underline that most of the previous applications of ANNs to kinetic modelling were for regression (Kayala and Baldi, 2012; Amato et al., 2012; Wei et al., 2016;

---

* Corresponding author.
  E-mail address: f.galvanin@ucl.ac.uk (F. Galvanin).

Acronyms

ANN   Artificial neural network
DE   Differential evolution
DoE   Design of experiments
MBDoE   Model-based design of experiments

Chakkingal et al., 2021). Kayala and Baldi (2012) proposed a machine learning approach to determine the reaction mechanism given the reactant and products by using ANNs trained on different reaction types. Amato et al. (2012) used an ANN regressor as a surrogate model of the system and then exploited it for designing informative kinetic experiments, in order to avoid the multiple solution of ODEs systems. Wei et al. (2016) explored the use of ANNs for predicting reaction products given a set of reactants. Chakkingal et al. (2021) used ANNs for interpreting microkinetic data, demonstrating the approach on Fischer-Tropsch synthesis case study.

A recent work by Quaglio et al. (2020b) presented a model selection approach where an artificial neural network is trained for recognising kinetic model structures given the available experimental data, therefore not using the ANN for regression but instead tackling the model selection as a classification problem. The method does not necessitate fitting kinetic parameters and is particularly suitable in scenarios where a large number of potential kinetic mechanisms is involved. The approach presented in Quaglio et al. (2020b) was characterised by a fixed design of the experiments, though further study showed a relevant impact of the experimental design on the ANN ability in correctly classifying the model structures. This led to the idea of combining the ANN-based kinetic model recognition method with an optimal design of experiments procedure, firstly introduced in Sangoi et al. (2022). The methodology proposed for optimal design of experiments requires an optimisation, such as in standard model-based design of experiments (MBDoE) procedures (Galvanin et al., 2016), but differs from MBDoE approaches on how the optimisation problem is defined and solved numerically. Sangoi et al. (2022) presented preliminary results on the application of optimal DoE to enhance the ANN accuracy in classification.

This paper is an extension of the concept introduced in Sangoi et al. (2022), to provide a formal analysis and discussion of all the steps involved in the design of experiments optimisation methodology and new results including a) the sensitivity analysis of ANN performance to different sets of experimental conditions; and b) comparison of optimal experimental design solutions under different scenarios of experimental noise, degrees of freedom on the experimental variables and their impact on model identification.

This paper is aimed at describing the framework proposed to optimise the design of experiments for kinetic model identification and show extensively the results obtained on the case study considered. The manuscript is structured as follows. The proposed approach to design the experiments coupled to ANNs-based kinetic model recognition is detailed in Section 2. The case study on which the framework has been tested is described in Section 3. Results are presented and discussed in Section 4.

2. Proposed optimal experimental design approach

This section presents the proposed procedure for optimal design of experiments aimed at improving the performance of the ANN-based method for kinetic model recognition. Moreover, a comparison with standard model building procedures is also given in this section.

The section is structured as follows. Firstly, an overview of ANN classifiers is provided in subSection 2.1. Then subSection 2.2 follows by presenting the proposed procedure, with a special focus on the steps

involved in the optimisation of the experimental conditions to provide the reader with a comprehensive description of the experimental design procedure.

2.1. Artificial neural network classifier

Before starting with the description of the optimal DoE procedure, a brief overview of ANN classifiers is provided to the reader (Walczak and Cerpa, 2003) to have a clear understanding of the methods employed, summarising the most relevant information reported in Quaglio et al. (2020b). The key element in an Artificial Neural Network is the so-called *perceptron* (Rosenblatt, 1958), which is a function that transforms a $N_n \times 1$ input array $\mathbf{n}$ of real numbers into a scalar output p. The single layer perceptron model is mathematically described by

$$p = \psi\left(\mathbf{w}^T \mathbf{n} + b\right) \tag{1}$$

where $\mathbf{w}$ is an $N_n \times 1$ array of parameters, b is a bias parameter (scalar), and ψ represents the activation function of the neuron.

In general, an ANN is composed of many perceptrons that can be arranged and connected in several ways. Fig. 1 depicts the structure of a fully connected feedforward ANN, where the perceptrons are arranged in different layers and every neuron in the ith layer is connected with all the neurons in the $(i - 1)$th and $(i + 1)$th layers. Let consider a two-layers ANN as shown in Fig. 1 and be N_h and N_o the number of neurons respectively in the hidden and output layer. The ANN model is then defined as

$$\mathbf{p} = \psi_o\left[\mathbf{W}_o^T \psi_h\left(\mathbf{W}_h^T \mathbf{n} + \mathbf{b}_h\right) + \mathbf{b}_o\right] \tag{2}$$

In (2), $\mathbf{p}$ is the $N_m \times 1$ array of output values, $\mathbf{W}_h$ [$N_n \times N_h$] and $\mathbf{W}_o$ [$N_h \times N_m$] are matrices of parameters (weights) associated with the hidden and output layer respectively, $\mathbf{b}_h$ [$N_h \times 1$] and $\mathbf{b}_o$ [$N_m \times 1$] are the vectors with the bias parameters. The functions ψ_h and ψ_o are the activation functions chosen for the hidden layer and output layer.

When using an ANN for classification, the objective is to assign a label $l \in \{1, \dots, N_m\}$ to the input array among the set of N_m available categories. For this application a typical choice of activation function in the output layer ψ_o is the *softmax* function (Arbib, 2002). With this choice of activation function the output $\mathbf{p}$ satisfies the conditions $0 \leq p_k \leq 1 \ \forall k \in \{1, \dots, N_m\}$, $\sum_{k=1}^{N_m} p_k = 1$, therefore the values p_k can be interpreted as a probability of the class k to be associated with the input
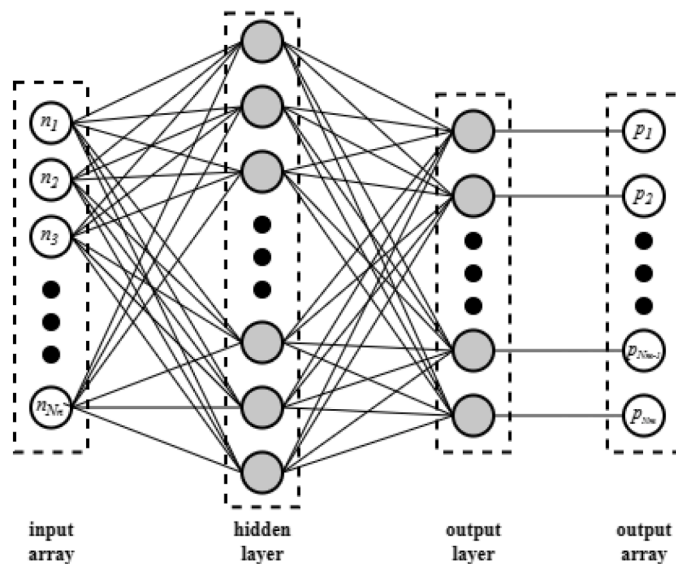

Fig. 1. Graphical representation of a two-layers feedforward artifiacial neural network. Gray-coloured circles represent the perceptrons, based on Quaglio et al. (2020b).

vector $\mathbf{n}$. The predicted label $\widehat{l} \in \{1, \ldots, N_m\}$ is then chosen as the class with the highest probability associated

$$\widehat{l} = \arg\max_{k} p_k \tag{3}$$

The construction of the ANN requires the definition of some hyperparameters, such as the number of layers and the number of perceptrons in each layer. The ANN is then trained on a training dataset, so that the weights are adjusted based on the information available in the data. Subsequently, new data (testing dataset) are fed to the trained ANN to test its ability in making predictions on unseen data, i.e. data not used for training.

2.2. Overview of the proposed procedure

In this project the ANN classifier is used as a tool for the selection of kinetic models, as formulated by Quaglio et al. (2020b) that referred to this application of ANNs as kinetic model recognition. In general, it is assumed that a set of N_m candidate kinetic models are proposed to characterise the dynamic behaviour of a reacting system of interest. In the experimental setup $\mathbf{u}$ is defined as the $N_u \times 1$ array of controlled input and $\mathbf{y}$ is the $N_y \times 1$ array of measured system outputs (state variables sampled over time). Time is indicated with the variable t. For every $l \in \{1, \ldots, N_m\}$ the respective model can be expressed in a general form as

$$\begin{aligned} \mathbf{f}_l(\dot{\mathbf{x}}_l, \ \mathbf{x}_l, \mathbf{u}, t, \boldsymbol{\theta}_l) &= 0 \\ \widehat{\mathbf{y}}_l &= \mathbf{h}_l(\mathbf{x}_l) \end{aligned} \tag{4}$$

where $\mathbf{x}_l$ is the array of $N_{x,l} \times 1$ system state variables, $\dot{\mathbf{x}}_l$ is the array of $N_{x,l} \times 1$ time derivatives of the state variables. In this formulation of the dynamic model, $\boldsymbol{\theta}_l$ is the $N_{\theta,l} \times 1$ array of model parameters, while $\widehat{\mathbf{y}}_l$ is the $N_{y,l} \times 1$ array of model predictions of the measured system states.

In the procedure proposed by Quaglio et al. (2020b) all the candidate models are used to build a large dataset of *in silico* simulated kinetic experiments. If N_m are the candidate models, N_p is the number of different sets of parameter values considered per model, N_y the number of observed states and N_{sp} the total number of samples from all the simulated experiments, then the dataset $\boldsymbol{\Psi}$ has dimension $\left[(N_m \cdot N_p) \times (N_y \cdot N_{sp} + 1) \right]$, since also the label is included as the last column, as depicted in Fig. 3. These labelled data are then used to train an ANN to associate the input experimental data (species concentration measurements) to the kinetic model structure that was used to generate them. Once the ANN is trained, validated, and tested, it can be applied for the kinetic model recognition on the physical reacting system: the experiments are conducted on the system under investigation at the same operating conditions of the simulated ones and the collected data are fed to the ANN. The softmax values obtained as output from the network are then taken as a probability associated to each model structure to best describe the reacting system.

Since in the parametric study by Quaglio et al. (2020b) a limited effect of the hyperparameters was observed on the ANN accuracy, in the proposed procedure the optimisation is conducted to identify the optimal experimental conditions while the ANN architecture is not optimised, i.e. the network hyperparameters are an input in the proposed procedure.

The framework firstly proposed in Sangoi et al. (2022) and described in detail in this paper is shown in Fig. 2. More specifically, the discussion is here extended to formally describe the formulation of the optimisation problem (Section 2.2.1), the definition of the hyperparameters for the optimiser (Section 2.2.2), and the equations used in the generation of the *in silico* data set (Section 2.2.3).

In the framework the ANN-based method is extended by including a procedure for the optimal design of experiments (step 2 in Fig. 2) before the *kinetic model recognition* (step 4) being applied on the real reacting system. In the scheme of Fig. 2 the main steps of the procedure are represented: step (1) is the definition of the inputs required to optimise the experiments in the context of ANN-based model recognition; step (2) the DoE optimization is performed which consists of three key tasks at every iteration of the algorithm; step (3) the results of the optimisation are obtained; and lastly in step (4) the kinetic model recognition approach introduced by Quaglio et al. (2020b) is used on the chemical system starting from the DoE optimisation outputs.

The inputs required in the procedure are the following: i) the definition of the experimental design space in terms of manipulated experimental design variables and their range of variability; ii) the library of candidate kinetic model structures formulated for describing the system; iii) a set of feasible parameters Θ_l for all the N_m candidate models; iv) a characterisation of the experimental noise; v) the set of hyperparameters describing the ANN structure (e.g., number of layers, number of nodes, activation functions).

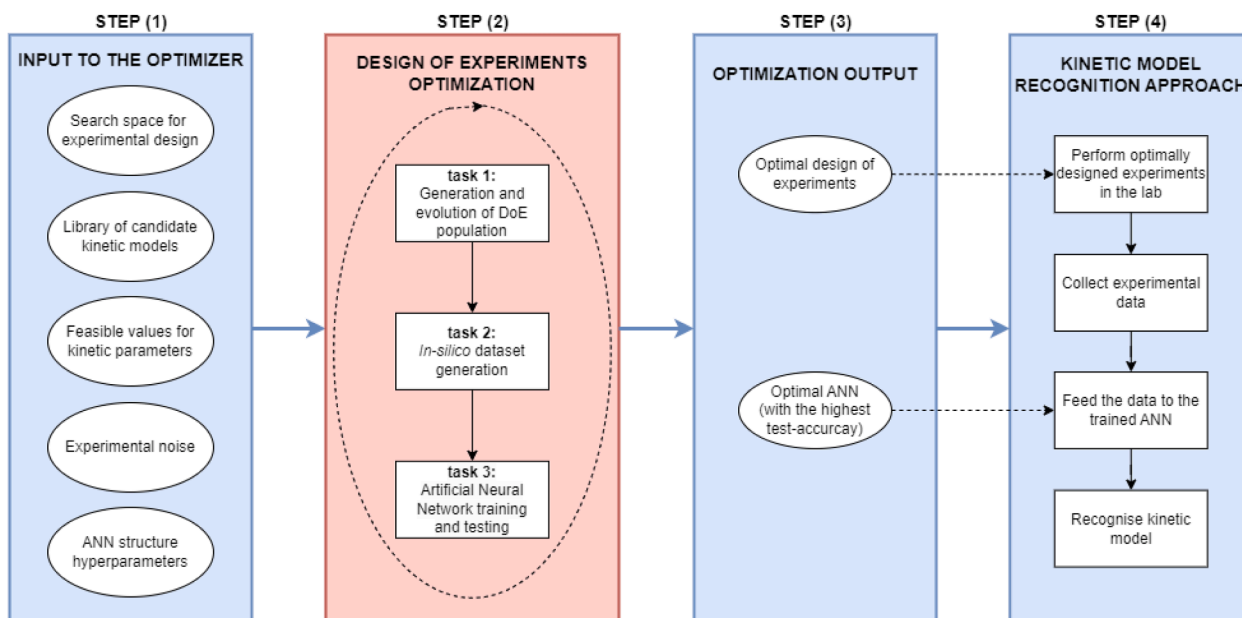Once all the inputs are defined, the DoE optimisation can start. The

Fig. 2. Proposed procedure for optimal design of experiments in the ANN-based kinetic model recognition framework.

experimental design optimisation step (Step 2) is an iterative process which consists of three key sequential tasks: (1) the generation and evolution of a population of possible DoE, (2) the experimental conditions are used to simulate kinetic experiments in order to create multiple datasets (one for each DoE in the population), (3) which are used to train and test ANNs (one per dataset). Therefore for every candidate DoE, a new ANN with the same structure is trained and tested, while fine-tuning of the network is not part of the proposed procedure.

Two outputs are obtained from the DoE optimisation, i.e. the optimal experimental conditions and the respective ANN, which is the one with the best performance in classifying new data. These two optimisation outputs are then used as inputs in the *kinetic model recognition* step, where the same experiments are conducted on the chemical system under assessment. The data obtained from the experiments are subsequently used as an input to the trained ANN classifier, which provides in output a probability associated to every candidate model structure.

2.2.1. Formulation of the optimisation problem

The metric chosen to optimise the experimental design is the ANN accuracy on the test set, evaluated as the percentage of models correctly classified, so that the optimisation problem is formulated as:

$$\min_{\boldsymbol{\varphi}} - \Gamma_{test}$$
$$\text{s.t.}$$
$$\boldsymbol{\varphi} \in \Phi \tag{5}$$

$$\Gamma_{test} = \frac{|\{i \in \{1, \ldots, N_{\Psi_{test}}\} \text{ s.t. } (\mathbf{n}_i, l_i) \in \Psi_{test} \wedge \hat{l}_i = l_i\}|}{|\{i \in \{1, \ldots, N_{\Psi_{test}}\} \text{ s.t. } (\mathbf{n}_i, l_i) \in \Psi_{test}\}|} \times 100\%$$

where the objective function Γ_{test} is the ANN accuracy in the predictions on the testing-set Ψ_{test}, $\boldsymbol{\varphi}$ is the experimental design vector and Φ is the experimental space defined by the experimental design variables and their range of variability.

The objective function is influenced by the intrinsic stochasticity in the ANN training process, therefore the optimisation turns into the problem of maximising a stochastic function that cannot be expressed mathematically as a function of the optimisation variables (i.e. the experimental conditions). Due to these features of the objective function in this context, it is chosen to apply a direct-search method to solve the problem rather than gradient-based algorithms. In particular, it is chosen to use the *differential evolution* (DE) algorithm by Storn and Price (1997), a population-based evolutionary algorithm. Along with robustness and easiness to make use of the solver, the most important advantage of using DE algorithm, with respect to the problem tackled, is that no information on the shape of the objective function is required, but only a measure of the cost function for every choice of the optimisation variables, i.e. the ANN's test accuracy when varying the DoE. In fact, DE was chosen because stochastic direct search approaches are particularly suitable for optimising nonlinear and non-differentiable objective functions. Moreover, DE algorithm satisfies the following requirements: i) it is able to handle nonlinear, non-differentiable and multimodal cost functions, ii) it can be parallelised to deal with computationally expensive problems, iii) it is easy to use and iv) it has good convergence properties, as reported by Storn and Price (1997). On the other hand, a disadvantage that must be considered is that the size of the optimisation variables vector, i.e. the number of experiments and the variables defining each experiment, has to be fixed and cannot be optimised with the experimental conditions. However, in order to also minimise the number of experiments required, a possibility is to repeat the optimisation starting with one experiment and then gradually increase the number of experiments. By doing this, the profile of the optimal solution as a function of the number of experiments is obtained and the point where the objective function first reaches its maximum can be chosen as the minimum number of experiments.

The following subsections provide a more detailed description of the three key tasks used to solve the optimisation problem (i.e. step 2 of

Fig. 2).

2.2.2. Task 1: generation and evolution of DoE population

The optimiser randomly generates a population of experimental designs, named *individuals*, which are then evolved at every iteration. The individuals are moved in the experimental space towards the region of expected optimum, aiming to make them all converge to the optimum within a user-specified tolerance. The general optimisation problem can be stated as a minimisation problem as

$$\min_{\boldsymbol{\varphi}} obj(\boldsymbol{\varphi}) \tag{6}$$

subject to some constraints

$$LB_j \leq \varphi_j \leq UB_j \ \forall j = 1, \ldots, N_{\varphi} \tag{7}$$

where $obj(\boldsymbol{\varphi})$ is the objective function, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_{N_{\varphi}})^T$ is the N_{φ}-dimensional array of optimisation variables and $(LB_j, UB_j) \ \forall j \in \{1, \ldots, N_{\varphi}\}$ are the lower and upper bounds defining the search space for the experimental design variables. In this study, $\boldsymbol{\varphi}$ is the vector of experimental conditions and $obj(\boldsymbol{\varphi})$ is the opposite of ANN test-accuracy (i.e., $-\Gamma_{test}$) as formulated in Eq. (5), so that minimising $obj(\boldsymbol{\varphi})$ corresponds to maximising Γ_{test}. The constraints in the form of Eq. (7) are given in terms of upper/lower bounds on operating conditions for the experimental design variables. A vector $\boldsymbol{\varphi}$ satisfying Eq. (7) will be called *feasible*.

To solve Eq. (6), the DE solver generates a number of individuals or target vectors, i.e. feasible vectors $\boldsymbol{\varphi}_{i,G} \ \forall i = 1, \ldots, NP$, where *NP* is the number of individuals in the population and *G* is an integer index indicating the generation number, which evolve in the next generation through the steps of *mutation, crossover* and *selection*. For the first iteration (i.e., $G = 1$) the target vectors are randomly selected within the feasible space. A detailed description of how the target vectors are evolved in the DE algorithm through mutation, crossover and selection is available in Appendix B.

An advantageous characteristic of differential evolution optimiser is its flexibility and easiness of use. A few parameters of the DE solver can be adjusted by the user, along with the strategy as previously mentioned, in order to adapt the algorithm to the problem needs. Among these parameters, the most relevant are the number of individuals in the population *NP*, the mutation factor *F* and the crossover probability *CR* (Appendix B). Moreover, the tolerance and the maximum number of iterations over which the entire population could be evolved are other important factors for tuning the optimiser. The number of individuals in the population is usually defined through the parameter *P,* which is a proportionality factor between N_{φ}, characteristic of the problem, and *NP*.

$$NP = P \cdot N_{\varphi} \tag{8}$$

According to Storn and Price (1997) a choice of *P* between 5 and 10 is reasonable to guarantee and adequate exploration of the space of variables.

The mutation constant *F* , a real number in the range [0,2], impacts on both the search radius and rate of convergence. Large values of *F* allow to increase the search radius, thus reducing the probability of stopping in local minima; however, the larger the mutation factor the slowest the convergence. The factor *F* can also be defined through an interval [*Fmin, Fmax*], so that the mutation factor randomly changes at every generation *G* of individuals, in a process called *dithering*. Dithering can help to significantly increase the rate of convergence. Storn and Price (1997) suggest that values of *F* smaller than 0.4 or larger than 1 are usually ineffective.

The crossover probability, or recombination constant, $CR \in [0,1]$, impacts on the mutated elements that can pass to the next generation of individuals. A large value of crossover probability usually leads to faster convergence while low values of *CR* will lead to a more stable behaviour

when approaching to the solution, therefore the choice of an appropriate value of *CR* is not trivial. A good starting choice can be *CR*=0.1 based on Storn and Price (1997), however larger values could be chosen to speed up the convergence.

Every time a DoE is generated or mutated, the respective value of the objective function is computed: the dataset is generated by simulating the corresponding experiments *in silico*, and split into training-validation-test sets to train the ANN, and evaluate the ANN accuracy (optimisation objective function) on the test dataset.

2.2.3. Task 2: in silico dataset generation

For every target DoE of the population, a labelled dataset $\mathbf{\Psi}$ is built as in Eq. (9) by simulating experimental measurements (see Fig. 3). The set of simulated measurements is the input array for the ANN, $\mathbf{n}_i$ $[N_{sp} \cdot N_y \times 1]$, and is constructed by integrating the equations of the kinetic model with model structure $l = l_i$.

$$\mathbf{\Psi} = \{(\mathbf{n}_i, \ l_i) \ \forall i = 1, ..., N_\psi \ \text{s.t.} \ l_i \in \{1, ..., N_m\}\} \tag{9}$$

To avoid the ANN predictions to be biased towards some of the kinetic models, it is important for the full dataset $\mathbf{\Psi}$ to be balanced and contain a comparable number of elements for each of the candidate N_m model structures. To reproduce real experimental conditions from the simulation, the elements $\mathbf{n}_i$ of the dataset are obtained by summing two contributions, i.e. the array of model predictions $\widehat{\mathbf{n}}_i$ $[N_{sp} \cdot N_y \times 1]$ and a random factor $\epsilon(\widehat{\mathbf{n}}_i)$ $[N_{sp} \cdot N_y \times 1]$ which emulates the experimental error on the measurements.

$$\mathbf{n}_i = \widehat{\mathbf{n}}_i + \epsilon(\widehat{\mathbf{n}}_i) \ \forall i \in \{1, ..., N_\psi\} \tag{10}$$

The prediction term $\widehat{\mathbf{n}}_i$ is defined as

$$\widehat{\mathbf{n}}_i = \left[\widehat{\mathbf{y}}_l(\boldsymbol{\varphi}_1, \boldsymbol{\theta}_l)^{\mathrm{T}}, \ ..., \ \widehat{\mathbf{y}}_l(\boldsymbol{\varphi}_{N_{sp}}, \boldsymbol{\theta}_l)^{\mathrm{T}} \right]_{\theta_l = U(\Theta_l), \ l = l_i} \ \forall i \in \{1, ..., N_\psi\} \tag{11}$$

The N_{sp} samples in Eq. (11) are computed from the l_i-th kinetic model structure, using the kinetic parameters $\boldsymbol{\theta}_l = U(\Theta_l)$, where the function

$U(\Theta_l)$ returns a randomly sampled parameter set from the feasible parameter domain Θ_l. The experimental error $\epsilon(\widehat{\mathbf{n}}_i)$ introduced in (10) is assumed to be normally distributed, therefore expressed as

$$\epsilon(\widehat{\mathbf{n}}_i) \sim \mathcal{N}(0, \ \mathbf{\Sigma}(\widehat{\mathbf{n}}_i)) \ \forall i \in \{1, ..., N_\psi\} \tag{12}$$

In (12) $\mathcal{N}(0, \mathbf{\Sigma})$ represents a multivariate normal distribution with zero mean and covariance matrix $\mathbf{\Sigma}$ $[N_{sp} \cdot N_y \times N_{sp} \cdot N_y]$. The covariance matrix is chosen as a diagonal matrix, the elements of which are defined in the general form as

$$\sigma_{jk} = \begin{cases} \sigma_r^2 \cdot \dfrac{1}{100} \cdot \widehat{n}_{ij} + \sigma_c^2 & \text{if} \ j = k \\ 0 & \text{if} \ j \neq k \end{cases} \ \forall j, k \tag{13}$$

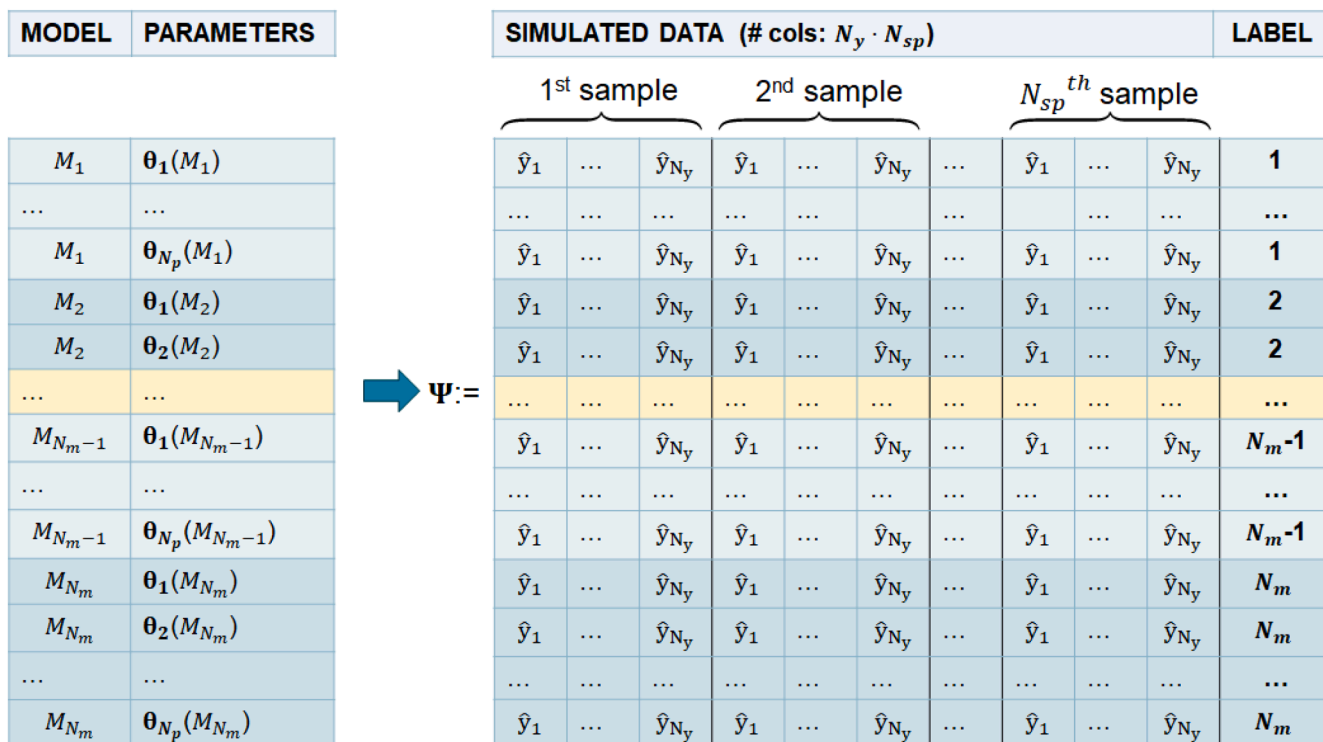Where $\widehat{n}_{ij}$ is the j-th element of the array $\widehat{\mathbf{n}}_i$, σ_r is a scalar representing the relative variance and σ_c is a scalar indicating the contribution of the constant variance to the overall noise on the simulated measurements.

2.2.4. Task 3: ANN training and testing

After defining the experimental design and building the dataset though simulated experiments, the procedure continues with the training and testing of the ANN. The ANN hyperparameters (e.g. number of layers, number of nodes, activation function) used to build the network are user-defined inputs, while the training process aims at identifying the ANN parameters, i.e. the weights associated to the connections between the nodes ($\mathbf{W}_h$ and $\mathbf{W}_o$ in Eq. (2)). It must be underlined that the selection of the network hyperparameters is important to avoid overfitting or underfitting issues; however, in this paper the number of neurons in the hidden layer is fixed based on the results obtained in the previous work by Quaglio et al. (2020b).

The full dataset $\mathbf{\Psi}$ (Fig. 3, Eq. (9)) is split in three sub-datasets based on a 60-20-20 proportion (Géron, 2019), i.e. a training set $\mathbf{\Psi}_{\text{training}}$, a validation set $\mathbf{\Psi}_{\text{validation}}$ and a testing set $\mathbf{\Psi}_{\text{testing}}$, as depicted in the scheme of Fig. 4.

The two datasets $\mathbf{\Psi}_{\text{training}}$ and $\mathbf{\Psi}_{\text{validation}}$ are used in the learning stage

MODEL	PARAMETERS	SIMULATED DATA (# cols: $N_y \cdot N_{sp}$)									LABEL
		1st sample		2nd sample			$N_{sp}{}^{th}$ sample				
M_1	$\boldsymbol{\theta}_1(M_1)$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	1
...	...	...	...	...	...	...	...	...	...	...	...	...
M_1	$\boldsymbol{\theta}_{N_p}(M_1)$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	1
M_2	$\boldsymbol{\theta}_1(M_2)$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	2
M_2	$\boldsymbol{\theta}_2(M_2)$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	2
...	...	...	...	...	...	...	...	...	...	...	...	...
M_{N_m-1}	$\boldsymbol{\theta}_1(M_{N_m-1})$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	N_m-1
...	...	...	...	...	...	...	...	...	...	...	...	...
M_{N_m-1}	$\boldsymbol{\theta}_{N_p}(M_{N_m-1})$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	N_m-1
M_{N_m}	$\boldsymbol{\theta}_1(M_{N_m})$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	N_m
M_{N_m}	$\boldsymbol{\theta}_2(M_{N_m})$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	N_m
...	...	...	...	...	...	...	...	...	...	...	...	...
M_{N_m}	$\boldsymbol{\theta}_{N_p}(M_{N_m})$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	...	$\widehat{y}_1$	...	$\widehat{y}_{N_y}$	N_m

$\mathbf{\Psi} :=$

Fig. 3. Structure of the full dataset built from simulated measurements. Every instance (i.e. row) of the dataset includes all the data obtained from a reacting system, i.e. a specific model structure and values of the model parameters.

Fig. 4. Dataset organization on a 60–20–20 basis for ANN training, validation and testing.

for the identification of the network parameters. The profiles of accuracy vs. epochs and loss vs. epochs, obtained with the training and validation sets, are monitored to ensure that the ANN is not overfitting the training data. Once the network parameters are identified, the testing dataset is employed to test the ANN performance on new data, *unseen* during the network training process.

The accuracy of the ANN in predicting the right label for the testing data is defined as in Eq. (5), which expresses the objective function for the DoE optimisation. This procedure of splitting the dataset and then training and testing an ANN is repeated for every dataset generated in the previous step, i.e. for every target vector in the experimental design space form which a dataset is built by simulating the kinetic experiments.

The tasks 1 to 3 here described are iterated until the solver converges to a solution, i.e. until the optimal experimental conditions that lead to the maximum ANN test-accuracy are found. Once the optimiser has converged, the modelling procedure continues (steps 3 and 4 in Fig. 2) by conducting the experiments identified by the optimiser on the real reacting system. The data collected are then fed to the ANN which provides in output a probability associated to each one of the possible kinetic model structures to characterise the system under assessment.

3. Case study

The methodology here proposed for optimal design of experiments has been tested on a case study, in particular, the chemical system introduced in the work by Quaglio et al. (2020b) is used.

In this system, three species, generically denoted as A, B, and C, are reacting in liquid phase within a perfectly mixed isothermal batch reactor. Direct reactions are assumed, such that

$$A \xrightarrow{r_1} B; \ A \xrightarrow{r_2} C; \ B \xrightarrow{r_3} C \tag{14}$$

where r_j (mol m^{-3} s^{-1}) is the reaction rate of the j-th reaction. The material balances that describe the dynamic evolution of the concentration C_i (mol m^{-3}) of the three species are expressed as

$$\frac{dC_i}{dt} = \sum_{j=1}^{N_r} \nu_{ij} r_j \ \forall i = A, B, C \tag{15}$$

In (15) the symbol ν_{ij} (-) indicates the stoichiometric coefficient of species i in reaction j. The functional form of the kinetic model is uncertain, and $N_m = 8$ candidate model structures are formulated as summarised in Table 1. Models with label $l = 1, …, 4$ are describing a system where the reactions occur in series, while $l = 5, …, 8$ are

associated to parallel reactions mechanism. The kinetic rates k_j are expressed as Arrhenius-type constants as in (16), where A_j and $E_{a,j}$ are the model parameters, respectively pre-exponential factor and activation energy.

$$k_j = A_j \exp\left(-\frac{E_{a,j}}{RT}\right) \forall j = 1, …, N_r \tag{16}$$

It is important to underline that all the kinetic models, formulated as in Table 1 are structurally non-identifiable: on the one hand, the kinetic constant k_2 is multiplied by zero in the models with series mechanism, $l \in \{1, …4\}$, thus the value of parameters A_2 and $E_{a,2}$ cannot be uniquely determined; on the other hand, the kinetic factor k_3 is multiplied by zero when parallel mechanism of reaction is modelled, i.e. for label $l \in \{5, …,8\}$, so that the value of parameters A_3 and $E_{a,3}$ cannot be uniquely estimated. Therefore, sequential model discrimination methods (Schwaab et al., 2008) would need a reformulation of the models to be applied, since non-identifiability issues translate into large variances in the parameter estimates and as a consequence also on the model predictions. Conversely, the structural non-identifiability of the candidate models is not an issue for the ANN-based model recognition since the fitting of kinetic parameters is not required. The proposed approach focuses on the identification of the kinetic model structure (i.e. set of equations), not on the estimation of the parameters. If a non-identifiable model structure is selected as an output of the ANN method, a reparameterisation (Quaglio et al., 2020a) of the model to satisfy the parameter identifiability requirements is needed.

3.1. Definition of experimental design space

The methodology presented in Section 2 requires as input the definition of the experimental design space. The manipulated variables in the batch reactor experiments here considered with the respective range of variability are: i) temperature T in the reactor [520 – 720] K, ii) sampling time t [50 – 350] s, and iii) initial concentration of reactant A, C_{A0} [0 – 250] mol m^{-3}. The initial concentration of B and C is 0 mol m^{-3}, i.e. A is the only species initially present in the reaction system.

In practical applications, when defining the experimental design variables and their range, the user has to make use of prior knowledge on the system to be modelled, imposing physically meaningful constraints. In this case study, since experiments are conducted *in silico*, no information on a real system is assumed to be available, therefore the experimental space has been arbitrarily chosen.

Table 1

Power-law rate expression for the candidate kinetic models. Models with label $l = 1,…,4$ consider a series mechanism, while models $l = 5, …,8$ consider a parallel mechanism.

Label:	Series				Parallel			
	1	2	3	4	5	6	7	8
r_1	$k_1 \cdot C_A$	$k_1 \cdot C_A$	$k_1 \cdot C_A^2$	$k_1 \cdot C_A^2$	$k_1 \cdot C_A$	$k_1 \cdot C_A$	$k_1 \cdot C_A^2$	$k_1 \cdot C_A^2$
r_2	$k_2 \cdot 0$	$k_2 \cdot 0$	$k_2 \cdot 0$	$k_2 \cdot 0$	$k_2 \cdot C_A$	$k_2 \cdot C_A^2$	$k_2 \cdot C_A$	$k_2 \cdot C_A^2$
r_3	$k_3 \cdot C_B$	$k_3 \cdot C_B^2$	$k_3 \cdot C_B$	$k_3 \cdot C_B^2$	$k_3 \cdot 0$	$k_3 \cdot 0$	$k_3 \cdot 0$	$k_3 \cdot 0$

3.2. Experimental noise

A noise factor is added to the simulated measurements to mimic real experimental conditions as presented in Eq. (10). For this case study, three different scenarios of experimental error are considered, respectively σ=1.00 mol m^{-3}, σ=4.00 mol m^{-3}, and σ=10.00 mol m^{-3} which will be referred to as *low, medium,* and *high* noise level scenarios. The effect of the noise level on the concentration profile, obtained from simulated experiments, can be observed in Fig. 5. Graphs are obtained from the simulation of a system modelled through a label-1 kinetic model structure, where first-order reactions occur in series. It can be observed how the noise level impacts on the quality of measurements, thus on the data that are fed to the ANN.

3.3. Feasible parameter values and dataset generation

The dataset generation step requires to simulate a high number of reacting systems, in order to consider parametric uncertainty in the model identification framework. In this case study, 125 different choices of feasible parameter values are used for each candidate kinetic model. Therefore, the cardinality of the dataset is equal to 1000 ($N_M = 8$, $N_p = 125$), where each instance (i.e. a row in the dataset, see Fig. 3) corresponds to a simulated reacting system, i.e. a kinetic model and a set of parameter values for that model, and contains all the simulated measurements sampled for that reacting system. The 1000 instances in the dataset are randomly ordered and then split in training-validation-testing sets on a 60-20-20 base. Therefore, the training set contains 600 elements, while validation and testing sets are built with 200 simulated chemical systems each.

The feasibility condition for the kinetic model parameters is expressed as

$$\Theta_l = \{\boldsymbol{\theta}_l \in \mathbb{R}^{N_{\theta,l}} \quad \text{s.t.} \quad \begin{array}{lll} 100 \le A_j \le 200 & \forall j = 1,\ldots,N_r & \wedge \\ 45,000 \le E_{a,j} \le 90,000 & \forall j = 1,\ldots,N_r & \wedge \\ 0.05 \le \widehat{X}_{A,l} \le 0.95 & \forall \varphi \in \overline{\Phi} & \wedge \\ 0.05 \le \widehat{S}_{B,l} & \forall \varphi \in \overline{\Phi} & \wedge \\ 0.05 \le \widehat{S}_{C,l} & \forall \varphi \in \overline{\Phi} & \} \end{array}$$

$$\forall l = 1,\ldots,N_m \quad (17)$$

where $\boldsymbol{\theta}_l \in \Theta_l$ is a feasible choice of parameter values for kinetic model l, for all $l = 1,\ldots,N_m$.

The definition of feasibility conditions for model parameter values should be based on some prior knowledge on the system behaviour at the reference experimental conditions defined by $\overline{\Phi}$ (preliminary experiments). It is here assumed that the model parameters must satisfy the following feasibility constraints: *i)* pre-exponential factors and activation energies lie between lower and upper physical bounds; *ii)* the conversion of species A predicted by the model l, $\widehat{X}_{A,l}$ (mol/mol), is between 0.05 and 0.95 at the experimental conditions defined by $\overline{\Phi}$; *iii)* the predicted selectivity of product B, $\widehat{S}_{B,l}$ (mol/mol), and the predicted selectivity towards species C, $\widehat{S}_{C,l}$ (mol/mol), are above 0.05 at all the reference conditions $\varphi \in \overline{\Phi}$. The definition used for conversion of reactant A, X_A, and selectivity of product i, S_i, are
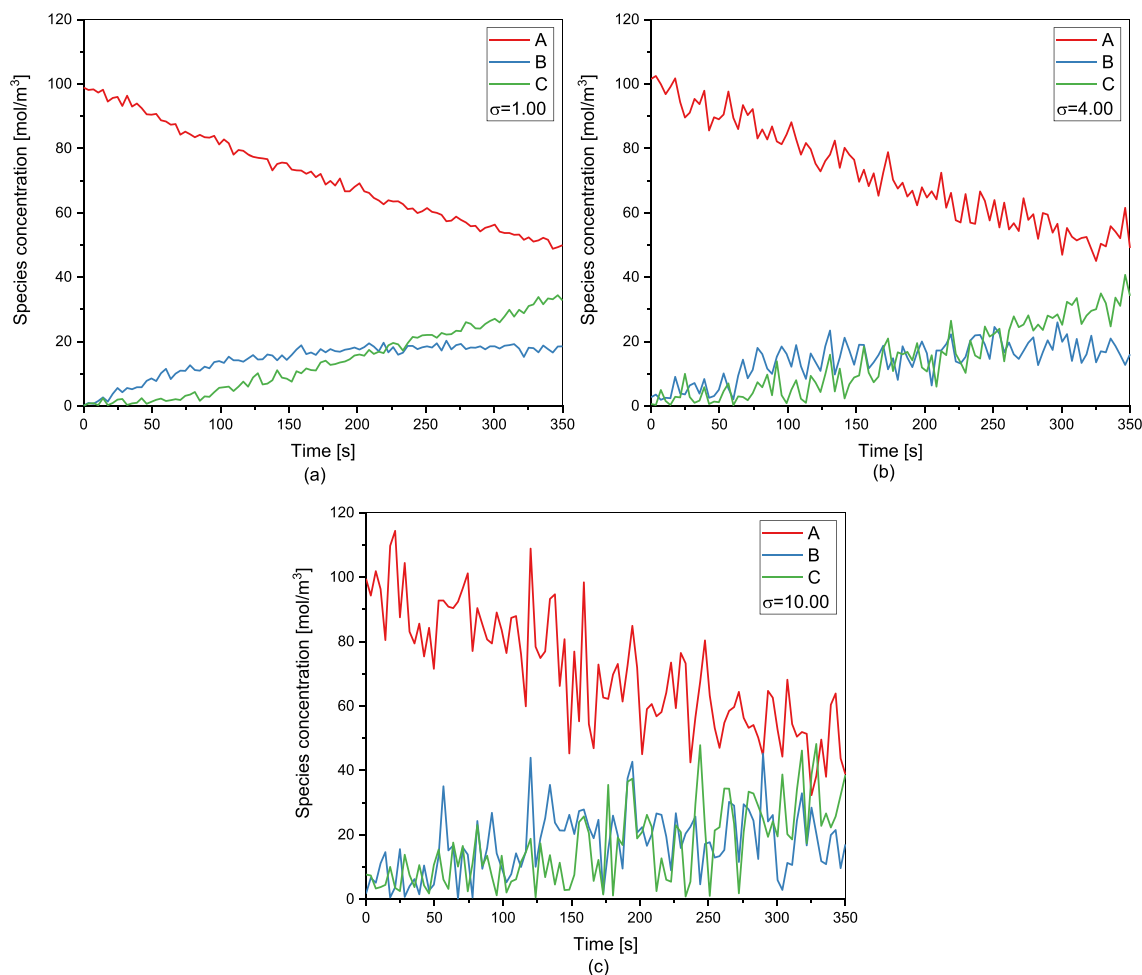

Fig. 5. Concentration profiles at different noise levels obtained with $C_{A,0} = 100$ mol/m^3 and $T = 620$ K from a kinetic model of label 1. The standard deviation values considered in the experimental error model are (a) *low* σ=1.00 mol/m^3, (b) *medium* σ=4.00 mol/m^3 and (c) *high* σ=10.00 mol/m^3.

$$X_A = \frac{C_{A,0} - C_A}{C_{A,0}}; \; S_i = \frac{C_i - C_{i,0}}{C_{A,0} - C_A} \tag{18}$$

In this case study, $\overline{\Phi}$ defines two simulated experiments, each one characterised by two sampling times. The reference experimental conditions for the first experiment are: *i)* temperature T equal to 573 K; *ii)* the initial concentration of reactant $C_{A,0}$ is 100 mol m^{-3}; *iii)* samples taken after 100 s and 300 s. The reference experimental conditions defining the second experiment are: *i)* temperature T equal to 673 K *ii)* initial concentration of reactant $C_{A,0}$ equal to 100 mol m^{-3}; *iii)* sampling times equal to 100 s and 300 s.

Conducting the experiments outside of the conditions imposed by $\overline{\Phi}$ could lead to limiting cases, such as complete or almost null reactant conversion. However, in this case a wider range for the experimental variables (see Section 3.2) has been assumed to extensively explore the experimental space while optimising the DoE.

3.4. Artificial neural network hyperparameters

For this study a two-layers feedforward artificial neural network is used, structured with one hidden layer and one output layer, based on the study presented in Quaglio et al. (2020b). The hyperparameters characterising the ANN architecture and how the training process is carried out are summarised in Table 2. Following Quaglio et al. (2020b), the hidden layer is built with N_h=100 neurons. For the activation function in the hidden layer nodes, rectified linear unit (ReLU) function is used. The activation function in the output layer is the softmax function. The size of the input array N_n depends on the number of experiments to be conducted and on the number of samples in each experiment: for every sample, 3 concentration values are measured (C_A, C_B, C_C), and accordingly the input array size is $N_n = 3 \cdot N$, where N is the total number of samples. The number of nodes in the output layer, N_o, is the same as the number of candidate kinetic model structures for the classification problem, therefore N_o=8.

3.5. Optimisation settings

The optimal DoE methodology has been implemented in Python. The optimisation algorithm *differential evolution* (Storn and Price, 1997) is available in the python library "*scipy.optimize*" (Virtanen et al., 2020), which requires the specification of certain parameters as input. The values used in this case study are reported in Table 3 along with the description of the parameters.

4. Results and discussion

A sensitivity analysis is conducted to study the effect of the design variables on the ANN performance, before implementing and testing the optimal DoE procedure described in Section 2 of this paper. The results of the sensitivity analysis are available for the reader in Appendix A.

Table 2
Artificial neural network hyperparameters defining the structure and training process.

Variable	Value/selection
Number of nodes in the input layer	Depends on the input array size
Number of nodes in the hidden layer	100
Number of nodes in the output layer	8
Activation function in the hidden layer	Rectified linear unit (*ReLU*)
Activation function in the output layer	Softmax
Optimiser	Adaptive moment estimation (*Adam*)
Initialiser	Normal
Loss function	Categorical cross-entropy
Dropout size	0.1
Learning rate	0.1
Batch size	40
Epochs	100

Table 3
The optimization settings choosen for the differential evolution algorithm in this case study.

Parameter	Value/ selection	Description
strategy	'best1bin'	The strategy used by the optimiser for evolving the individuals to the next generation.
maxiter	30	The maximum number of iterations (generations).
popsize	15	The factor defining the population size. The number of individuals is "popsize" times the number of variables to be optimised.
tol	0.1	The tolerance defining the stopping criterion for the algorithm.
mutation	(0.5, 1)	The mutation factor employed in the mutation step of the algorithm. A random value is chosen from the range at every iteration.
recombination	0.7	The recombination factor used for the crossover.
workers	-1	The solver uses multiprocessing to exploit the maximum computational power.

From the analysis it emerged that the experimental design chosen had a strong impact on the ANN accuracy and the regions of maximum accuracy in the design space emerged clearly, supporting the idea of using the ANN accuracy as the objective function.

In this section, the optimally designed experiments are reported along with the respective ANN accuracy obtained and the major achievements are discussed for different cases depending on the set of experimental design variables. For each case also then effect of the noise level on the measurements is investigated.

4.1. Optimal design of experiments

The preliminary results presented in Sangoi et al. (2022) are extended by comparing optimal experimental design solutions under different scenarios of experimental noise and degrees of freedom on experimental variables. The DoE optimisation has been conducted for three cases characterised by different sets of experimental design variables:

- Case A: temperature and sampling time, i.e. $\boldsymbol{\varphi} = [\mathbf{T}, \mathbf{t}]$;
- Case B: temperature and initial concentration of reactant A, i.e. $\boldsymbol{\varphi} = [\mathbf{T}, \mathbf{C}_{A,0}]$;
- Case C: temperature, initial concentration of reactant A, and sampling time, i.e.
  $\boldsymbol{\varphi} = [\mathbf{T}, \mathbf{C}_{A,0}, \mathbf{t}]$.

For each case different noise levels are considered as discussed in Section 3.3, identified by standard deviations equal to σ=1.00 mol m^{-3}, σ=4.00 mol m^{-3}, and σ=10.00 mol m^{-3} which are referred to as *low, medium*, and *high* noise. Results are also discussed considering an increasing number of experiments to be simultaneously optimised, ranging from 1 to 6 points in the experimental design space.

4.1.1. Case A: temperature and sampling time

The first case considered for testing the proposed DoE optimisation approach is characterised by experiments defined by temperature and sampling time. The amount of reactant A, the only species present in the batch reactor at time zero, is fixed, assuming a value of 100 mol/m^3 for each simulated experiment. It is assumed that only one sample is collected per experiment at the sampling time defined by the optimisation.

The results of the DoE optimisation are shown in Fig. 6 for the low and high noise scenarios, with 1, 3 and 6 experiments. Given that the ANN employed for model recognition is a black-box model, the interpretation of the chosen experimental design is not straightforward; however, some insights can still be extracted from the distribution of the designed experiments in the search space. Looking at Fig. 6, it is
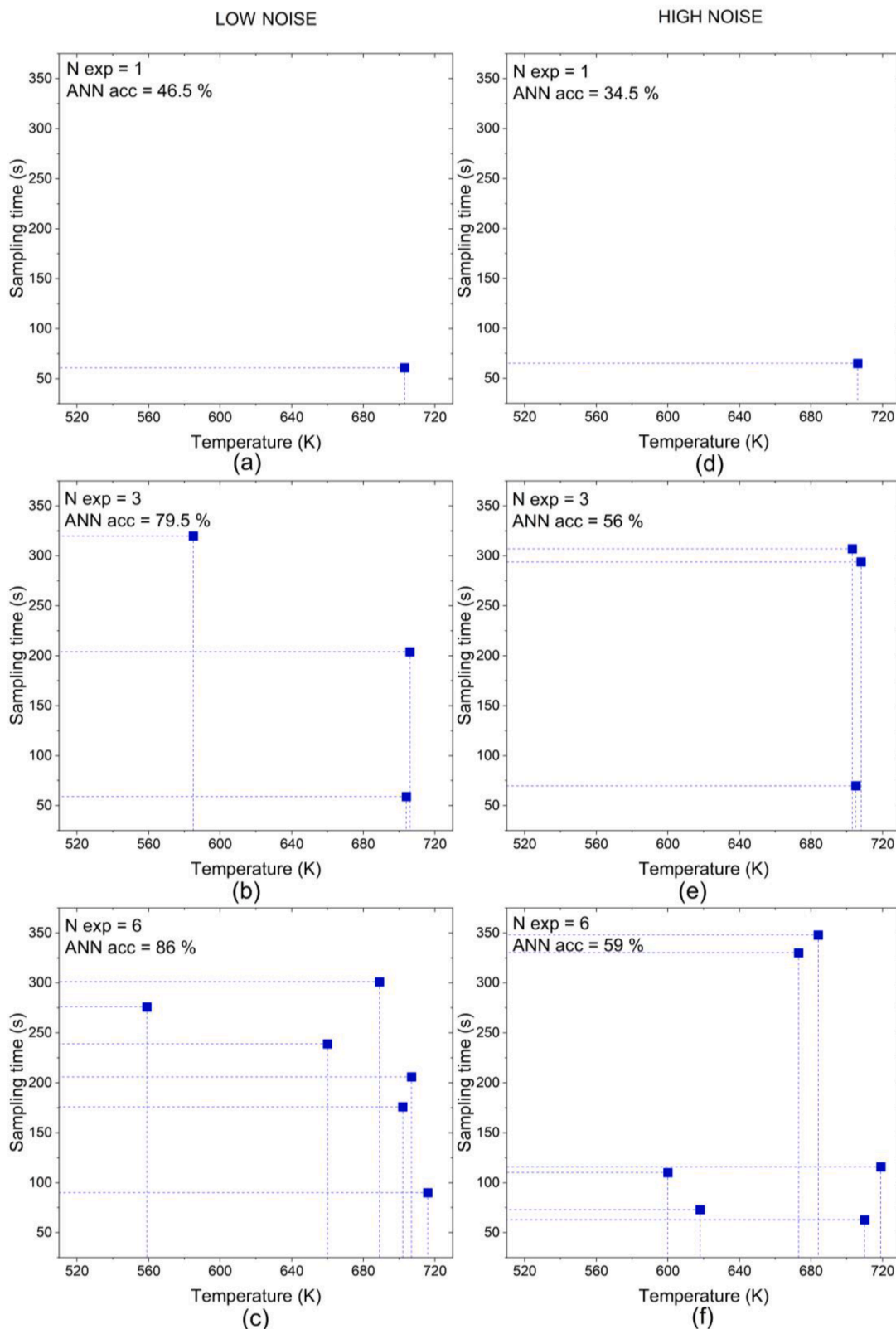
LOW NOISE

HIGH NOISE

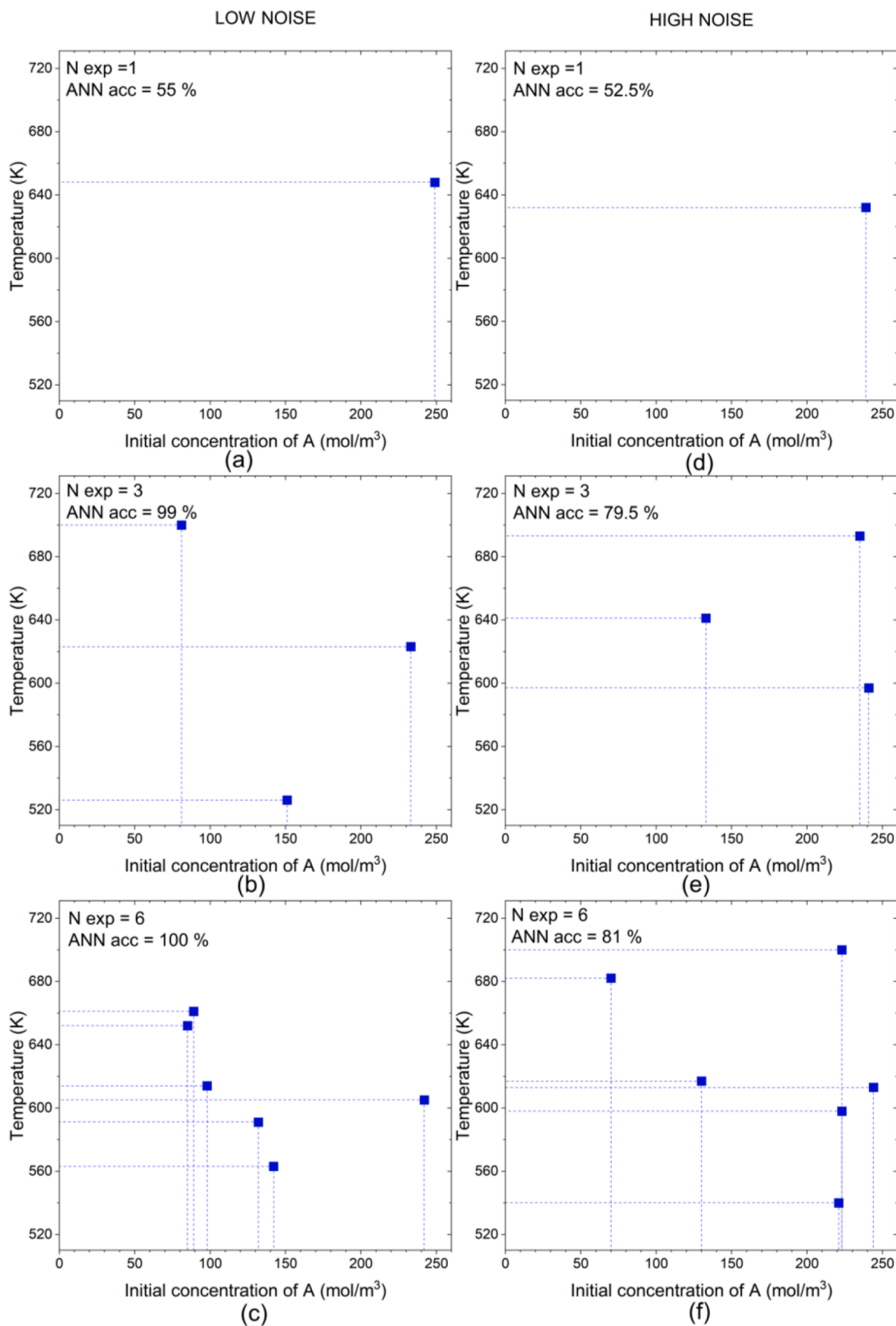**Fig. 6.** Optimally designed experiments with temperature and sampling time as experimental design variables. Results are reported for low (6.a-c) and high (6.d-f) experimental noise scenarios, and for the optimisation of 1, 3 and 6 experiments. The respective ANN test-accuracy is also reported.

observed that one experiment was always designed at a temperature around 705 K (high temperature) with the sample collected at about 75 s for all the 6 cases considered. This area of the design space corresponds to the maximum ANN-accuracy for a single experiment observed in the

preliminary studies (Appendix A, Figure A.1), therefore indicating a good performance of the methodology in identifying the optimal DoE conditions.

From the comparison of low and high noise levels with multiple

experiments (Fig. 6b, c, e and f), it is observed that higher temperature values are chosen in the case of high noise. In fact the minimum temperature value observed in the low noise case is 560 K, while in the high noise case the minimum reaction temperature value is 600 K. A possible explanation for this is that higher temperature leads to higher conversion, therefore higher concentration values for the reaction products. Since the measurement error is defined by a constant variance model, choosing high temperature values turn into a lower relative impact of

Fig. 7. Optimally designed experiments with temperature and initial concentration of species A as experimental design variables. Results are reported for low (7.a-c) and high (7.d-f) experimental noise scenarios, and for the optimisation of 1, 3 and 6 experiments. The respective ANN test-accuracy is also reported.

the noise with respect to the experimental data fed to the ANN.

In Fig. 6 the respective values of ANN accuracy (Eq. (5)) in identifying the correct model structures are reported for every scenario of noise and number of experiments. It is observed that the noise level, so the data quality, has a strong impact on the performance: the maximum accuracy achieved in the low noise case is 86% (Fig. 6.c), while in the high noise scenario it is 59% (Fig. 6.f), both obtained with 6

experiments. In all the noise scenarios a sharp increase in performance is observed from 1 to 3 optimal experiments, while the additional data from 3 to 6 experiments lead to a small increase in ANN accuracy (79.5% to 86% for low noise, 56% to 59% for high noise).

4.1.2. Case B: initial concentration of A and temperature

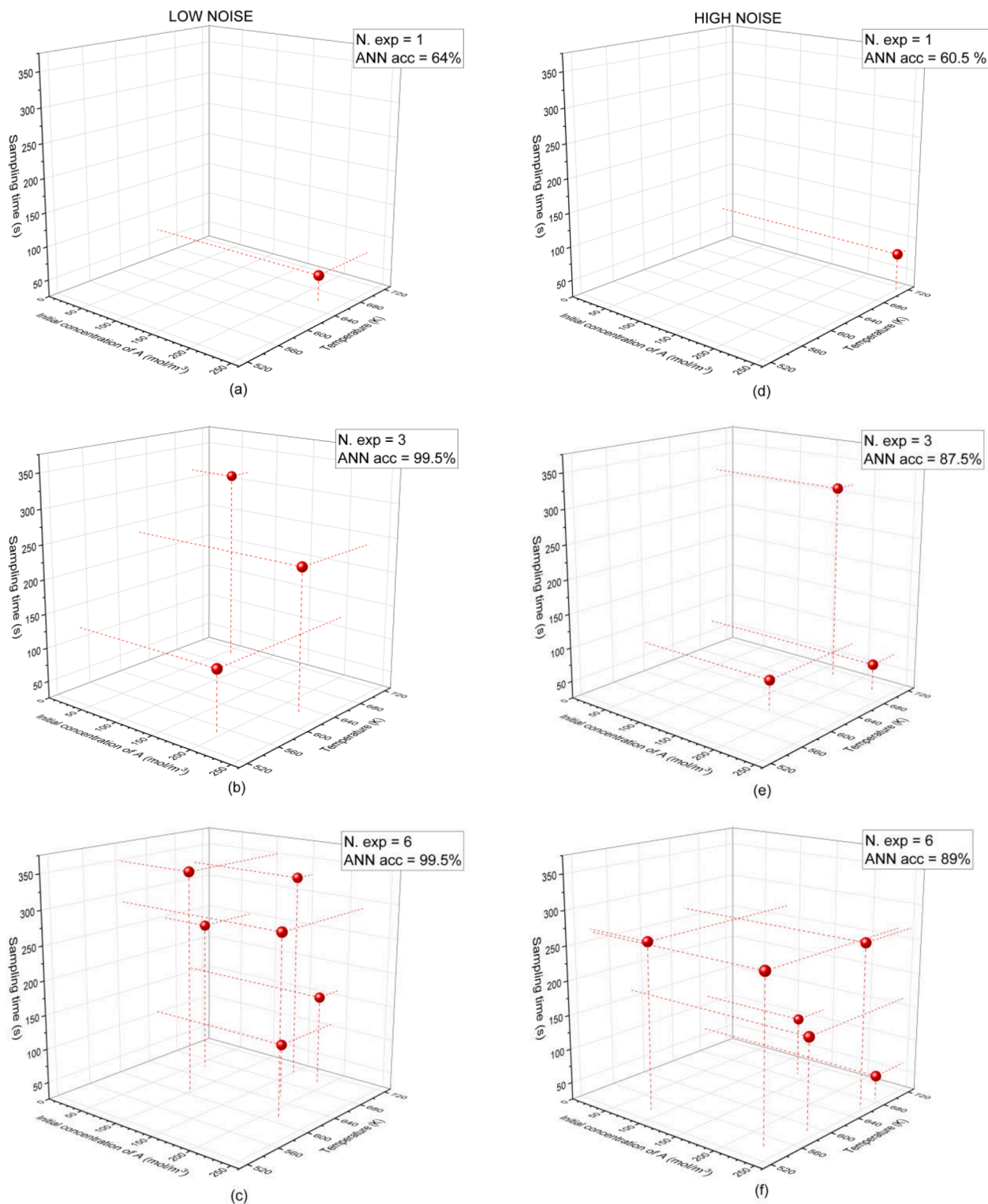The second scenario considered for testing the proposed DoE

Fig. 8. Optimally designed experiments with temperature, initial concentration of species A and sampling time as experimental design variables. Results are reported for low (8.a-c) and high (8.d-f) experimental noise scenarios, and for the optimisation of 1, 3 and 6 experiments. The respective ANN test-accuracy is also reported.

optimisation approach is characterised by experiments where the controlled variables are the initial concentration of reactant A and the temperature. For a fair comparison with case A, also in case B one sample per experiment is collected, and in this case the sampling time is fixed at 100 s. In this way the dimension of the dataset used for ANN training and testing is kept the same for a given number of experiments.

Fig. 7 shows the optimally design experiments in case B for the low and high noise scenario, and for 1, 3 and 6 experiments simultaneously optimised. Similarly to case A, also in this case the optimal experimental conditions identified for a single experiment are located in the region of maximum observed in the preliminary studies (Appendix A, Figure A.2), supporting the choice of the optimiser for this task.

When 3 experiments are simultaneously optimised (Fig. 7b and e) it is observed that design points are spread in the design space and higher temperature values are preferred when the noise level increases. Conversely, when 6 experiments are designed (Fig. 7c and f), some experiments are defined by very similar conditions. By analysing the ANN accuracy, in the low noise scenario an accuracy of 99% is achieved with only 3 experiments and the accuracy can be improved further to 100% if 3 additional experiments are designed.

In the high noise level scenario instead the ANN accuracy increases from 79.5% to 81% from 3 to 6 experiments. The fact that the objective function improvement is marginal from 3 to 6 experiments is indicative of a saturation in the amount of information that can be obtained on the kinetic models for the ANN. It is expected therefore that different choices of the experimental design can lead to the same objective function value when the number of experiments is increased further.

When comparing Case B and Case A results it can be concluded that, for a fixed number of experiments, controlling the initial concentration of A as an experimental design variable is more important than controlling the sampling time to obtain more informative data for the ANN.

4.1.3. Case C: initial concentration of A, temperature, and sampling time

The last scenario considered for testing the proposed DoE optimisation approach is characterised by experiments where all the experimental design variables can be manipulated. Also in this case one sample is collected per experiment to keep the consistency with case A and B, therefore an experiment is defined by its constant temperature, the initial concentration of species A and the sampling time that will be reported with the triplet (T, $C_{A,0}$, st).

Fig. 8 shows the optimally design experiments in case C for the low and high noise scenario, and for 1, 3 and 6 experiments simultaneously optimised. Comparing the optimisation of a single experiment, the designed experiment in the low noise case Fig. 8a is defined by moderate temperature, high reactant concentration and low sampling time($T = 652$ K, $C_{A,0} = 229$ mol/m^3, $st = 68$ s), while in the high noise scenario Fig. 8d the DoE is ($T = 714$ K, $C_{A,0} = 249$ mol/m^3, $st = 79$ s). It is observed that higher temperature and concentration of A are preferred when higher noise levels are affecting the measurements.

When N_{exp} is increased to 3 and 6, the designed experiments are widely distributed, showing a tendency to extensively explore the design space to gain more information about the kinetic system.

It must be noted that in the high noise case the initial concentration of A is generally chosen at higher values compared to the low noise case. It is also observed in Fig. 8f (high noise scenario, N_{exp}=6) that the maximum sampling times chosen is 265 s, while in Fig. 8c (low noise scenario, N_{exp}=6) the boundaries of the design space are reached (349 s).

In Fig. 8 the performance in terms of ANN accuracy is reported. When a single experiment is designed the accuracy that can be obtained is similar in the two noise scenarios, i.e. 64% (low noise scenario) and 60.5% (high noise scenario). This result is remarkable considering that the ANN is able to classify correctly 60% of the 200 model structures in the testing set based on the data collected from a single experiment with one sample. When N_{exp} is increased, the noise has a significant impact on the achievable accuracy. In the low noise scenario 99.5% accuracy is obtained with just 3 experiments, while in the high noise scenario the

accuracy increases to 87.5% with 3 experiments and then to 89% when 6 experiments are designed.

4.2. Maximum artificial neural network accuracy

This section wants to analyse the dependency of the ANN performance with respect to the number of designed experiments. Table 4 reports the value of objective function achieved with the optimally designed experiments for the three cases A, B and C and at the three noise levels considered. In each one of these experiments only one sample of the mixture was collected. Results are also illustrated in Fig. 8a-c for the three cases described above.

Through the comparison of different scenarios based on the set of experimental variables to be optimised, the new results included in this paper show that both the specific choice of experimental design variables and the noise level show a relevant influence on the ANN ability in classifying the kinetic models. Regardless of the noise level, the best results are obtained with a larger set of experimental design variables to be optimised (case C), while the worst performance is achieved in case A, where the initial concentration of reactant A is fixed. From these observations it is concluded that the information in the data useful for the ANN to recognise the model structures is mainly dependent on temperature and initial concentration of A, while the sampling time has a lower impact. However, from the comparison of the ANN test accuracy results of case B and C, it can be observed that optimising also the allocation of sampling points in time in the experimental design vector can significantly improve the ANN performance when the data are affected by high noise.

From Fig. 9 and Table 4 the impact of the experimental error on the objective function is clearly evident. The best results are obtained when all the experimental variables are optimised (Fig. 8c) where 100% accuracy in ANN predictions is achieved in the low noise case; with medium noise levels the accuracy of 95% is reached, while in the high noise scenario the maximum ANN test-accuracy recorded is 90%. These results are valuable, considering that the ANN was employed for recognising the right kinetic model among a set of 8 candidate models characterised by similar structures, and that the test was conducted on 200 different simulated reacting systems. It must be remarked also that these results are obtained with a limited number of experiments and with just one sample per experiment.

Fig. 9 show that in all the scenarios considered the accuracy of the ANN reaches a plateau after 3 or 4 optimally designed experiments and cannot be increased further. Still, small fluctuations in the objective function values are observed, e.g. for the case C - low noise level scenario (Table 4), a slight decrease is observed from 100% to 99.5% with 5 and 6 experiments. A 0.5% decrease in the accuracy means that 1 of the 200 reacting systems in the test-set was misclassified, a variation considered to be due to the intrinsic randomness in the ANN training. Similar considerations can be done for the other scenarios.

5. Conclusions

This work extends the analysis on the framework for optimal design of experiments using artificial neural networks by detailing the formal definition of the optimisation problem, the *in silico* dataset structure, and the adjustable optimiser settings. The optimal DoE is coupled with the artificial neural network-based method for the identification of kinetic model. The proposed method requires as inputs the definition of the experimental design space, the library of candidate kinetic model structures, a set of feasible parameter values, the characterisation of the experimental noise, and the ANN hyperparameters. A genetic algorithm (differential evolution) is then employed to perform the DoE optimisation, aiming to find the conditions that lead to the best performance in classifying the kinetic model structures, i.e. maximising the ANN test-accuracy.

The method was demonstrated on a simulated case study,

Table 4
Highest value of ANN test-accuracy obtained after the design of experiments optimisation. Results for the three scenarios A, B,C (set of expeimental design variables) and at the different noise levels. Green colour highlights conditions where the accuracy was higher than 95%, black between 80 and 95%, red is for accuracy below 80%.

Scenario	Noise level		Optimal ANN test-accuracy (%)					
		# Experiments:	1	2	3	4	5	6
A	Low		50.5	96.0	97.0	98.5	98.0	97.0
	Medium		46.5	79.0	79.5	84.0	83.0	86.0
	High		34.5	54.0	56.0	55.0	53.0	59.0
B	Low		55.0	96.5	99.0	99.0	99.0	100
	Medium		53.5	87.5	93.5	93.0	94.0	96.5
	High		52.5	75.5	79.5	79.0	81.5	81.0
C	Low		64.0	97.5	99.5	100	99.5	99.5
	Medium		64.0	94.0	93.0	95.0	95.5	95.0
	High		60.5	79.0	87.5	88.0	90.0	89.0
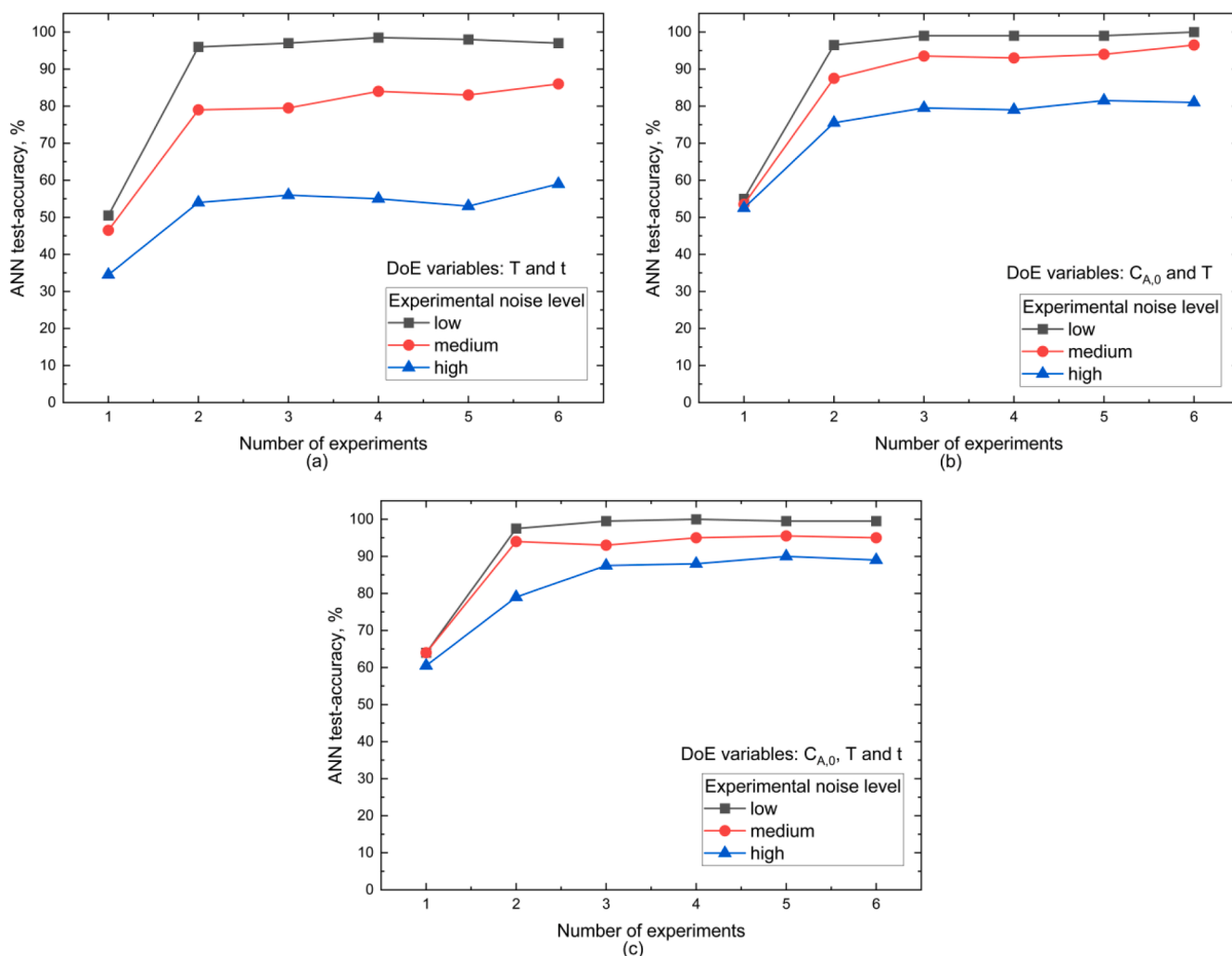
Fig. 9. Profiles of optimal ANN accuracy with respect to the number of experiments, reported for the three noise levels: low noise in black, medium noise in red and high noise in blue. Experimental design variables: (a) temperature and sampling time, (b) temperature and initial concentration of A, (c) temperature, initial concentration of A and sampling time.

considering a three-species reacting systems in a batch isothermal reactor and eight candidate model structures. In the case study, the choice of the set of free experimental design variables resulted of strong importance for achieving the best performance in the model selection. The analysis on the distribution of optimal experimental design variables in the experimental design space underlined that at low noise lower temperature values are chosen, compared to the high noise case

where the optimal experiments are characterized by higher temperature values. The results obtained when optimising all the possible experimental variables (case C) showed that not only high temperature, but also high initial concentration of reactant is preferred at high noise conditions.

Moreover, it was shown the impact of the system noise on the ANN test-accuracy on unseen data, i.e. not used for ANN training. Notable

results were that i) the ANN accuracy increased sharply with the number of experiments reaching a plateau after four optimally designed experiments in all the scenarios considered in this study, ii) optimising the sampling time is less important than temperature and reactant concentration, however its effect is relevant for noisy systems, iii) the algorithm tend to prefer experiments conducted at high temperatures when the noise is high. The highest values of ANN accuracy registered for the three noise levels were 100% (low noise), 96.5% (medium noise) and 90% (high noise).

Future research will assess the comparison between sequential and parallel strategies for the design of multiple experiments within the framework presented in this paper, considering the pros and cons in terms of computational burden and the robustness of the optimal solution with the two approaches. Future work needs also to include the application of the proposed DoE framework for ANN kinetic model recognition to more complex reacting systems (e.g. kinetics in multiphase systems and in the presence of catalysts) and to real chemical systems to validate the methodology.

In this work the methodology proposed starts from a fixed ANN architecture and focuses on optimising the DoE, which led to good results for the ANN accuracy. Another possible extension of the work by Quaglio et al. (2020b) is to evaluate the benefit of optimising the ANN architecture (Zhang et al., 2023) instead, leading to a comparison between the optimal DoE and optimal ANN architecture approaches.

CRediT authorship contribution statement

Enrico Sangoi: Writing – original draft, Software, Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Marco Quaglio:** Writing – review & editing, Supervision, Software, Methodology, Investigation, Conceptualization. **Fabrizio Bezzo:** Writing – review & editing, Supervision, Conceptualization. **Federico Galvanin:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

GitHub link to the code has been made available

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2024.108752.

References

Amato, F., González-Hernández, J.L., Havel, J., 2012. Artificial neural networks combined with experimental design: a "soft" approach for chemical kinetics. Talanta 93, 72–78. https://doi.org/10.1016/j.talanta.2012.01.044.

Arbib, M.A., 2002. The Handbook of Brain Theory and Neural Networks, 2nd ed. The MIT Press. https://doi.org/10.7551/mitpress/3413.001.0001.

Asprey, S.P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Comput. Chem. Eng. 24, 1261–1267. https://doi.org/10.1016/S0098-1354(00)00328-8.

Chakkingal, A., Janssens, P., Poissonnier, J., Barrios, A.J., Virginie, M., Khodakov, A.Y., Thybaut, J.W., 2021. Machine learning based interpretation of microkinetic data: a Fischer–Tropsch synthesis case study. React. Chem. Eng. 7, 101–110. https://doi.org/10.1039/D1RE00351H.

Galvanin, F., Cao, E., Al-Rifai, N., Gavriilidis, A., Dua, V., 2016. A joint model-based experimental design approach for the identification of kinetic models in continuous flow laboratory reactors. Comput. Chem. Eng. 95, 202–215. https://doi.org/10.1016/j.compchemeng.2016.05.009.

Géron, A., 2019. Hands-On Machine Learning With Scikit-Learn, Keras, and TensorFlow, 2nd ed. O'Riley Media, Inc.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw 2, 359–366. https://doi.org/10.1016/0893-6080(89)90020-8.

Kayala, M.A., Baldi, P., 2012. ReactionPredictor: prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning. J. Chem. Inf. Model. 52, 2526–2540. https://doi.org/10.1021/ci3003039.

Klatt, K.-U., Marquardt, W., 2009. Perspectives for process systems engineering—Personal views from academia and industry. Comput. Chem. Eng. 33, 536–550. https://doi.org/10.1016/j.compchemeng.2008.09.002. Selected Papers from the 17th European Symposium on Computer Aided Process Engineering held in Bucharest, Romania, May 2007.

Quaglio, M., Fraga, E.S., Galvanin, F., 2020a. A diagnostic procedure for improving the structure of approximated kinetic models. Comput. Chem. Eng. 133, 106659 https://doi.org/10.1016/j.compchemeng.2019.106659.

Quaglio, M., Roberts, L., Bin Jaapar, M.S., Fraga, E.S., Dua, V., Galvanin, F., 2020b. An artificial neural network approach to recognise kinetic models from experimental data. Comput. Chem. Eng. 135, 106759 https://doi.org/10.1016/j.compchemeng.2020.106759.

Rosenblatt, F., 1958. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. 65, 386–408. https://doi.org/10.1037/h0042519.

Russell, S., Norvig, P., 2021. Artificial Intelligence: A Modern Approach, Global Edition, 4th ed. Pearson.

Sangoi, E., Quaglio, M., Bezzo, F., Galvanin, F., 2022. Optimal design of experiments based on artificial neural network classifiers for fast kinetic model recognition. In: Yamashita, Y., Kano, M. (Eds.), Computer Aided Chemical Engineering, 14 International Symposium On Process Systems Engineering. Elsevier, pp. 817–822. https://doi.org/10.1016/B978-0-323-85159-6.50136-6.

Schwaab, M., Luiz Monteiro, J., Carlos Pinto, J., 2008. Sequential experimental design for model discrimination: taking into account the posterior covariance matrix of differences between model predictions. Chem. Eng. Sci. 63, 2408–2419. https://doi.org/10.1016/j.ces.2008.01.032.

Storn, R., Price, K., 1997. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. J. Glob. Optim. 11, 341–359. https://doi.org/10.1023/A:1008202821328.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Walczak, S., Cerpa, N., 2003. Artificial Neural Networks. In: Meyers, R.A. (Ed.), Encyclopedia of Physical Science and Technology (Third Edition). Academic Press, New York, pp. 631–645. https://doi.org/10.1016/B0-12-227410-5/00837-1.

Wei, J.N., Duvenaud, D., Aspuru-Guzik, A., 2016. Neural Networks for the Prediction of Organic Chemistry Reactions. ACS Cent. Sci. 2, 725–732. https://doi.org/10.1021/acscentsci.6b00219.

Zhang, S., Vassiliadis, V.S., Hao, Z., Cao, L., Lapkin, A.A., 2023. Heuristic optimisation of multi-task dynamic architecture neural network (DAN2). Neural Comput. Appl. 35, 4775–4791. https://doi.org/10.1007/s00521-022-07851-9.