

Journal Pre-proof

An artificial neural network approach to recognise kinetic models from experimental data

Marco Quaglio, Louise Roberts, Mohd Safarizal Bin Jaapar, Eric S. Fraga, Vivek Dua, Federico Galvanin

PII: S0098-1354(19)30229-7
DOI: <https://doi.org/10.1016/j.compchemeng.2020.106759>
Reference: CACE 106759



To appear in: *Computers and Chemical Engineering*

Received date: 26 February 2019
Revised date: 7 October 2019
Accepted date: 23 January 2020

Please cite this article as: Marco Quaglio, Louise Roberts, Mohd Safarizal Bin Jaapar, Eric S. Fraga, Vivek Dua, Federico Galvanin, An artificial neural network approach to recognise kinetic models from experimental data, *Computers and Chemical Engineering* (2020), doi: <https://doi.org/10.1016/j.compchemeng.2020.106759>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

An artificial neural network approach to recognise kinetic models from experimental data

Marco Quaglio^a, Louise Roberts^a, Mohd Safarizal Bin Jaapar^a, Eric S. Fraga^a, Vivek Dua^{a,*}, and Federico Galvanin^{a,*}

^a*Department of Chemical Engineering, University College London (UCL), Torrington Place, WC1E 7JE London, United Kingdom*

*corresponding authors e-mail: v.dua@ucl.ac.uk; f.galvanin@ucl.ac.uk

Abstract

The quantitative description of the dynamic behaviour of reacting systems requires the identification of an appropriate set of kinetic model equations. The selection of the correct model may pose substantial challenges as there may be a large number of candidate kinetic model structures. In this work, a model selection approach is presented where an Artificial Neural Network classifier is trained for recognising appropriate kinetic model structures given the available experimental evidence. The method does not require the fitting of kinetic parameters and it is well suited when there is a high number of candidate kinetic mechanisms. The approach is demonstrated on a simulated case study on the selection of a kinetic model for describing the dynamics of a three-component reacting system in a batch reactor. The sensitivity of the approach to a change in the experimental design and to a change in the system noise is assessed.

keywords: model selection, model discrimination, identifiability, machine learning, design of experiment

1 Introduction

Modelling the kinetic behaviour of chemical reactions requires the construction of systems of differential and algebraic equations potentially involving a high number of state variables and kinetic parameters. The identification of kinetic models requires *i*) the selection of an appropriate functional form for the model equations and *ii*) the estimation of its kinetic parameters from experimental data (Bonvin et al., 2016). Both stages may pose significant challenges to the modeller. More specifically, there may be significant uncertainty on the relevant reactions occurring in the system and on the most appropriate functional forms for describing their dynamics. Furthermore, even if an appropriate model structure is selected, the estimation of its kinetic parameters may be impossible to perform due to identifiability problems associated to the proposed kinetic model structure (Raue et al., 2009).

A variety of tools for model validation have been proposed in the literature to leverage modelling and experimental efforts in kinetic modelling studies. Model building procedures

based on data fitting start with the construction of a number of candidate model structures (Asprey and Macchietto, 2000). An identifiability analysis is then performed to evaluate if the parameters involved in the candidate models can be estimated from experimental data (Cobelli and Di Stefano, 1980; Galvanin et al., 2013). Models which do not pass the identifiability check are rejected at this stage. It is important to observe that even the *exact* model may be rejected if it does not satisfy the identifiability requirement. The kinetic parameters of the remaining models are then estimated fitting available measurements (Bard, 1974) and the fitting quality is assessed with a statistical test on the goodness-of-fit (MacKay, 1992; Silvey, 1975). Information-theoretic approaches for model selection can be employed to choose the best fitting model penalising unnecessarily complex model structures (Burnham and Anderson, 2002). Popular criteria for model selection are the Akaike information criterion (AIC) (Akaike, 1974), which lies its foundations in frequentist inference, and the Bayesian information criterion (BIC) (Schwarz, 1978). If more than one model is found adequate to represent the data, one may proceed by designing additional experiments with the aim of discriminating among the competing model structures (Buzzi-Ferraris et al., 1990; Olofsson et al., 2019) and then for improving parameter precision (Franceschini and Macchietto, 2008).

Whenever the selection of a *physics-based* kinetic model is impractical, e.g. because of an extremely high number of possible reaction pathways, one may prefer to invest modelling efforts in the identification of a *data driven* model. This may be any parametric model, e.g. a polynomial or a response surface (Box and Draper, 1987), which provides a convenient representation of the data (Bonvin et al., 2016). The structure of data driven models typically does not reflect the inner mechanisms of the physical system. Data driven modelling approaches represent one of the foundational paradigms in machine learning technologies (Barber, 2011; LeCun et al., 2015). Data driven models and machine learning approaches have already found successful application in many contexts in chemical engineering (Venkatasubramanian, 2019), particularly in the fields of materials design (Janet et al., 2018; Hou et al., 1997), process operations (Lopes et al., 2018; Molga et al., 2006; Petsagkourakis et al., 2019; Quadros et al., 2005) and fault diagnosis (Zhao et al., 2019).

A class of machine learning models which has recently seen an increase in popularity is the Artificial Neural Network (ANN) model (Krizhevsky et al., 2012; Russell and Norvig, 2016). The recent success of ANNs is associated primarily with 1) their flexibility in approximating any nonlinear continuous function (Hornik et al., 1989) 2) the development of efficient algorithms for ANN training (Geron, 2017; Hinton et al., 2006) and 3) a steady decrease in the cost of computational power (Russell and Norvig, 2016). In chemical engineering, ANNs have been applied to address both regression and classification problems (Himmelblau, 2008; Lee et al., 2018). ANNs were employed for nonlinear system identification (Dua, 2011; Kramer, 1991; Petsagkourakis et al., 2019; Traver et al., 1999), model reduction (Prasad and Bequette, 2003) and process control (Bloch and Denoeux, 2003; Hussain and Kershenbaum, 2000). ANN-based classifiers have been used to support drug discovery (Wang et al., 2005), catalyst design (Goldsmith et al., 2018), reaction prediction (Coley et al., 2019; Kayala and Baldi, 2012; Wei et al., 2016) and fault detection (Rengaswamy and Venkatasubramanian, 2000; Suewatanakal, 1993). In contrast to physics-based models, the estimation of parameters in ANNs typically requires substantial amounts of data. Furthermore, it is extremely challenging to assign physical significance to the ANNs parameters and accurate extrapola-

tion beyond the conditions used for the identification of the ANN is generally not possible.

In this work, a novel framework for the selection of physics-based kinetic models is proposed. In the proposed framework, an ANN-based classifier is trained from in-silico experimental data with the aim of *recognising* the most appropriate kinetic model given the available experimental evidence. It is shown that the approach is effective for discriminating among rival model structures even when the kinetic models are not structurally identifiable. In fact, the estimation of kinetic parameters is not required in the procedure.

The present manuscript is structured as follows. A general overview on the ANN model is given in Section 2. The proposed ANN-based kinetic model recognition framework is detailed in Section 3. The framework is demonstrated on a simulated case study, which is detailed in Section 4. Results are presented and discussed in Section 5.

2 Artificial Neural Network classifier

Artificial Neural Networks are parametric models whose structure is loosely inspired by biological neural networks. In biological brains, a high number of interacting neural cells respond to input electrical stimuli by *firing* (i.e. transmitting) an electrical signal to downstream neural cells in the network. First attempts of modelling the logic behaviour of neural cells led to the development of the single layer perceptron (Rosenblatt, 1962), represented in Figure 1a. The perceptron is a function which transforms an $N_n \times 1$ input array \mathbf{n} of real values into a scalar output p . In the analogy with the neural cell, the elements of \mathbf{n} represent the signals received from the upstream neurons and p represents the signal transmitted to the downstream neuron. The single layer perceptron is described by

$$p = \varphi(\mathbf{w}^T \mathbf{n} + b) \quad (1)$$

In (1), \mathbf{w} is an $N_n \times 1$ array of parameters, b is a bias parameter and φ represents an activation function. Popular choices for the activation function are the rectified linear unit (ReLU), the sigmoid and the hyperbolic tangent (Russell and Norvig, 2016).

A feedforward ANN can be built by arranging a set of perceptrons into layers where the output of each layer is fed as input to the following layer in the network. The structure of a two-layers feedforward ANN is given in Figure 1b, where grey-coloured circles represent the perceptrons. The last layer in the network is called the output layer, while the previous layer is a hidden layer.

Let N_h be the number of perceptrons in the hidden layer. The two-layers feedforward ANN represented in Figure 1b is described by the expression

$$\mathbf{p} = \varphi_o[\mathbf{W}_o^T \varphi_h(\mathbf{W}_h^T \mathbf{n} + b_h \mathbf{1}_h) + b_o \mathbf{1}_o] \quad (2)$$

In (2), \mathbf{p} is a $N_m \times 1$ array of output values; \mathbf{W}_o [$N_h \times N_m$] and \mathbf{W}_h [$N_n \times N_h$] are matrices of network parameters associated with the output layer and with the hidden layer respectively; scalars b_h and b_o are bias parameters and $\mathbf{1}_h$ [$N_h \times 1$] and $\mathbf{1}_o$ [$N_m \times 1$] are arrays whose entries are all equal to 1. The functions φ_o and φ_h represent respectively the activation function in the output layer and the activation function in the hidden layer.

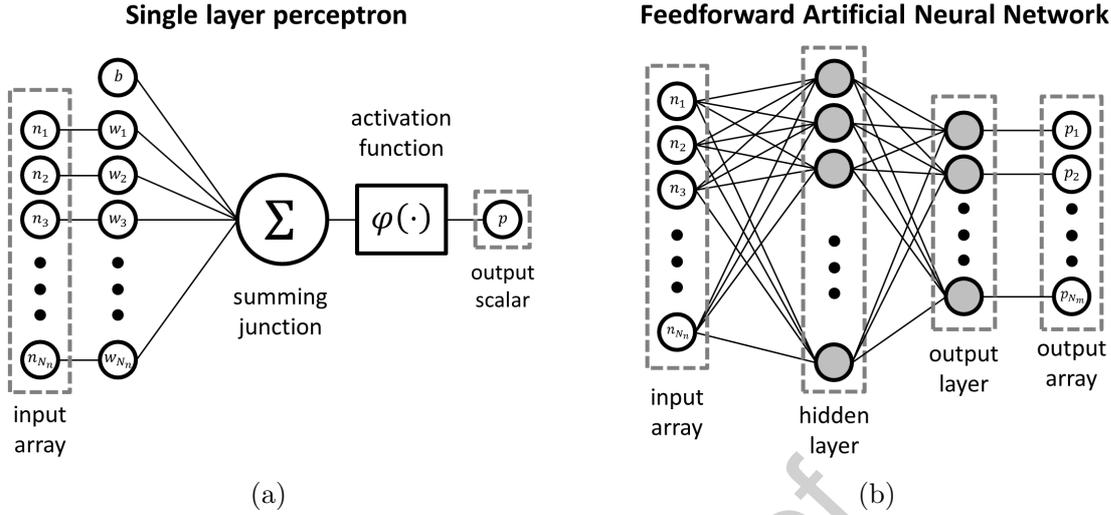


Figure 1: (a) Single layer perceptron; (b) Two-layers Feedforward Artificial Neural Network, i.e. a multilayer perceptron structure. In (b), grey-coloured circles represent the perceptrons.

When the ANN is employed for classification purposes, the aim is to assign a label, i.e. a categorical variable $l \in \{1, \dots, N_m\}$, to the input array from a set of N_m possible labels. For such applications, a popular choice for the activation function φ_o in the output layer is the *softmax* (Arbib, 2003). The softmax function returns a normalised $N_m \times 1$ output \mathbf{p} , whose elements p_k with $k = 1, \dots, N_m$ satisfy the condition $\sum_{k=1}^{N_m} p_k = 1$. Hence, $p_k \forall k = 1, \dots, N_m$ may be interpreted as a predicted probability that the input belongs to the k -th class. The label \hat{l} assigned by the ANN classifier to the input is then computed as the index associated to the largest element of \mathbf{p} .

$$\hat{l} = \arg \max_k p_k \quad (3)$$

The construction of an ANN-based classifier requires the selection of a structure for the network, i.e. the selection of the number of hidden layers and the number of perceptrons per layer (Dua, 2010). Once the structure of the ANN is fixed, the *training* of the ANN (i.e. the tuning of the ANN parameters) is performed through a process of regularised regression (Russell and Norvig, 2016). The training involves the fitting of a dataset of labelled input arrays $\Psi = \{(\mathbf{n}_i, l_i) \forall i = 1, \dots, N_\psi\}$, where N_ψ represents the number of elements in the dataset. Since the number of parameters involved in an ANN structure is typically large, the training may require substantial amounts of data. If data are obtained from experiments, the identification of an ANN may be impractical because of a high requirement for experimental resources. However, simulated experiments offer an interesting alternative and can be used for generating in-silico large amounts of data at negligible cost. The possibility of training ANNs from synthetic data has been demonstrated in several works in the literature (Bates et al., 2018; Jaderberg et al., 2014; Le et al., 2017; Mnih et al., 2015).

3 Kinetic model recognition

We assume that a setup is available for conducting kinetic experiments on a reacting system of interest. In the setup, \mathbf{u} is an $N_u \times 1$ array of manipulated system inputs and \mathbf{y} is an $N_y \times 1$ array of state variables that can be sampled over time. The variable time is denoted as t . N_m potential model structures are proposed for describing the dynamic behaviour of the system:

$$\begin{aligned} \mathbf{f}_l(\dot{\mathbf{x}}_l, \mathbf{x}_l, \mathbf{u}, t, \boldsymbol{\theta}_l) &= \mathbf{0} \\ \hat{\mathbf{y}}_l &= \mathbf{h}_l(\mathbf{x}_l) \end{aligned} \quad \forall l = 1, \dots, N_m \quad (4)$$

In (4), quantities appearing with subscript l refer to the l -th candidate model. More specifically, $\mathbf{f}_l [N_{f,l} \times 1]$ and $\mathbf{h}_l [N_y \times 1]$ are arrays of model equations characterising the structure of the l -th kinetic model. $\mathbf{x}_l [N_{x,l} \times 1]$ and $\dot{\mathbf{x}}_l [N_{x,l} \times 1]$ are, respectively, an array of state variables and an array of time derivatives for the state variables involved in the l -th model. $\hat{\mathbf{y}}_l$ is the $N_y \times 1$ array of predictions associated to the l -th model for the measurable system states \mathbf{y} . Quantity $\boldsymbol{\theta}_l \in \Theta_l$ represents the $N_{\theta,l} \times 1$ array of kinetic parameters involved in the l -th model structure. The aim of the modeller is to select the most appropriate kinetic model among the N_m candidates.

A model selection approach is proposed in this work where an ANN-based classifier is used to recognise the most appropriate kinetic model based on available experimental evidence. The framework is given in Figure 2. The procedure starts from the definition of a library of possible kinetic model as in (4) and an experimental design, which defines the conditions for the collection of N samples of \mathbf{y} in the experimental design space Φ . It is also assumed that the measurement error present in the system is fully characterised as uncorrelated Gaussian noise.

Let $\boldsymbol{\varphi}_j \in \Phi \forall 1, \dots, N$ be the experimental conditions for the j -th designed sample. The designed experiments are conducted on the available experimental setup leading to the collection of the samples \mathbf{y}_j with $j = 1, \dots, N$. The procedure then involves the identification of an ANN-based classifier for assigning the most appropriate model class $\hat{l} \in \{1, \dots, N_m\}$ to the array of experimental data $\mathbf{n} = [\mathbf{y}_1^T, \dots, \mathbf{y}_N^T]^T$.

The identification of the classifier is performed in two stages:

1. *In-silico data generation stage.* Samples are generated in-silico by integrating the candidate model equations at the conditions defined by the chosen experimental design. For each model structure l ($\forall l = 1, \dots, N_m$), the in-silico experimental campaign is repeated multiple times using different values for the parameters $\boldsymbol{\theta}_l \in \Theta_l$. A labelled dataset Ψ is then built from the simulated samples where the data generated in each simulated campaign are labelled with the model structure l used to generate the data.
2. *ANN construction and identification stage.* The labelled dataset is used for the construction, training, validation and testing of an ANN-based classifier. The aim at this stage is to make the ANN *learn* about the physical system by using experimental data generated in-silico.

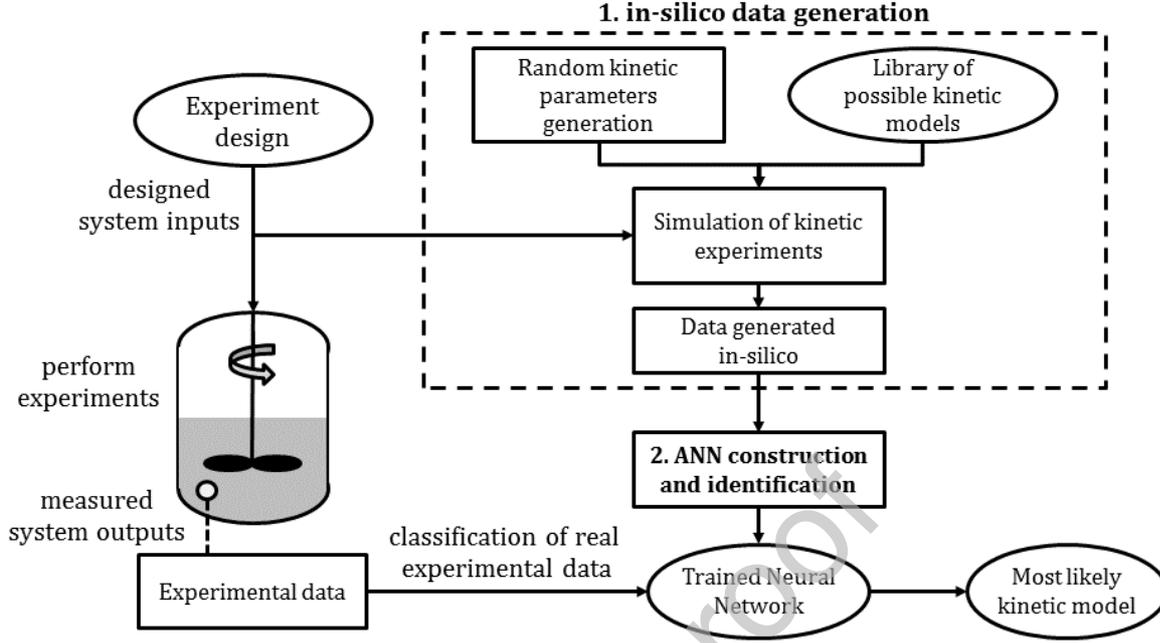


Figure 2: Proposed ANN-based approach for kinetic model recognition. In the framework, an ANN-based classifier is trained from in-silico experimental data for selecting the most appropriate kinetic model based on the available experimental evidence.

In the following subsections, the aforementioned steps in the procedure are further detailed. The steps are demonstrated on a simulated case study in Section 4 on a three-components reaction conducted in a batch reactor.

3.1 In-silico data generation

A labelled dataset Ψ is constructed from experimental data generated in-silico as in (5). In Ψ , the elements in the generic labelled array \mathbf{n}_i [$N \cdot N_y \times 1$] represent simulated measurements, which are obtained by integrating the equations of the kinetic model $l = l_i$.

$$\Psi = \{(\mathbf{n}_i, l_i) \forall i = 1, \dots, N_\psi \quad \text{s.t.} \quad l_i \in \{1, \dots, N_m\}\} \quad (5)$$

Dataset Ψ shall be built including a comparable number of elements for each model structure $l = 1, \dots, N_m$. In (6), the generic labelled array \mathbf{n}_i is defined as the sum of two contributions: 1) a prediction for the array of experimental data $\hat{\mathbf{n}}_i$ [$N \cdot N_y \times 1$] and 2) a random function ϵ [$N \cdot N_y \times 1$].

$$\mathbf{n}_i = \hat{\mathbf{n}}_i + \epsilon(\hat{\mathbf{n}}_i) \quad \forall i = 1, \dots, N_\psi \quad (6)$$

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

$$\hat{\mathbf{n}}_i^T = [\hat{y}_l(\boldsymbol{\varphi}_1, \boldsymbol{\theta}_l)^T, \dots, \hat{y}_l(\boldsymbol{\varphi}_N, \boldsymbol{\theta}_l)^T]_{\boldsymbol{\theta}_l = \mathcal{U}(\Theta_l), l=l_i} \quad (7)$$

$$\forall i = 1, \dots, N_\psi$$

In (7), predictions for the N samples are computed from the l_i -th kinetic model using a random value for its kinetic parameters Θ_{l_i} . The parameter value is set equal to a random function $\mathcal{U}(\cdot)$, which returns a realisation of a multivariate uniform distribution defined over its argument. More specifically, the function $\mathcal{U}(\Theta_{l_i})$ returns a random parameter set from the feasible parameter domain Θ_{l_i} . A uniform distribution is employed to avoid the introduction of bias on the possible values of the model parameters.

The random function ϵ is introduced in (6) to simulate measurement noise and it should be chosen to represent the level of noise in the physical system. In this work measurement errors are assumed to be uncorrelated and normal with zero mean. However, different noise models are considered to take into account both a constant noise variance and a noise variance proportional to the measured value. The noise model is defined as

$$\epsilon(\hat{\mathbf{n}}_i) \sim \mathcal{N}(\mathbf{0}, \Sigma(\hat{\mathbf{n}}_i)) \quad (8)$$

In (8), $\mathcal{N}(\mathbf{0}, \Sigma)$ is a multivariate normal distribution with mean zero and covariance Σ $[N \cdot N_y \times N \cdot N_y]$. The covariance Σ is a diagonal matrix whose diagonal elements are computed as the sum of a term proportional to the predicted data and a constant term. The jk -th element of Σ , namely σ_{jk} , is defined as

$$\sigma_{jk} = \begin{cases} \sigma_r^2 \cdot \frac{1}{100} \cdot \hat{n}_{i,j} + \sigma_c^2 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \quad \forall j, k \quad (9)$$

In (9), $\hat{n}_{i,j}$ represents the j -th element in $\hat{\mathbf{n}}_i$. Scalar quantities σ_r and σ_c are user-defined parameters that respectively quantify the contribution of relative and constant variance to the total noise variance.

3.2 ANN construction and identification

Initially the dataset is split into three subsets: *i*) a training set Ψ_{training} ; *ii*) a validation set $\Psi_{\text{validation}}$; *iii*) a test set Ψ_{test} . Typically the dataset is split in the subsets following the 60-20-20 rule (Geron, 2017). The accuracy of ANN-based classifiers in representing a labelled dataset is quantified with the categorical *cross entropy* (Russell and Norvig, 2016). The parameters of the ANN are estimated with the aim of minimising the cross entropy on the training set Ψ_{training} . The cross entropy on the validation set $\Psi_{\text{validation}}$ is monitored during the regression to support the detection of overfitting (Geron, 2017). The accuracy on the validation set is also used to perform cross-validation and select among different ANN structures (Geron, 2017). More specifically, the structure of the ANN associated to the highest validation accuracy at the end of the training process is selected as the optimal ANN structure.

When the network is identified, the accuracy is evaluated on the test set Ψ_{test} . This provides an unbiased quantification of the ANN accuracy on the classification of *unseen* data, i.e. data not used in the training process nor in the ANN structure selection process. The test accuracy is defined as the percentage of correctly labelled elements in the test set as in (10).

$$\text{Test Accuracy} = \frac{|\{i \in \{1, \dots, N_\psi\} \text{ s.t. } (\mathbf{n}_i, l_i) \in \Psi_{\text{test}} \wedge \hat{l}_i = l_i\}|}{|\{i \in \{1, \dots, N_\psi\} \text{ s.t. } (\mathbf{n}_i, l_i) \in \Psi_{\text{test}}\}|} \cdot 100\% \quad (10)$$

In (10), \hat{l}_i and l_i are respectively the predicted and the actual model structure for the i -th input in the dataset Ψ ; the symbol $|\cdot|$ denotes the cardinality operator. The accuracy on the test set is presented by means of a $N_m \times N_m$ table $\mathbf{\Gamma}$, namely a *confusion matrix*. In this manuscript, the jk -th element of $\mathbf{\Gamma}$, i.e. γ_{jk} , is defined as

$$\gamma_{jk} = \frac{|\{i \in \{1, \dots, N_\psi\} \text{ s.t. } (\mathbf{n}_i, l_i) \in \Psi_{\text{test}} \wedge l_i = j \wedge \hat{l}_i = k\}|}{|\{i \in \{1, \dots, N_\psi\} \text{ s.t. } (\mathbf{n}_i, l_i) \in \Psi_{\text{test}} \wedge l_i = j\}|} \quad (11)$$

In (11), γ_{jk} represents the ratio between the number of class j elements in the test set Ψ_{test} that are classified as class k elements and the number of class j elements in Ψ_{test} . When no misclassification occurs, matrix $\mathbf{\Gamma}$ is the $N_m \times N_m$ identity matrix. As an example, a confusion matrix is reported in Figure 3 assuming $N_m = 3$ candidate kinetic models. In the given example, 85% of model type 1 in the test set are correctly classified while 5% are misclassified as type 2 and 10% as type 3; 96% of model type 2 in the test set are correctly classified while 4% are misclassified as type 1; 100% of model type 3 in the test set are classified correctly.

Under the assumption that the noise model implemented in ϵ is representative of the measurement noise in the system, the test accuracy and the confusion matrix can be interpreted as indexes of ANN reliability on the classification of real system data. A dense confusion matrix $\mathbf{\Gamma}$ indicates that the ANN is not reliable and that models are not distinguishable given the level of system noise and the selected experimental design. The ANN is considered reliable if $\mathbf{\Gamma}$ tends to the identity matrix, i.e. most/all extra-diagonal elements in $\mathbf{\Gamma}$ are null or close to zero. A reliable ANN may be employed to recognise the most appropriate kinetic model from unlabelled data collected in kinetic experiments.

If the confusion matrix is non-diagonal for any choice of the experimental design, this may highlight the presence of structural distinguishability issues in the kinetic model structures (Walter and Pronzato, 1997). As an example, it may happen that an extra-diagonal element γ_{jk} is nearly equal to the corresponding diagonal element γ_{jj} on the same row (i.e. $\gamma_{jk} \simeq \gamma_{jj}$ for some j and $k \neq j$). This highlights that model j and model k may produce coincident predictions. In different words, models j and k may be equivalent representations of the phenomenon and a reformulation of their model structures may be required.

4 Case study

The procedure described in Section 3 is demonstrated in this section on a simulated case study to demonstrate that ANN classifiers can successfully be employed to recognise kinetic model structures from experimental data. The performance of the proposed framework is evaluated at variable experimental conditions and different levels of measurement noise in the system.

4.1 Reacting system

The considered chemical system is a three-component reacting mixture, where chemical species are denoted as A, B and C. The concentration of the components in the mixture is

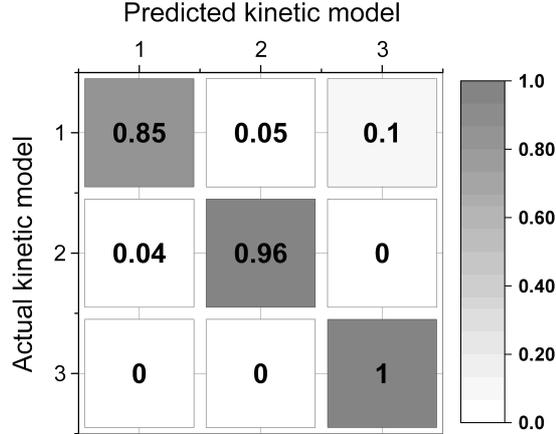


Figure 3: Example of confusion matrix with $N_m = 3$ kinetic model structures.

denoted as C_i [mol m^{-3}] $\forall i \in \{A, B, C\}$. The reaction network involves $N_r = 3$ reactions as shown in (12), where r_j [$\text{mol m}^{-3} \text{s}^{-1}$] with $j = 1, \dots, N_r$ denote the reaction rates.



The experiments are assumed to be conducted in a perfectly mixed and isothermal batch reactor. A sample is constituted by a set of measured concentration for all the components in the mixture, i.e. $\mathbf{y} = [C_A, C_B, C_C]^T$. Initial concentrations in the experiments are assumed to be fixed, i.e. $C_A(0) = 100 \text{ mol m}^{-3}$, $C_B(0) = 0 \text{ mol m}^{-3}$, $C_C(0) = 0 \text{ mol m}^{-3}$. The experimental conditions that can be controlled for the collection of a sample are 1) the temperature T [K] in the range 573 - 673 K and 2) the sampling time t_s [s] in the range 100.0 - 300.0 s. These constitute independent directions in the experimental design space $\Phi = \{(T, t) \text{ s.t. } 573 \text{ K} \leq T \leq 673 \text{ K} \wedge 100.0 \text{ s} \leq t_s \leq 300.0 \text{ s}\}$. It is also assumed that for all conditions in the experimental design space, the observed conversion of reactant A, i.e. X_A , is between 0.05 and 0.95 and that the selectivity with respect to product B and C, i.e. S_B and S_C respectively, is always above 0.05. Constraints on conversion and selectivity are assumed for preventing the system from exhibiting a limit-case behaviour, e.g. cases in which A does not react or cases in which the system tends to the single-reaction behaviour and only B or C is produced. Such constraints represent prior knowledge on the system kinetics and will be employed to bound the feasible parameter domains of the candidate models.

4.2 Candidate kinetic models

The system is described by the general set of equations

$$\frac{dC_i}{dt} = \sum_{j=1}^{N_r} \nu_{ij} r_j \quad \forall i \in \{A, B, C\} \quad (13)$$

In (13), ν_{ij} is the stoichiometric coefficient associated to the i -th component in the j -th reaction. Nonetheless, there is uncertainty on which is the appropriate functional form for the reaction rate terms r_j with $j = 1, \dots, N_r$. A set of $N_m = 8$ candidate kinetic models is assumed. The possible functional forms for the rates in the N_m model structures are reported in Table 1. Kinetic models $l = 1, \dots, 4$ assume a series kinetic mechanism while models $l = 5, \dots, 8$ assume a parallel kinetic mechanism. $k_j \forall j = 1, \dots, N_r$ represent Arrhenius-type rate constants and are given by

$$k_j = A_j e^{-\frac{E_{a,j}}{RT}} \quad \forall j = 1, \dots, N_r \quad (14)$$

where $A_j \forall j = 1, \dots, N_r$ are the pre-exponential factors and $E_{a,j}$ [J mol⁻¹ K⁻¹] $\forall j = 1, \dots, N_r$ represent activation energies. The parameters involved in the kinetic model structures are $\Theta_l = [A_1, A_2, A_3, E_{a,1}, E_{a,2}, E_{a,3}] \forall l = 1, \dots, N_m$.

In kinetic models $l = 1, \dots, 4$, the rate constant k_2 is multiplied by zero. Therefore, the predicted concentration profiles are insensitive to a change in the kinetic parameters A_2 and $E_{a,2}$. Analogously to models $l = 1, \dots, 4$, in models $l = 5, \dots, 8$ the kinetic constant k_3 is multiplied by zero and concentrations predictions are insensitive to a change in A_3 and $E_{a,3}$. Therefore, none of the candidate models satisfies the requirement of structural identifiability (Cobelli and Di Stefano, 1980; Raue et al., 2009).

The feasible parameter domain for each kinetic model $l = 1, \dots, N_m$ is defined as the set of real values such that: *i*) pre-exponential factors and activation energies lie between lower and upper physical bounds; *ii*) the predicted conversion of reactant A is between 0.05 and 0.95 at all conditions in the experimental design space; *iii*) the predicted selectivity towards components B and C is above 0.05 at all conditions in the design space. Mathematically, this is described by the set of constraints in the form (15), where symbols $\hat{X}_{i,l}$ and $\hat{S}_{i,l}$ represent model l predictions for conversion and selectivity for the i -th component in the mixture.

$$\Theta_l = \{ \boldsymbol{\theta}_l \in \mathbb{R}^{N_{\theta,l}} \text{ s.t. } \begin{array}{lll} 100 \leq A_j \leq 200 & \forall j = 1, \dots, N_r & \wedge \\ 45 \leq E_{a,j} \leq 90 & \forall j = 1, \dots, N_r & \wedge \\ 0.05 \leq \hat{X}_{A,l} \leq 0.95 & \forall \boldsymbol{\varphi} \in \Phi & \wedge \\ 0.05 \leq \hat{S}_{B,l} & \forall \boldsymbol{\varphi} \in \Phi & \wedge \\ 0.05 \leq \hat{S}_{C,l} & \forall \boldsymbol{\varphi} \in \Phi & \} \end{array} \quad (15)$$

$$\forall l = 1, \dots, N_m$$

Prior knowledge on conversion and selectivity is included in the definition of the parameter domains primarily to generate datasets for ANN training using parameter values that are in agreement with the known physical constraints on the system behaviour.

4.3 Methods

An algorithm reflecting the framework illustrated in Section 3 was implemented in Python 3.5 (Python Core Team, 2018). The aim in this case study is to assess the classification capabilities of ANNs in recognising the most appropriate kinetic model from experimental data. Different cases are proposed to assess the sensitivity of the ANN accuracy to the choice of the experimental design and to the measurement noise in the system.

Table 1: Candidate functional forms for the reaction rates. Kinetic models $l = 1, \dots, 4$ assume a series mechanism while models with $l = 5, \dots, 8$ assume 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	$k_3 \cdot C_B$	$k_3 \cdot C_B^2$	$k_3 \cdot C_B$	$k_3 \cdot C_B^2$	$k_3 \cdot 0$	$k_3 \cdot 0$	$k_3 \cdot 0$	$k_3 \cdot 0$

- *Case 1.* A full factorial experiment design is assumed with three levels for the temperature $T = \{573, 623, 673\}$ K and three levels for the sampling time $t_s = \{100, 200, 300\}$ s. Hence, the design consists of $N = 9$ samples of \mathbf{y} , i.e. $N \cdot N_y = 27$ concentration measurements. The measurement noise in the system is assumed to be characterised by $\sigma_r = 2.0$ and $\sigma_c = 0.04$.
- *Case 2.* A factorial design with three levels for the temperature $T = \{573, 623, 673\}$ K and one level for the sampling time $t_s = \{300\}$ s is considered. The measurement noise is the same as in case 1 with $\sigma_r = 2.0$ and $\sigma_c = 0.04$.
- *Case 3.* The same design as in case 1 is adopted but assuming a higher measurement noise, characterised by $\sigma_r = 5.0$ and $\sigma_c = 10.0$.

In all cases, a dataset Ψ is constructed with $N_\psi = 1000$ labelled elements. The dataset Ψ is split including 600 elements in the training set, 200 elements in the validation set and 200 elements in the test set. The Python library *Keras 2.2.4* is used for the implementation and training of the ANN classifiers (Chollet et al., 2015). The Adam algorithm (Kingma and Ba, 2014) is employed for the training of the ANNs setting a dropout regularisation of 10% (Geron, 2017). A two-layer ANN structure is adopted in all cases where the Rectified Linear Unit (ReLU) activation function is used in the hidden layer (Geron, 2017). The number of perceptrons in the hidden layer, i.e. N_h , is optimised through grid search over the range 1 – 400 (Geron, 2017). Data are mean-centred and normalised before being fed to the ANN by using the function *StandardScaler* implemented in the Python library *scikit-learn 0.20.1* (Pedregosa et al., 2011).

The classification capabilities of the ANN-based classifier are measured in terms of accuracy in predicting the *unseen* data in the test set Ψ_{test} . The test accuracy is evaluated according to (10). The confusion matrix, whose elements are evaluated according to (11), is also reported for each of the aforementioned cases.

5 Results

The number of nodes in the hidden layer of the ANN is optimised through grid search. The accuracy of the ANN in representing the validation set is reported in Figure 4 as a function

of the number of hidden nodes for all the cases considered in the study. In each case, the smallest number of hidden neurons associated with the highest validation accuracy is selected as the optimal number of neurons in the hidden layer.

The designed samples in case 1 are plotted in Figure 5a. In case 1, a maximum validation accuracy was achieved with a two-layer ANN with $N_h = 100$. A test accuracy of 98.0% was achieved on the classification of 200 *unseen* kinetic models in the test set. The confusion matrix $\mathbf{\Gamma}$ related to case 1 is proposed in Figure 5b. As one can see from Figure 5b, no misclassification occurred in relation to kinetic models 1, 3, 4, 5, 6 and 8, i.e. the diagonal elements associated to the aforementioned models are equal to 1. A fraction of kinetic models 2 and 7 was misclassified in case 1. More specifically, 4% of model 2 instances in the test set were misclassified as model 1 and 14% of model 7 were misclassified as model 8.

The three samples designed in case 2 are plotted in Figure 6a. An optimal number of hidden nodes $N_h = 400$ was selected in this case and the achieved test accuracy was 75%. The corresponding confusion matrix is shown in Figure 6b. One can see that the confusion matrix is denser than in case 1, i.e. more extra-diagonal elements differ from zero. From Figure 6, it can be appreciated that only the diagonal element γ_{33} is equal to 1. In different words, all model 3 elements in the test set are correctly classified by the ANN while a certain degree of misclassification is observed for the other model classes. In particular, element γ_{67} in the confusion matrix in Figure 6 is equal to 0.62. Hence, 62% of model 6 elements in the test set were misclassified as model 7.

In case 3, the same experimental design considered in case 1 is assumed and it is shown in Figure 7a. An optimised number $N_h = 220$ of hidden nodes was selected through grid search in case 3. The high system noise assumed in case 3 has a negative impact on the test accuracy, which reduces to 64.5%. As one can see from the confusion matrix shown in Figure 7b, a certain degree of misclassification is present for all model classes, i.e. none of the diagonal elements in $\mathbf{\Gamma}$ is equal to 1.

Table 2: Summary of results for the considered cases. The test accuracy is measured on the classification of 200 unseen arrays of data.

Case number	Case description	Hidden nodes N_h	Test Accuracy
1	base case	100	98.0%
2	reduced samples	400	75.0%
3	increased noise	220	64.5%

5.1 Results discussion

The results obtained in the three cases are summarised in Table 2. From Table 2, one can see that either a reduction in the number of designed samples or an increase in the system noise contribute to a reduction of the test accuracy. Some degree of misclassification was present in all cases, i.e. some extra-diagonal elements in the confusion matrices are different from

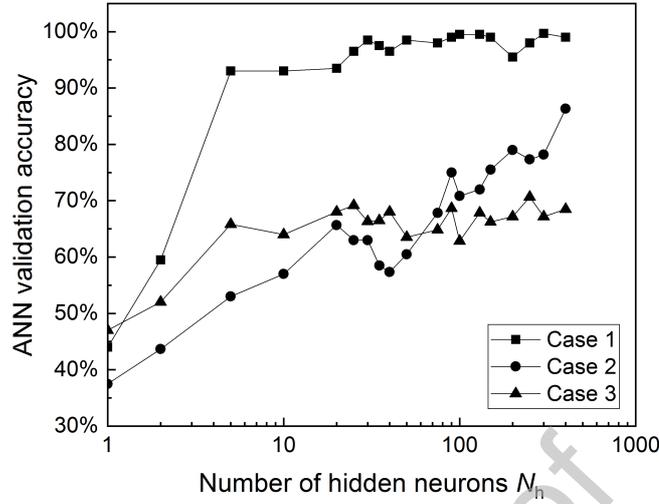


Figure 4: Accuracy of a two-layers ANN in representing the validation set as a function of the number of neurons in the hidden layer N_h . The validation accuracy is reported for the three cases considered in this study.

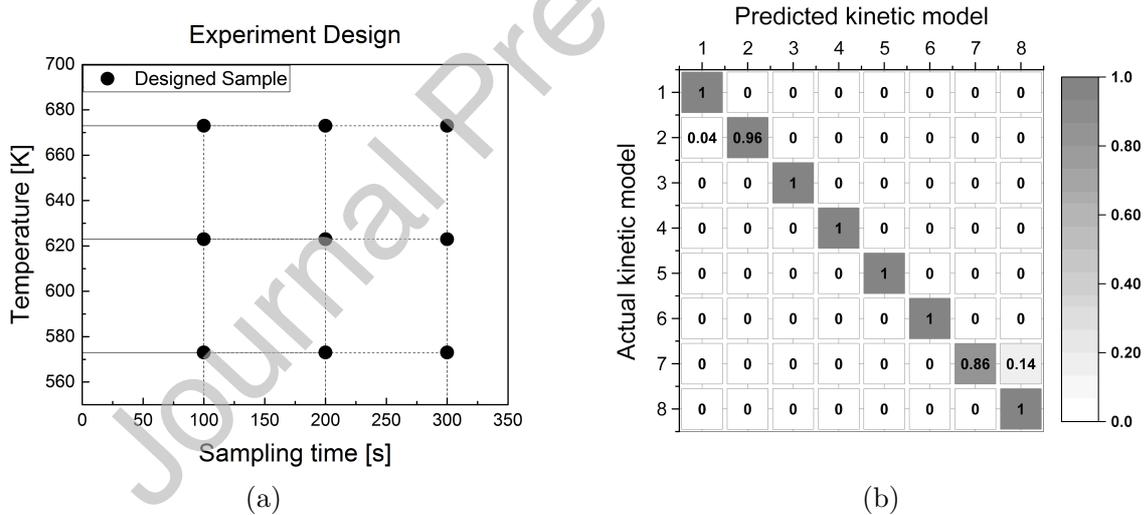


Figure 5: Case 1: (a) Adopted experiment design; (b) Confusion matrix Γ . In Case 1, the system noise is characterised by a $\sigma_r = 2.0$ and $\sigma_c = 0.04$. A test accuracy of 98.0% was achieved by the ANN classifier in Case 1.

zero (see Figure 5b, Figure 6b and Figure 7b). Nonetheless, it is observed that the confusion matrix in all cases is block diagonal, where the bottom-left and the top-right blocks are equal to the 4×4 null matrix. This observed patterns in the confusion matrix Γ indicate that models with $l = 1, 2, 3, 4$ are never misclassified as models with $l = 5, 6, 7, 8$ and vice-versa. In different words, the ANN classifiers never confound series with parallel reaction

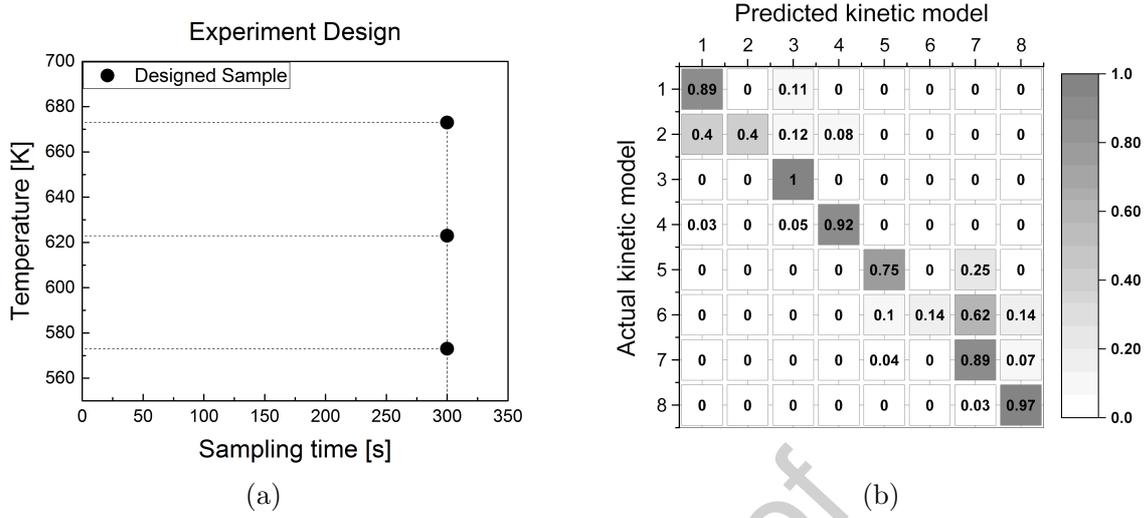


Figure 6: Case 2: (a) Considered experiment design; (b) Confusion matrix Γ . In Case 2, the system noise is characterised by a $\sigma_r = 2.0$ and $\sigma_c = 0.04$. A test accuracy of 75.0% was achieved in Case 2.

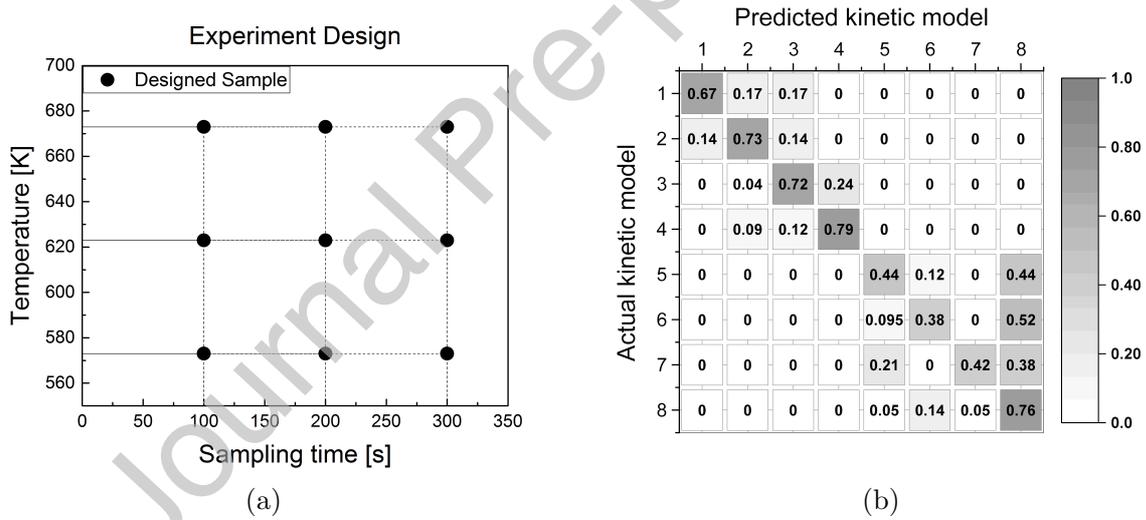


Figure 7: Case 3: (a) Adopted experiment design; (b) Confusion matrix Γ . In Case 3, the system noise is higher with respect to the noise in Case 1 and in Case 2 and it is characterised by $\sigma_r = 5.0$ and $\sigma_c = 10.0$. A test accuracy of 64.5% was achieved by the ANN classifier in Case 3.

mechanisms and vice-versa.

The highest test accuracy, i.e. 98.0%, was achieved in case 1 where the largest number of designed samples was assumed and a low level of system noise was present. Nonetheless, as one can see from Figure 5b, a fraction of kinetic models 2 was misclassified as model 1 and some model 7 elements were classified as model 8. The output $\mathbf{p} = [p_1, \dots, p_8]$ of the softmax

function, computed as in (2), is reported in Table 3 for the misclassified instances. For all the misclassified elements in the test set (indicated in Table 3 as α , β , γ and δ), the softmax output associated to the actual model type is comparable in magnitude to the largest p , which determines the predicted model type. As an example, element α in Table 3 is a model 2 element that is misclassified as model 1 with $p_1 = 0.48$ and $p_2 = 0.42$. Hence, a high score is assigned also to the actual model class, indicating that the ANN cannot distinguish the predicted and the actual model given the chosen experimental design and noise level in the system.

A principal component analysis (PCA) (Jackson, 2003) is presented in this section to further diagnose the reason for the misclassification in case 1. A multivariate latent variable model was used to visualise the 200 labelled inputs in the test set. Since in case 1 each labelled input is a 27×1 array of measurements, the multivariate model includes 27 principal components. The first and the second principal component, selected for visualisation purposes, account for 83.7% of the variance in the test set.

The scores in the latent space are shown in Figure 8 as circle-shaped symbols. More specifically, in Figure 8, hollow circles represent correctly classified models in the test set and solid dots represent the misclassified instances. Arrows in the centre of Figure 8 represent the projections of the 27 directions of the original input space on the two-dimensional latent space. As one can see from Figure 8, the 27 directions appear as clustered in three distinct groups. More specifically, the clustered directions are associated to measurements of component A (left), measurements of B (bottom-right) and measurements of C (top-right).

The misclassified elements on the bottom-right in Figure 8, i.e. β , γ and δ , represent model 7 elements that were classified as model 8. In the misclassified model 7 input arrays, the concentration of B is dominant in the samples, in fact the points lie in the bottom-right part of the latent variable space. Hence, in the misclassified model 7 instances, reaction $A \rightarrow B$ is much faster than reaction $A \rightarrow C$, i.e. $r_1 \gg r_2$, and the system tends to the single reaction behaviour. As one can see from Table 1, the difference between the structures of model 7 and model 8 is in the functional form of r_2 . Thus, when $r_1 \gg r_2$, model 7 and model 8 tend to become indistinguishable. In these limit conditions, the chosen experiment design may not be optimal for allowing a clear discrimination between the structures of model 7 and model 8.

The misclassified element on the left in Figure 8, i.e. α , represents a model 2 element that was misclassified as a model 1. Element α is aligned to the directions accounting for the measurements of A in the samples. This misclassified element represents a limit case where the concentration of A in the samples is high and it is associated to a model 2 instance where both reaction $A \rightarrow B$ and $B \rightarrow C$ are slow. Also in this case misclassification occurs at limit conditions and the selected experiment design may not be appropriate for discriminating between type 1 and type 2 kinetic model structures.

In addition to the previous diagnosis, one shall also observe that the misclassified elements in Figure 8 lie on the outer limit in the distribution of the input arrays. Neural networks are data driven models whose predictive capabilities are as good as the data used for their identification. The information available in the training data may not be sufficient for identifying an ANN capable to accurately classify limit cases. Nonetheless, it is observed that the identification of a robust ANN-based classifier for kinetic model selection relies *i)* on the selection of an appropriate experimental design *ii)* on the presence of low system

noise and *iii*) on the construction of a comprehensive in-silico training set.

6 Final remarks

In the presented case study, the number of parameters involved in the ANN-based classifiers is substantially higher than the number of parameters involved in the candidate kinetic models. Nonetheless, it is observed that the structure of the feed-forward ANN allows for the employment of efficient training algorithms for regularised regression, e.g. the Adam, AdaGrad and RMSProp (Geron, 2017). Conversely, the estimation of parameters in kinetic models represents a case specific problem that may pose substantial challenges to numerical routines for parameter estimation even in the presence of a small number of kinetic constants.

One shall also observe that the ANN-based classifier selects as plausible any model for which a feasible parameter set exists such that it could have generated the experimental observations. In contrast to other inference methods, e.g. model selection approaches based on Bayesian inference (Barber, 2011), the ANN does not penalise models involving irrelevant parameters. Nevertheless, the proposed framework represents an effective complementary tool to standard model discrimination criteria as it decouples the problem of model selection from the problems of identifiability associated with the candidate kinetic models. After a preliminary model selection is performed through an ANN-based approach, the scientist may focus on addressing only the identifiability problems associated with the plausible model structures. It is then possible to apply standard inference techniques, e.g. the Akaike or the Bayesian information criterion (Burnham and Anderson, 2002), to select the best available model penalising unnecessary complexity.

In the cases presented in this study, both the dataset generation step and the neural network identification step required only a few seconds of CPU time. It is recognised that the computational burden associated with the ANN identification and validation process will increase in the presence of a higher number of candidate models. Especially if models involve a substantial number of kinetic parameters, the in-silico data generation stage in the procedure may require the construction of a labelled dataset Ψ involving a substantial number N_ψ of labelled arrays to effectively cover high-dimensional parameter domains. Nonetheless, it is also observed that if the experimental design and the library of possible kinetic models are fixed, the illustrated procedure for ANN construction and identification must be performed only once. In fact, under such circumstance, the trained ANN can be employed to classify experimental data obtained from multiple reacting systems without repeating the ANN identification process.

The experimental designs considered in the case study were not optimised with the aim of discriminating among rival kinetic models. The integration in the proposed procedure of a step of optimal design of experiments for model discrimination will be the main objective of future research activities. As in standard frameworks for model discrimination (Buzzi-Ferraris et al., 1990), it is also expected that the discriminant power of optimally designed experiments and, consequently, the achievable ANN accuracy for a given experimental design is influenced by *1*) the explorable experimental design space (i.e. the degrees of freedom available to control the system), *2*) the observable state variables of the system, *3*) the number of candidate kinetic models and *4*) the prior knowledge on the feasible parameter

domains associated with the candidate models. Assessing the sensitivity of the ANN accuracy to a change in the aforementioned factors will also be object of future research.

Table 3: Case 1: Actual model class and Softmax output associated to the misclassified instances in the test set. Outputs marked with an asterisk indicate the highest score, which determines the predicted model type.

Misclassified instance	Actual model type	Softmax output							
		p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
α	2	0.48*	0.42	0.06	0.03	0.01	0.00	0.00	0.00
β	7	0.00	0.00	0.00	0.00	0.00	0.00	0.49	0.51*
γ	7	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.52*
δ	7	0.00	0.00	0.01	0.00	0.00	0.00	0.49	0.50*

7 Conclusion

In this work, a novel model selection framework is proposed where an Artificial Neural Network (ANN) is trained for selecting the most appropriate kinetic model given the available experimental data. The procedure starts with the formulation of a set of candidate model structures and the definition of an experimental design. The ANN is trained using experimental data generated in-silico. A dataset for ANN identification is constructed by simulating the designed experiments with all the candidate kinetic models, varying the values of their kinetic constants. In each simulated experiment, the values for the kinetic parameters are drawn from a distribution defined on the feasible parameter domain (i.e. the range of parameters that is compatible with the physical knowledge available on the system). Experiments are simulated a number of times to construct a sufficiently large dataset which covers all the possible kinetic model structures and the range of possible values for the kinetic constants. The dataset is then used for training an ANN-based classifier with the aim of matching the experimental data with the model structure used to generate the data.

The approach was demonstrated on a simulated case study on the selection of a kinetic model for a three-component reaction performed in a batch reactor. A number of candidate models were assumed both for series and parallel kinetic mechanisms. An optimised two-layer ANN achieved a 98% accuracy on the classification of unseen data (i.e. data not used in the training process). It was shown that the performance of ANN classifiers on the model recognition task relies on an appropriate experimental design and on the presence of contained measurement noise in the system. In the presented case study, a reduction in the number of designed samples or an increase in the system noise resulted in a reduction of the accuracy. Nonetheless, in all the cases the ANNs were able to distinguish between a series and a parallel kinetic mechanism.

The proposed model selection approach does not require the fitting of kinetic parameters. Therefore, it is well-suited for selecting a kinetic model even when available model

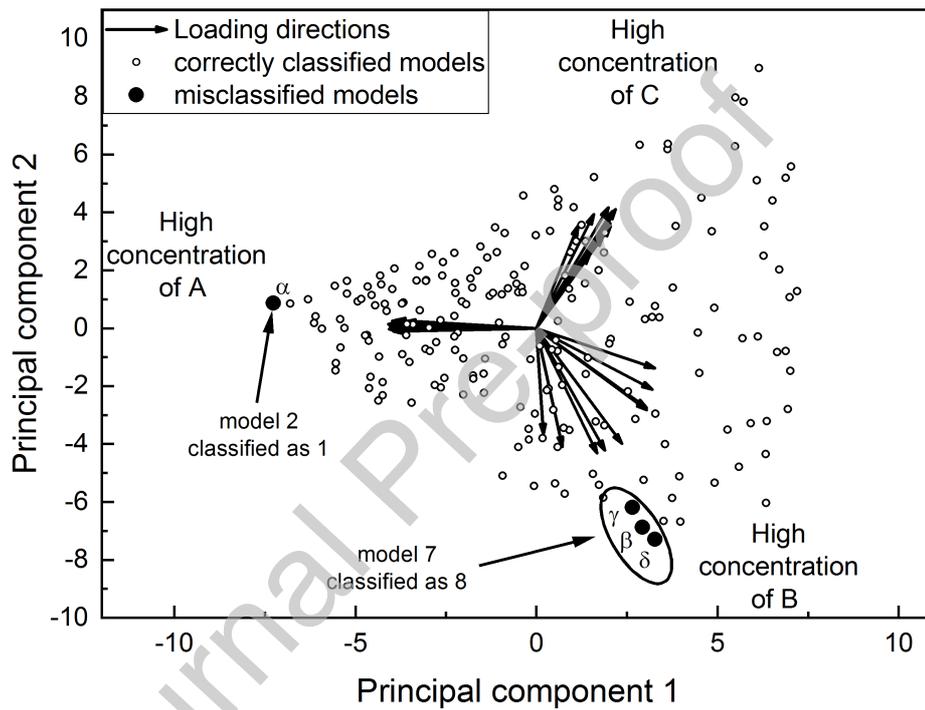


Figure 8: Case 1: projection of 200 27×1 labelled input arrays in the test set on the latent space defined by the two main principal components. Arrows represent the projection of the 27 original dimensions on the two-dimensional latent variable space. Hollow circles represent the scores for the correctly classified elements in the test set while solid dots represent the scores for the misclassified elements.

structures are affected by identifiability issues (i.e. when parameters in the candidate models cannot be estimated through regression because of a low sensitivity of model responses to a parameter change and/or an extreme parameter correlation). The primary objective of future works is to include an optimal design of experiment step in the proposed framework and test the approach on the recognition of kinetic models from real experimental data. The proposed framework will also be validated in the presence of more complex candidate kinetic mechanisms and broader reaction networks. Further extensions of the work will also focus on improving the robustness of the ANN classifiers in handling non-ideal scenarios. In particular, novelty detection approaches (Markou and Singh, 2003a,b) will be included in the framework to handle situations where none of the candidate kinetic models is appropriate and/or the coupling of two or more postulated reaction mechanisms is required to model the system behaviour.

Acknowledgements

This work was supported by the Hugh Walter Stern PhD Scholarship, University College London; the Department of Chemical Engineering, University College London.

Symbols used

Latin symbols

A_j	pre-exponential factor for the j -th reaction
b	bias parameter in single layer perceptron
b_h	bias parameter in the hidden layer of the ANN
b_o	bias parameter in the output layer of the ANN
C_i	concentration of species i
$E_{a,j}$	activation energy for the j -th reaction
k_j	kinetic constant of the j -th reaction
l	categorical variable
l_i	label associated to the i -th element in Ψ
\hat{l}_i	label prediction for the i -th element in Ψ
$\hat{n}_{i,j}$	j -th element in $\hat{\mathbf{n}}_i$
N	number of available samples
$N_{f,l}$	number of functions in the l -th kinetic model
N_h	number of neurons in the hidden layer of the ANN
N_m	number of candidate kinetic models
N_n	number of elements in ANN input array \mathbf{n}
N_r	number of chemical reactions
N_u	number of independent inputs in the system
$N_{x,l}$	number of state variables in the l -th kinetic model
N_y	number of output variables in the system
$N_{\theta,l}$	number of parameters in the l -th kinetic model

N_ψ	number of instances in the labelled dataset Ψ
p	output of single layer perceptron
p_k	k -th element in the ANN output \mathbf{p}
r_j	rate of the j -th reaction
R	ideal gas constant
t	time
t_s	set of sampling times
T	temperature
\mathcal{U}	random function
S_i	selectivity of i -th species in the mixture
$\hat{S}_{i,l}$	model l prediction for the selectivity of species i
X_i	conversion of i -th species in the mixture
$\hat{X}_{i,l}$	model l prediction for the conversion of species i

Greek symbols

α, β, \dots	misclassified instances in Ψ_{test}
γ_{jk}	jk -th element in $\mathbf{\Gamma}$
Θ_l	feasible parameter space associated to model l
ν_{ij}	stoichiometric coefficient of the i -th species in the j -th reaction
σ_{jk}	jk -th element of $\mathbf{\Sigma}$
σ_c	constant noise variance parameter
σ_r	relative noise variance parameter
φ_h	activation function in the hidden layer of the ANN
φ_o	activation function in the output layer of the ANN
Φ	experimental design space for a sample \mathbf{y}
Ψ	dataset of labelled input arrays
Ψ_{test}	dataset for ANN testing
$\Psi_{training}$	dataset for ANN training
$\Psi_{validation}$	dataset for ANN validation

Matrices and vectors

$\mathbf{0}$	null column array [v.u.]
$\mathbf{1}_h$	array whose entries are all equal to 1 [$N_h \times 1$]
$\mathbf{1}_o$	array whose entries are all equal to 1 [$N_m \times 1$]
\mathbf{f}_l	array of functions [$N_f, l \times 1$]
\mathbf{h}_l	array of functions [$N_y \times 1$]
\mathbf{n}	ANN inputs [$N_n \times 1$]
\mathbf{n}_i	i -th labelled input in Ψ [$N \cdot N_y \times 1$]
$\hat{\mathbf{n}}_i$	model l_i predictions for the experimental data [$N \cdot N_y \times 1$]
\mathbf{p}	output of ANN classifier [$N_m \times 1$]
\mathbf{u}	independent control variables (model inputs) [$N_u \times 1$]
\mathbf{w}	array of perceptron parameters [$N_n \times 1$]

\mathbf{W}_h	parameter matrix associated to the hidden layer in the ANN [$N_n \times N_h$]
\mathbf{W}_o	parameter matrix associated to the output layer in the ANN [$N_n \times N_h$]
\mathbf{x}_l	state variables in model l [$N_{x,l} \times 1$]
$\dot{\mathbf{x}}_l$	time derivative for state variables in model l [$N_{x,l} \times 1$]
\mathbf{y}	sample - array of measured system outputs [$N_y \times 1$]
\mathbf{y}_i	i -th sample [$N_y \times 1$]
$\hat{\mathbf{y}}_l$	model l prediction for the system outputs [$N_y \times 1$]
$\mathbf{\Gamma}$	confusion matrix associated to the ANN classifier [$N_m \times N_m$]
$\boldsymbol{\epsilon}$	random array [$N \cdot N_y \times 1$]
$\boldsymbol{\theta}_l$	array of parameters in model l [$N_{\theta,l} \times 1$]
$\boldsymbol{\Sigma}$	covariance matrix associated to \mathbf{n} [$N \cdot N_y \times N \cdot N_y$]
$\boldsymbol{\varphi}_j$	experimental conditions associated to sample \mathbf{y}_j [v.u.]

Acronyms

ANN	Artificial Neural Network
PCA	Principal Components Analysis

References

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (6), 716–723.
- Arbib, M. A., 2003. *The Handbook of Brain Theory and Neural Networks*, 2nd Edition. MIT Press, Cambridge, Mass.
- Asprey, S. P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. *Computers & Chemical Engineering* 24 (2), 1261–1267.
- Barber, D., 2011. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, Cambridge ; New York.
- Bard, Y., 1974. *Nonlinear Parameter Estimation*. Academic Press, New York.
- Bates, C. J., Yildirim, I., Tenenbaum, J. B., Battaglia, P., 2018. Modeling human intuitions about liquid flow with particle-based simulation. [arXiv:1809.01524](https://arxiv.org/abs/1809.01524) [cs, q-bio].
- Bloch, G., Denoeux, T., 2003. Neural networks for process control and optimization: Two industrial applications. *ISA Transactions* 42 (1), 39–51.
- Bonvin, D., Georgakis, C., Pantelides, C. C., Barolo, M., Grover, M. A., Rodrigues, D., Schneider, R., Dochain, D., 2016. Linking Models and Experiments. *Industrial & Engineering Chemistry Research* 55 (25), 6891–6903.
- Box, G. E. P., Draper, N. R., 1987. *Empirical model-building and response surfaces*. Wiley.
- Burnham, K. P., Anderson, D. R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd Edition. Springer-Verlag, New York.

- Buzzi-Ferraris, G., Forzatti, P., Paolo, C., 1990. An improved version of a sequential design criterion for discriminating among rival multiresponse models. *Chemical Engineering Science* 45 (2), 477–481.
- Chollet, F., et al., 2015. Keras. <https://keras.io>.
- Cobelli, C., Di Stefano, J. J., 1980. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *The American Journal of Physiology* 239 (1), R7–24.
- Coley, C. W., Jin, W., Rogers, L., Jamison, T. F., Jaakkola, T. S., Green, W. H., Barzilay, R., Jensen, K. F., 2019. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chemical Science* 10 (2), 370–377.
- Dua, V., 2010. A mixed-integer programming approach for optimal configuration of artificial neural networks. *Chemical Engineering Research and Design* 88 (1), 55–60.
- Dua, V., 2011. An Artificial Neural Network approximation based decomposition approach for parameter estimation of system of ordinary differential equations. *Computers & Chemical Engineering* 35 (3), 545–553.
- Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science* 63 (19), 4846–4872.
- Galvanin, F., Ballan, C. C., Barolo, M., Bezzo, F., 2013. A general model-based design of experiments approach to achieve practical identifiability of pharmacokinetic and pharmacodynamic models. *Journal of Pharmacokinetics and Pharmacodynamics* 40 (4), 451–467.
- Geron, A., 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly, Beijing Boston Farnham Sebastopol Tokyo.
- Goldsmith, B. R., Esterhuizen, J., Liu, J.-X., Bartel, C. J., Sutton, C., 2018. Machine learning for heterogeneous catalyst design and discovery. *AIChE Journal* 64 (7), 2311–2323.
- Himmelblau, D. M., 2008. Accounts of Experiences in the Application of Artificial Neural Networks in Chemical Engineering. *Industrial & Engineering Chemistry Research* 47 (16), 5782–5796.
- Hinton, G. E., Osindero, S., Teh, Y.-W., 2006. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation* 18 (7), 1527–1554.
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2 (5), 359–366.
- Hou, Z.-Y., Dai, Q., Wu, X.-Q., Chen, G.-T., 1997. Artificial neural network aided design of catalyst for propane ammoxidation. *Applied Catalysis A: General* 161 (1), 183–190.

- Hussain, M. A., Kershenbaum, L. S., 2000. Implementation of an Inverse-Model-Based Control Strategy Using Neural Networks on a Partially Simulated Exothermic Reactor. *Chemical Engineering Research and Design* 78 (2), 299–311.
- Jackson, J. E., 2003. *A User's Guide to Principal Components*. Wiley-Interscience, Hoboken, N.J.
- Jaderberg, M., Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition. arXiv:1406.2227 [cs].
- Janet, J. P., Chan, L., Kulik, H. J., 2018. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *The Journal of Physical Chemistry Letters* 9 (5), 1064–1071.
- Kayala, M. A., Baldi, P., 2012. ReactionPredictor: Prediction of Complex Chemical Reactions at the Mechanistic Level Using Machine Learning. *Journal of Chemical Information and Modeling* 52 (10), 2526–2540.
- Kingma, D. P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv:1412.6980 [cs].
- Kramer, M. A., 1991. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal* 37 (2), 233–243.
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira, F., Burges, C. J. C., Bottou, L., Weinberger, K. Q. (Eds.), *Advances in Neural Information Processing Systems* 25. Curran Associates, Inc., pp. 1097–1105.
- Le, T. A., Baydin, A. G., Zinkov, R., Wood, F., 2017. Using synthetic data to train neural networks is model-based reasoning. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. pp. 3514–3521.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444.
- Lee, J. H., Shin, J., Realff, M. J., 2018. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering* 114, 111–121.
- Lopes, A. L. C. V., Ribeiro, A. F. G., Reis, M. P. S., Silva, D. C. M., Portugal, I., Baptista, C. M. S. G., 2018. Distribution models for nitrophenols in a liquid-liquid system. *Chemical Engineering Science* 189, 266–276.
- MacKay, D. J. C., 1992. Bayesian methods for adaptive models. Ph.D. Thesis, California Institute of Technology.
- Markou, M., Singh, S., 2003a. Novelty detection: a review—part 1: statistical approaches. *Signal Processing* 83 (12), 2481–2497.

- Markou, M., Singh, S., 2003b. Novelty detection: a review—part 2:: neural network based approaches. *Signal Processing* 83 (12), 2499–2521.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. *Nature* 518 (7540), 529–533.
- Molga, E., Cherbański, R., Szpyrkowicz, L., 2006. Modeling of an Industrial Full-Scale Plant for Biological Treatment of Textile Wastewaters: Application of Neural Networks. *Industrial & Engineering Chemistry Research* 45 (3), 1039–1046.
- Olofsson, S., Hebing, L., Niedenführ, S., Deisenroth, M. P., Misener, R., 2019. GPdoemd: A Python package for design of experiments for model discrimination. *Computers & Chemical Engineering* 125, 54–70.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Petsagkourakis, P., Sandoval, I. O., Bradford, E., Zhang, D., del Rio-Chanona, E. A., 2019. Reinforcement Learning for Batch Bioprocess Optimization. arXiv:1904.07292 [cs, math].
- Prasad, V., Bequette, B. W., 2003. Nonlinear system identification and model reduction using artificial neural networks. *Computers & Chemical Engineering* 27 (12), 1741–1754.
- Python Core Team, 2018. Python: A dynamic, open source programming language. <https://www.python.org/>.
- Quadros, P. A., Reis, M. S., Baptista, C. M. S. G., 2005. Different Modeling Approaches for a Heterogeneous Liquid-Liquid Reaction Process.
- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25 (15), 1923–1929.
- Rengaswamy, R., Venkatasubramanian, V., 2000. A fast training neural network and its updation for incipient fault detection and diagnosis. *Computers & Chemical Engineering* 24 (2), 431–437.
- Rosenblatt, F., 1962. *Principles of Neurodynamics*. Spartan Book.
- Russell, S., Norvig, P., 2016. *Artificial Intelligence: A Modern Approach*, Global Edition, 3rd Edition. Pearson.
- Schwarz, G., 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6 (2), 461–464.

- Silvey, S. D., 1975. *Statistical Inference*. CRC Press.
- Suewatanakal, W., 1993. A comparison of fault detection and classification using ann with traditional methods. Ph.D. thesis, Ph. D. Dissertation, Univ. of Texas, Austin, TX.
- Traver, M. L., Atkinson, R. J., Atkinson, C. M., 1999. Neural Network-Based Diesel Engine Emissions Prediction Using In-Cylinder Combustion Pressure. SAE Technical Paper 1999-01-1532, SAE International, Warrendale, PA.
- Venkatasubramanian, V., 2019. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal* 65 (2), 466–478.
- Walter, E., Pronzato, L., 1997. *Identification of Parametric Models from Experimental Data*. Springer-Verlag, London.
- Wang, Y.-H., Li, Y., Yang, S.-L., Yang, L., 2005. Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach. *Journal of Chemical Information and Modeling* 45 (3), 750–757.
- Wei, J. N., Duvenaud, D., Aspuru-Guzik, A., 2016. Neural Networks for the Prediction of Organic Chemistry Reactions. *ACS Central Science* 2 (10), 725–732.
- Zhao, H., Lai, Z., Chen, Y., 2019. Global-and-local-structure-based neural network for fault detection. *Neural Networks* 118, 43–53.

Declaration of interests

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.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof