Hybrid modelling of a batch separation process

Ulderico Di Caprio[a], Min Wu[a], Furkan Elmaz[b], Yentl Wouters[c], Niels Vandervorst[c], Ali Anwar[b], Siegfried Mercelis[b], Steffen Waldherr[d], Peter Hellinckx[e], M. Enis Leblebici[a,]*

[a]Center for Industrial Process Technology, Department of Chemical Engineering, KU Leuven, Agoralaan Building B, 3590 Diepenbeek, Belgium
[b]IDLab, Faculty of Applied Engineering, University of Antwerp-imec, Sint-Pietersvliet 7, 2000 Antwerp, Belgium
[c]Janssen Pharmaceutica N.V., Chemical Process R&D, Turnhoutseweg 30, B-2340 Beerse, Belgium
[d]KU Leuven, Department of Chemical Engineering, Celestijnenlaan 200F-bus 2424, Leuven 3001, Belgium
[e]Faculty of Applied Engineering, University of Antwerp, Groenenborgerlaan 171, 2000 Antwerp, Belgium

Abstract

Applying machine learning (ML) techniques is a complex task when the data quality is poor. Integrating first-principle models and ML techniques, namely hybrid modelling significantly supports this task. This paper introduces a novel approach to developing a hybrid model for dynamic chemical systems. The case in analysis employs one first-principle structure and two ML-based predictors. Two training approaches (serial and parallel), two optimisers (particle swarm optimisation and differential evolution) and two ML functions (multivariate rational function and polynomial) are tested. The polynomial function trained with the differential evolution showed the most accurate and robust results. The training approach does not significantly affect the hybrid model accuracy. However, the main effect of the training approach is on the robustness of the parameter predictions. The coefficients of determination (R^2) on the test batches are above 0.95. In addition, it showed satisfactory extrapolation capabilities on different production scales with R^2>0.9.

Keywords: hybrid modelling, dynamic system, solvent switch, optimization, statistical modelling, data value

1. Introduction

In recent years, the process industry is gaining several benefits in the application of big data and machine learning (ML) techniques to model their processes (Bogojeski et al., 2021; Chiang et al., 2017; Ge et al., 2021; Mohd Ali et al., 2015; Qin and Chiang, 2019). These techniques use experimental data to develop data-driven models (or *statistical models*) to predict the physical phenomena described in the dataset used for the training. Despite the augmented prediction accuracy and the interest from the chemical industry, the application of ML modelling is still limited to specific industrial-scale processes because of a data quality problem. The data should respect the 5Vs rule to apply ML techniques (BBVA, 2020; Demchenko et al., 2014), namely:

1. *Volume* – higher amount of data allows for higher model accuracy.
2. *Variety* – an observed phenomenon should be described using different kinds of data.
3. *Velocity* - the data should be available in real-time to refine the parameters of the model.
4. *Veracity* – the uncertainty related to the data should be available and low.
5. *Value* – the information contained in the data should be helpful and not redundant.

The data obtained from chemical processes lacks Value and often Veracity. The process industry can benefit from the high amount of data coming from production processes. However, they usually have little use in building prediction models since 1) they cover only a narrow parameter space and 2) the

isolated effect of each process parameter on the final output is often not captured by the data (Reis et al., 2016). In addition, the information obtained from the plant is highly intercorrelated (McBride et al., 2020) and often suffers missed information because of communication and storage problems (Fisher et al., 2020).
The hybrid modelling approach is a solution to apply statistical modelling to cases with low availability and quality of data. It employs first-principle modelling and statistical modelling to make its predictions. The statistical model can be used either to refine the prediction of the first-principle model or to predict the parameters utilised by it to perform the predictions. It increases the overall accuracy and extrapolation capabilities of the model (McBride et al., 2020; Sansana et al., 2021; von Stosch et al., 2014). For this reason, the hybrid modelling overcomes the abovementioned issues related to the narrow investigation space that characterises data coming from a chemical process. It is well known that the statistical models can perform their predictions only within the boundaries within which the data are located. On the contrary, the hybrid models can extend their predictions range beyond the data boundaries since they can rely on the presence of a first-principle model. Moreover, the integration of the first-principle model allows the incorporation of the first-principle correlation between the variables before the model training (Glassey and von Stosch, 2020; von Stosch et al., 2014). This reduces the amount of data required to train the model. For these reasons, we hypothesise that using a hybrid model is a solution to Value and Veracity insufficiencies in chemical process data.

This work utilises the constant volume solvent switch (CSS) as a case study. It is one of the primary downstream operations required to produce active pharmaceutical ingredients (API). It aims to switch the reaction solvent to the crystallisation solvent allowing the purification of the API (Elgue et al., 2006; Papadakis et al., 2016). Developing a model with high prediction capabilities of the CSS operation by using the first-principle approach is challenging to execute. The system has a significant dynamic behaviour, making it too complex to accurately model using thermodynamic equations. In addition, the interaction of the API with the various solvents utilised in this process is hard to predict with high accuracy. As a matter of fact, the first-principle model used in this work returns poor prediction accuracy (Table 4).

This paper proposes a new methodology to develop hybrid models in the dynamic system running under uncertainties, focusing on the separation systems. We intend to overcome the Value and Veracity issues mentioned above. The hybrid model utilised in this paper is composed of one first-principle model and two statistical models. Eight combinations of statistical modelling techniques have been investigated in this work. These are the combinations of two statistical functions, two optimisers, and two training strategies. The statistical functions are namely multivariate rational function (MRF) and polynomial. The optimisers are particle swarm optimisation (PSO) and differential evolution (DE). The training of the models was performed using parallel and serial strategies.
The usage of MRF has been briefly explored in chemical engineering for the approximation of functions not described by first-principle models (Willis and von Stosch, 2017), while it has shown excellent capabilities in other fields (Austin et al., 2021; Hong et al., 2018). The DE and PSO algorithms have already been applied to chemical engineering problems to perform both parameter identification and process optimisation (Aguitoni et al., 2018; Angira and Santosh, 2007; Biazi et al., 2020; Dragoi and Curteanu, 2016; Mariano et al., 2011; Schwaab et al., 2008). They are valid alternatives to the traditional gradient-based optimisation approaches due to their ability to escape from the local minimum and converge to the global optimum in the search area (Dragoi and Curteanu, 2016).
We hypothesise that combining these powerful techniques leads to a hybrid model with high prediction accuracy and robustness. This way, we resolve the Value and Veracity issues with high extrapolation capabilities. The new methodology relies on the capability of MRF and polynomial to predict and handle non-linear behaviours and the augmented quality of the parameter tuning of the

differential evolution approach compared to the traditional gradient-based approaches. This new hybrid model methodology focuses on batch and fed-batch operations. However, it applies to all the separations and reaction processes. The model developed with this approach is validated through data from an industrial plant provided by a pharmaceutical partner company.

2. Material and methods

2.1 The process

The data used in this study were obtained from a process in development. The process aims to swap the solvent from dichloromethane (DCM) to acetone (Ac). A low concentration (below 10% mol) of methanol (MeOH) is present within the reboiler at the beginning of the operation. Figure 1 shows the schematic representation of the equipment used for this operation.

The CSS operation is divided into two phases according to the optimisation strategy suggested in the literature (Elgue et al., 2006). These are, namely, the *first evaporation* and *constant volume* phases. During the *first evaporation* step, the reaction solvent (DCM in this study) is evaporated to a minimum volume optimised during the design phase. In the *constant volume* phase, the crystallisation solvent (Ac) is fed while the evaporation of the solvent blend (DCM+Ac) takes place at the designed minimum volume.

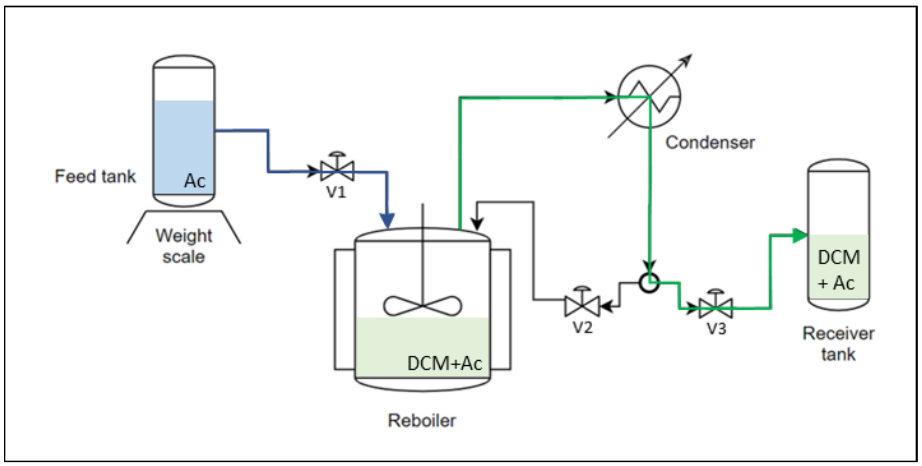

Figure 1: Schematic representation of the solvent switch process utilised as a case study in this work. At the initial state, the reboiler contains only dichloromethane (DCM). The acetone (Ac) is contained in a tank and fed to the reboiler. During the solvent switch operation, Ac and DCM evaporate from the reboiler while the Ac is fed to keep the liquid volume within the reboiler constant. This increases the concentration of Ac in the solution within the reboiler and decreases the DCM concentration.

2.2 Data description and input-output of the final model

Three batches are utilised in this study. All the batches were executed in the same reactor, with the same equipment and following the same control strategy. In addition, a supplementary batch was utilised to investigate the extrapolation capabilities of the model. This batch data was used following the training phase to test the extrapolation abilities of the model. This extrapolation batch was executed with half the volume compared to the training batches. A list of all the process variables contained in the datasets and utilised to develop the model is available in Supplementary Information (SI) section 1.2. The model utilises as input the flowrate of Ac, the temperature of the heat transfer fluid within the jacket, the reflux direction and the stirring speed. It predicts all the other variables contained in the dataset; however, the main ones are the temperature inside the reboiler, the volume inside the reboiler and the evaporated amount of solvent leaving the reboiler.

2.3 The first-principle model

The first-principle model contains material balance, phase equilibrium equations and heat balance. It was built with the following assumptions: 1) Thermodynamic equilibrium between liquid and vapour, 2) The pressure of the system is constant during the operation, 3) The viscous dissipation heating due to the stirring is negligible, 4) The fluid within the reboiler is perfectly mixed, 5) The condenser is ideal (all the vapour entering into it is condensate without any vapour-liquid equilibria), 6) No chemical reactions take place during the process, 7) Inert gas is not dissolved within the liquid, 8) The gas phase behaves as ideal. A detailed description of the first-principle model is available in SI section 1.3.
The energy balance was written by considering the evaporation of the solvent blend driven by the service fluid. The overall heat-balance equation is reported in (1):

$$\frac{d(M_L * h_L + M_V * h_V)}{dt} = \\ = F_{in} * h_F + R * F_{ref} * h_{ref} - F_{out} * h_V + F_{gas}^{inert} * h_{gas} + U * A * (T_{service} - T) \quad (1)$$

where M_L $[mol]$ is the molar amount of liquid within the reboiler, $h_L[J * mol^{-1}]$ is the liquid enthalpy, M_V $[mol]$ is the molar amount of vapour within the reboiler, $h_V[J * mol^{-1}]$ is the liquid enthalpy, $F_{in}[mol * s^{-1}]$ is the molar flowrate of crystallisation solvent flowing to the reboiler, h_F $[J * mol^{-1}]$ is the enthalpy of the crystallisation solvent flowing to the reboiler, $R[-]$ is the state of the reflux (0 or 1), $F_{ref}[mol * s^{-1}]$ is the flowrate of the reflux stream, $h_{ref}[J * mol^{-1}]$ is the enthalpy of the reflux liquid stream, F_{out} $[mol * s^{-1}]$ is the flowrate of vapour leaving the reboiler, $F_{gas}^{inert}[mol * s^{-1}]$ is the flowrate of inter gas flowing into the system, $h_{gas}[J * mol^{-1}]$ is the enthalpy of the inert gas, U $[J * s^{-1} * K^{-1} * m^{-2}]$ is the overall heat transfer coefficient between the liquid within the reboiler and the heat transfer fluid, A $[m^2]$ is the contact area between the liquid within the reboiler and the heat transfer fluid, $T_{service}[K]$ is the temperature of the heat transfer fluid, T $[K]$ is the temperature of the fluids (liquid and vapour) within the reboiler.
The heat-transfer coefficient from the jacket to the reboiler (U) was modelled utilising the heat resistance chain model (Bird et al., 2006). For the study case, the wall is composed of two solid layers (steel/glass enamel) and the convection layer inside the reboiler; the resistance related to the convection layer jacket-side was considered negligible. The convection heat transfer coefficient inside the reboiler was calculated by using the Nusselt equation correlation (2).

$$Nu = \frac{h * D}{k} = f(Re, Pr, Vi) = k_{UA} * Re^{\alpha} * Pr^{\beta} * Vi^{\gamma} \quad (2)$$

where Nu $[-]$ is the Nusselt number, h $[J * s^{-1} * K^{-1} * m^{-2}]$ is the thermal convection coefficient inside the reboiler, D $[m]$ is the characteristic dimension of the system, k $[W * m^{-1} * K^{-1}]$ is the thermal conductivity of the liquid inside the reboiler, $k_{UA}[-]$, $\alpha[-]$, $\beta[-]$ and $\gamma[-]$ are case-dependent parameters, $Re[-]$ is the Reynolds number, $Pr[-]$ is the Prandtl number, and $Vi[-]$ is the ratio between the viscosity of the liquid on the wall of the reboiler and in the liquid bulk.
To compute (2) four case-dependent parameters should be identified. In the literature, the following parameter values are suggested $k \in [0.3,1.5], \alpha = 0.66,\ \beta = 0.33, \gamma = 0.14$ (Mohan et al., 1992).

2.4 The structures of the statistical models

Applying a first-principle model with the parameter optimised for the analysis case returned poor prediction accuracy and low generalisation capabilities among the batches. Two main deviating parameters were identified from the sensitivity analysis on the parameters of the first-principle model. Namely, the heat transfer coefficient (3) and the liquid enthalpy (4). Two statistical functions were employed to correct the identified parameters and model their deviations:

$$k_{UA}^{HM} = k_{UA}^{IDEAL} + k_{UA}^{dev} \quad (3)$$

$$h_L^{HM} = h_L^{IDEAL} + h_L^{dev} \quad (4)$$

where k_{UA}^{HM} $[-]$ is the corrected proportional value in the Nusselt equation (2), $k_{UA}^{IDEAL}[-]$ is the ideal proportional value in the Nusselt equation, $k_{UA}^{dev}[-]$ is the correction function for the proportional factor in the Nusselt equation, $h_L^{HM}[J * mol^{-1}]$ is the corrected value of the liquid enthalpy, $h_L^{IDEAL}[J * mol^{-1}]$] is the ideal value of the liquid enthalpy, $h_L^{dev}[J * mol^{-1}]$ is the correction function for the liquid enthalpy.

We hypothesise that the heat transfer coefficient deviates because of the effect of the liquid volume change on the mixing efficiency. For this reason, k_{UA} was computed as a function of the liquid volume inside the reboiler.

The deviation of the liquid enthalpy from the first-principle model was computed as a function of the concentration of Ac and the temperature of the liquid within the reboiler. In this case, the h_{dev}^{HM} is not only the enthalpy of the mixture containing Ac, DCM and MeOH. However, it includes 1) the enthalpy excess of the mixture Ac-DCM, 2) the interaction between the solvents and the API, and 3) the deviation from the ideality of the model.

MRF and polynomial functions were employed to model the abovementioned deviations. The explicit expressions of the functional forms are reported in Table 1. In the equation contained in this table, $V[m^3]$ is the volume of liquid within the reboiler, $T\ [K]$ is the temperature within the reboiler, $x_{Ac}[mol * mol^{-1}]$ is the molar-fraction of Ac within the liquid contained in the reboiler, $k_{UA}^{IDEAL}, A_1, B_1, A_2, B_2$ are the fitting coefficients identified from the optimiser

Table 1: Explicit form of the functional structure used to perform the black-box function estimation.

	MRF	Polynomial
k_{UA}^{HM}	$k_{UA}^{IDEAL} + \frac{A_1 * V}{1 + B_1 * V}$	$k_{UA}^{IDEAL} + A_1 * V + B_1 * V^2$
h_L^{dev}	$\frac{A_1 * x_{Ac} + A_2 * x_{Ac} * T}{1 + B_1 * x_{Ac} + B_2 * x_{Ac} * T}$	$A_1 * x_{Ac} + A_2 * x_{Ac} * T + B_1 * x_{Ac}^2 + B_2 * x_{Ac}^2 * T^2$

2.5 The search algorithms

Differential evolution (DE) (Storn and Price, 1997) and particle swarm optimisation (PSO) (Kennedy and Eberhart, 1995) were employed to identify the coefficients of the black-box models. Both the algorithms were trained with 20 particles initially distributed on the coefficient space using a Latin hypercube sampling strategy. 30 iterations were used to perform the training of the models. Each search was repeated 10 times with different positions of the initial population. A more detailed description of the search algorithms is in the SI section 1.4. For the DE, the best/2/bin was used as mutation strategy with a mutation coefficient F=0.5. The cross-over probability factor was set to be 0.7. For the PSO, the inertia coefficient $\omega = 0.85$, the individual coefficient $c_p = 2$ and social coefficient $c_g = 1$.

2.6 The training approaches

Two different training approaches were used to train the coefficients of the two statistical models (k_{UA}^{HM} and h_L^{dev}) with the different optimisers. The statistical models were trained simultaneously (Parallel approach) or consecutively (Serial approach). The serial approach was possible due to the structure of the statistical models contained in the hybrid model. The h_{dev}^{HM} model does not have any impact during the *first evaporation* (see Section 2.1) since the concentration of Ac in this phase is zero. For this reason, the k_{UA}^{HM} model could be trained separately on this first step of the process. Once its coefficients were identified, the h_{dev}^{HM} model was trained on the *constant volume* step of CSS (see Section 2.1). The authors hypothesise that the performance of the training approaches will change with the functional form and the optimiser employed during the training. In addition, the authors hypothesise that the serial approach will return models with higher prediction accuracy and

robustness. Eight different combinations of training approaches, optimisers and functional forms were tested and compared (Table 2).

Table 2: Abbreviation of the eight combinations of training approach, optimiser and functional form tested and compared in this work

Abbreviation	***Approach***	***Optimiser***	***Functional form***
PSO/MRF/Parallel	Parallel	PSO	MRF
PSO/poly/Parallel	Parallel	PSO	Polynomial
DE/MRF/Parallel	Parallel	DE	MRF
DE/poly/Parallel	Parallel	DE	Polynomial
PSO/MRF/Serial	Serial	PSO	MRF
PSO/poly/Serial	Serial	PSO	Polynomial
DE/MRF/Serial	Serial	DE	MRF
DE/poly/Serial	Serial	DE	Polynomial

The training of the statistical models was performed utilising two batches; the fitness functions were calculated for each of them, and then the two values were averaged and fed back to the optimiser. The training loop and the variables used in it are shown in Figure 2.

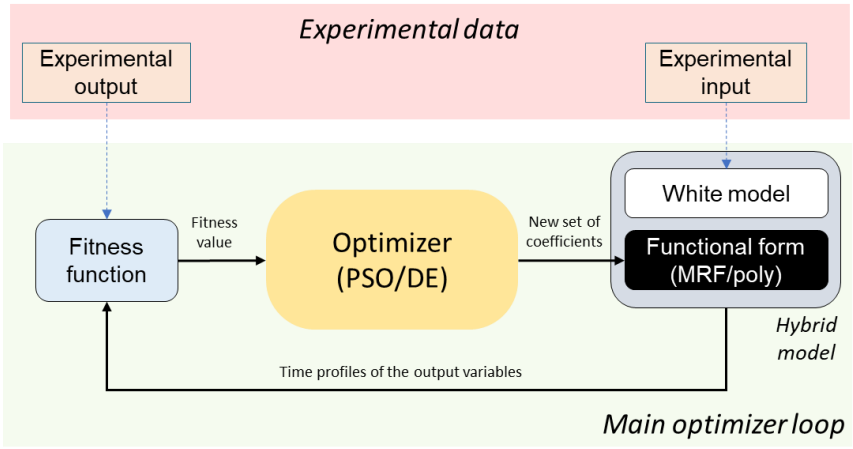

Figure 2: Hybrid model training structure. The coefficients set proposed by the optimiser is used to run the simulation driven by experimental input. The model returns the estimated time profile of the output variables. These are compared with the experimental output from the plant, and the fitness function is calculated using these deviations. The calculated value of the fitness function is given to the optimiser, which will propose a new set of coefficients to test.

2.7 Fitness function

The fitness function was composed of the most representative and stable variables of the process. Their experimental values were compared to the model predictions at each time step.

In case of premature interruption of the DAE resolution, the fitness function returns 10^{14}; on the contrary, the equation (5) is applied to compute its value.

$$C_{fitness}(\boldsymbol{\phi}) = C_{TempReb}(\boldsymbol{\phi}) + C_{RecTank}(\boldsymbol{\phi}) + 1e10 * C_{Time}(\boldsymbol{\phi}) \quad (5)$$

where $\boldsymbol{\phi}$ is the vector associated with the statistical model coefficients, $C_{TempReb}(\boldsymbol{\phi})$ is the fitness function part associated with predicting the temperature inside the reboiler; it is computed as the mean squared error (MSE) of the temperature. $C_{RecTank}(\boldsymbol{\phi})$ is the fitness function part associated with predicting the volume inside the receiver tank. This variable has a high uncertainty due to process-related unquantified losses (e.g. sensor noise, vent gas purge, sloshing); $C_{RecTank}(\boldsymbol{\phi})$ is thus computed taking into account these uncertainties as described in SI section 1.5. $C_{Time}(\boldsymbol{\phi})$ is fitness

function part associated with the prediction of the end time of the process. A more detailed description of the cost function is available in SI section 1.5.

2.8 Definition of the success criteria

The prediction accuracy of the hybrid models was evaluated utilising the fitness function values returned from each training run. Since each model was trained 10 times, the median and the variance of the fitness function values were employed as descriptors for each training group. Lower values of the variance indicate higher robustness of the training technique to the initial conditions. Lower values of the median indicate the better prediction accuracy of the technique.

A twofold approach was applied to evaluate the quality of the black-box models that drive the parameters of the first-principle model. The quality of the models that define k_{UA}^{HM} is evaluated utilising both the physical boundaries suggested in the literature (Mohan et al., 1992) and a *reference trend* proposed by the DE/poly/Serial.

The h_{dev}^{HM} statistical model was evaluated only utilising the trend proposed by DE/poly/Serial. It was used as a reference for two main reasons: 1) in the preliminary tests, it returned both the more robust fitness score median and variance in moving from the training set to the test set and 2) the statistical models estimated by this technique are less sensitive to the initial population distribution.

The index of dispersion (IoD) (6) was utilised to assess the quality of the black-box models returned by the various techniques and training approach

$$IoD = \frac{\sigma^2}{\mu} \tag{6}$$

where σ is the standard deviation of the training group and μ is the average of the training group.

For each training group, the IoD value was calculated at fixed independent variables of the black-box models. The average and maximum values of this descriptor are used as indicators: the lower their values, the more robust the technique with the initial population.

3. Results and discussion

3.1 Parallel training

3.1.1 Comparison of the complete hybrid model prediction in the parallel training

In the parallel training, the hybrid models based on the polynomial structure show higher accuracy and robustness than the MRF (Figure 3a). The presence of the denominator in the MRF structure increases the non-linearity of the coefficient search and narrows down the validity domain of the functions. It led to some runs to numerical instabilities or premature interruption of the hybrid model on the test set. In the polynomial form, the absence of the denominator increases the range of the solution, decreases the non-linearity of the search and reduces the instabilities of the system. All these factors allow the polynomial-based models to have higher accuracy and robustness than the MRF-based models (Figure 3).

The optimiser in the parallel training has a marginal effect on the overall hybrid model performances. In the literature, the DE algorithm is well-known to be more robust and performant in the coefficient search than the PSO (Civicioglu and sBesdok, 2013; Vesterstrom and Thomsen, 2004; Zhang and Wei, 2014). However, in this case, its results do not differ so much from the one of the PSO. We hypothesise that the robustness of the DE is not enough to stabilise the solution proposed by the MRF statistical functions for the case study.

The models obtained using DE/poly/Parallel return a lower median value of the fitness function on the training and test sets than the models trained with the PSO/poly/Parallel. This is possible because the DE utilises a more complex search procedure than the PSO. The DE has the possibility to both crossover and mutate the members of the population, facilitating the search and allowing escape from local

minima (exploration capabilities). On the other hand, the PSO has a simpler particle selection algorithm based only on the position of the particle and its score. It decreases the possibility of escaping from local minima (Stacey et al., 2003). Therefore, the search performed by the PSO is more inclined to examine the convexity in which the current best particle is allocated. This behaviour makes the search executed with the DE more inclined to overfit the training set when the MRF is employed as the statistical function within the hybrid model. On the contrary, when the polynomial functional form is employed as statistical function, this behaviour increases the prediction capabilities of the obtained hybrid model.

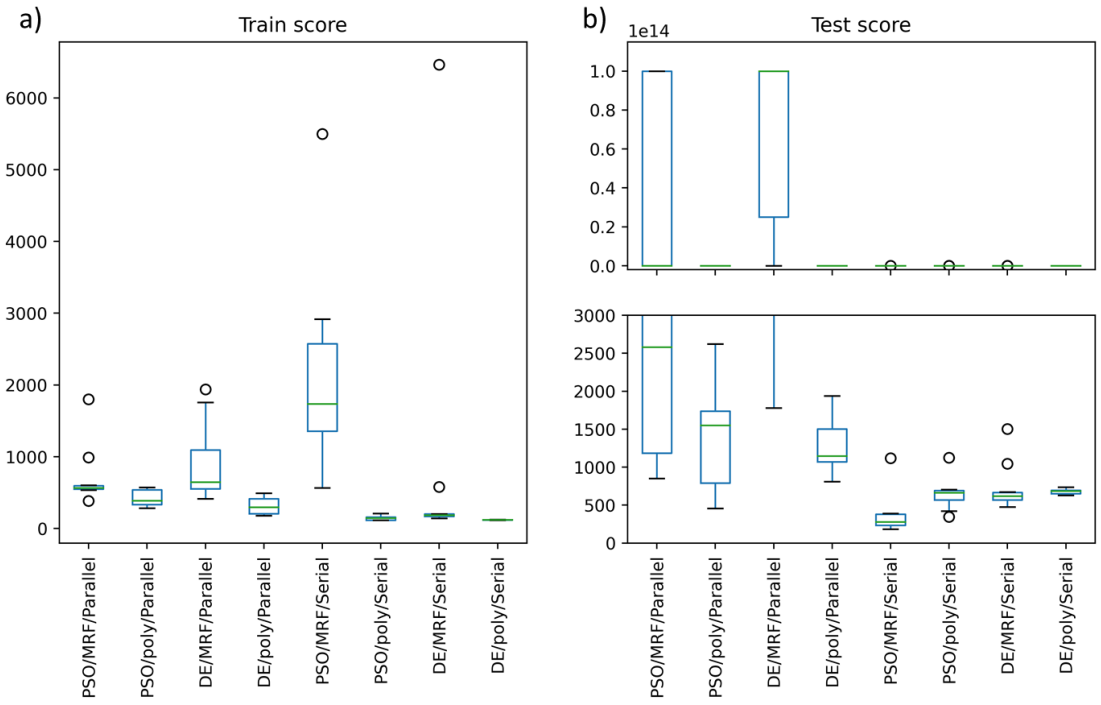

Figure 3: a) Boxplot of the fitness function for different approaches, optimisers and functions obtained on the train set. b) Boxplot of the fitness score for different approaches, optimisers and functions obtained on the test set. In these plots, numerical instabilities are detectable when the fitness function values are equal to 10^{14}. For the plot reporting the scores on the test set, the overall view is reported at the top, and the zoom of the lower fitness functions area is on the bottom. The polynomial function and the serial approach are the better performing techniques in this study. They return the lower median values and the lower variance on the test set. No significant differences are detectable in changing the optimiser. However, the usage of DE slightly increases the model accuracy and robustness.

3.1.2 Comparison of the black-box model parameters prediction in the parallel training

From Figure 4 and Table 3, one can observe that the polynomial functions show higher robustness with the initial condition and higher identification accuracy than the MRF when predicting k_{UA}^{HM}. The MRF structure shows numerical instabilities. The role of the optimiser is negligible when the MRFs are used as the statistical functional form. On the contrary, when the polynomial structure is used, the usage of DE increases the robustness of the predictions (Table 3) and their accuracies.

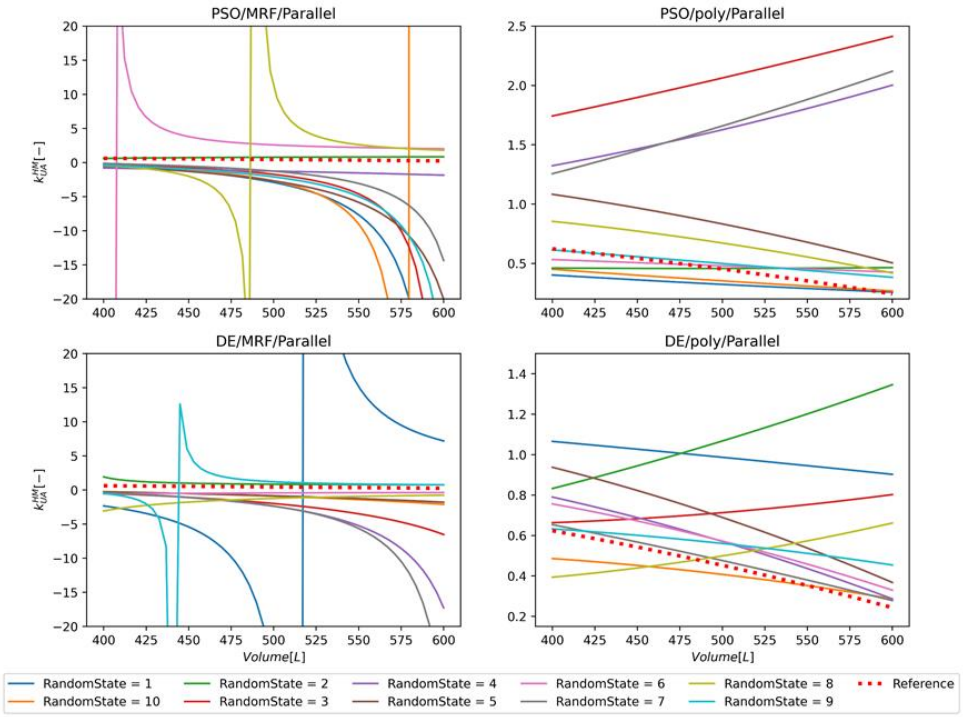

Figure 4: Profile of the k_{UA}^{HM} predicted with the parallel approach with the various combination of optimiser and black-box structure. MRF return poor prediction capabilities and low robustness since the numerical instabilities and wrong trend. The usage of the polynomial function returns better prediction capabilities than the MRF. In this case, when DE is used as the optimiser, the parameter estimation shows higher robustness with the initial population.

From Figure 5, one can observe the lower prediction accuracy and robustness of the MRF structure. It returns highly scattered solutions with different trends without any physical meaning, such as k_{UA} values lower than 0. The solutions delivered by the PSO/MRF/Parallel do not follow the trends of the reference for both k_{UA}^{HM} and h_{dev}^{HM}, furthermore the predicted values of k_{UA}^{HM} are outside the acceptable physical boundaries, or they show negative values. Despite these deficiencies, the fitness score of the PSO/MRF/Parallel on the training set is still low both in variance and median since the technique overfits on the training set. The overfitting is due to the synergistic training of the two models (parallel). To compensate for the deviation brought from one black-box model (e.g. k_{UA}^{HM}), the other (e.g. h_{dev}^{HM}) should adapt its values to reduce the value of the fitness function. The search carried out by the PSO is more affected by this synergistic effect as none of the solutions proposed by the PSO follows the reference trend, and most of the k_{UA}^{HM}values are negative. The usage of the DE optimiser slightly mitigates the effect of the non-linearity brought by the MRF, yet the majority of the solutions proposed by the DE/MRF/Parallel do not follow the reference. Only the models associated with the random state 8 return a solution that follows the reference trend. This unreliability of MRF trained with parallel training limits its usage and, in this case, prohibits it.

The usage of the polynomial increases the accuracy and robustness of the solution. When the polynomial structure is employed, both the optimisers return functions that follow the reference. However, the DE is more robust to the initial population than PSO because of the lower scattered profiles of the black-box estimated parameters. Using the polynomial structure drastically increases the capabilities to identify the coefficients of the black-box model. It allows the domain of the black-box models to be $\mathbb{R}$. Thus, the majority of the solutions proposed by the PSO are now able to follow the reference trends in k_{UA}^{HM} values except for the solution associated with random states 3, 4 and 7 (Figure 4). This also affects the quality of the models that describe h_{dev}^{HM} that now have the same trend of the reference. However, the PSO/poly/Parallel approach returns more scattered solutions than DE/poly/Parallel. In Table 3, the average and the maximum values of the IoD related to the

PSO/poly/Parallel technique are higher than the DE/poly/Parallel. This behaviour is due to the increased robustness of the DE algorithm to the initial population compared to the PSO optimiser (Vesterstrom and Thomsen, 2004). All the black-box models obtained with the DE follow the trend of the reference, and all the values of k_{UA}^{HM} are within the literature boundaries.

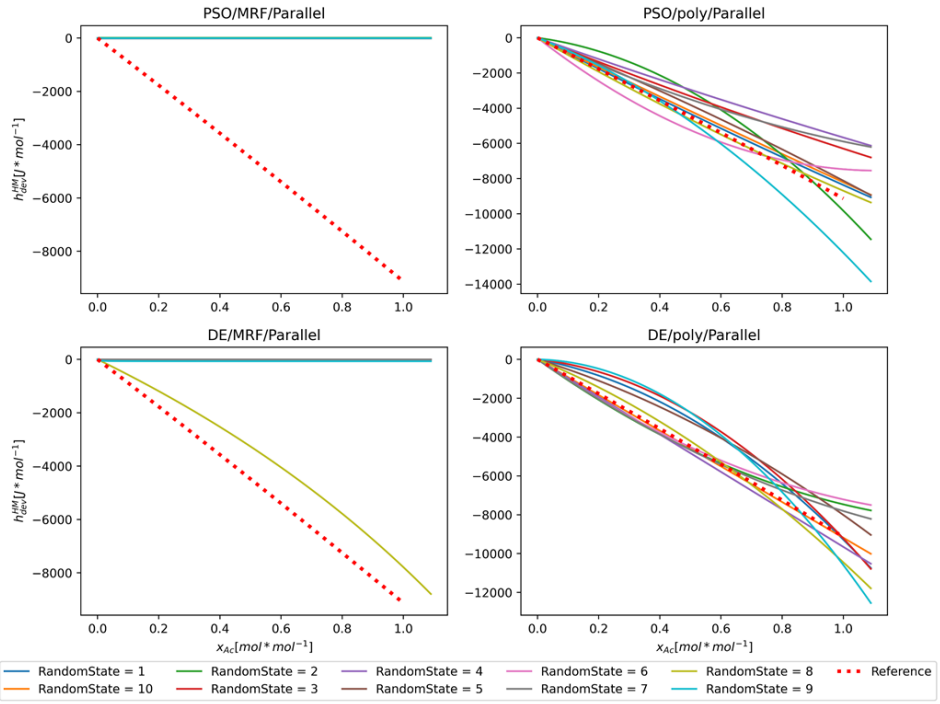

Figure 5: Profiles of the h_{dev}^{HM} predicted using the parallel approach with the various combination of optimiser and black-box structure calculated at 50°C. The polynomial functions show higher robustness and accuracy than the MRF. The models obtained with the MRF show numerical instabilities and poor accuracy regardless of the used optimiser. The polynomial function shows higher prediction accuracy and robustness with the random state with both the optimisers.

The superior prediction accuracies and robustness offered by the DE/poly compared to the other techniques is related to a combination of several factors: 1) the DE search is very robust to the variations in the initial population; 2) the DE has better capabilities in finding the global minimum compared to the PSO (Vesterstrom and Thomsen, 2004) and 3) the coefficient search for the polynomial structure is linear and less complex to perform compared to MRF.

Table 3: The IoD average and max absolute value for the various techniques and training approaches. The calculation is made at the same independent variable values for the different random states

	Parallel training				*Serial training*			
	k_{UA}^{HM}		h_L^{dev}		k_{UA}^{HM}		h_L^{dev}	
Technique	*Average*	*Max*	*Average*	*Max*	*Average*	*Max*	*Average*	*Max*
PSO/MRF	636.77	2.87e4	7.01	7.23	1.39e-2	2.80e-2	449.02	687.13
PSO/poly	0.43	0.74	194.49	592.99	2.96e-3	4.24e-3	86.58	217.71
DE/MRF	3.15	566.66	3.39e3	7.78e3	2.01e-2	3.47e-2	466.34	1328.65
DE/poly	8.57e-2	1.98e-1	153.75	266.02	6.16e-4	1.95e-3	5.24	7.14

3.2 Serial training

3.2.1 Comparison of the hybrid model prediction in the serial training

Due to mathematical instabilities, none of the hybrid models trained with the serial training approach show any premature interruption on the test set (Figure 3). The PSO/MRF/Serial returns a higher median fitness score on the training set than the PSO/poly/Serial, thus a lower accuracy on the training

set. However, on the test set, the result is the opposite favouring the MRF from the point of accuracy. This is related to the synergic effect of the lower exploration capabilities of the PSO technique and the non-linearity in the coefficient search provided by the MRF structure. These two factors do not allow the PSO/MRF/Serial to overfit the training set and have a better score on the test set compared to the PSO/poly/Serial.

The functional form marginalises when the DE is used as an optimiser. DE/MRF/Serial and the DE/poly/Serial return comparable variance and median fitness scores. This is linked to the robustness of the DE technique that mitigates the non-linear effects in the coefficient search of the MRF and lets it behave as a polynomial.

3.2.2 Comparison of the black-box model coefficient prediction in the serial training

All the techniques, when trained with the serial approach, return predicted k_{UA}^{HM} values within the boundaries proposed in the literature. The solutions obtained with the PSO technique are less accurate and robust than those proposed by the DE (Figure 6 and Table 3). All the solutions of h_{dev}^{HM} proposed from this training approach return values that follows the reference trend (Figure 7).

The MRF shows lower robustness compared to the polynomial. In Table 3, their IoD is one order of magnitude higher than the one associated with the polynomial. This is linked to the linearity of the coefficient search in the polynomial function; it makes the coefficients search less complex to perform. The usage of DE has higher robustness than the PSO because of the more complex algorithm that the DE uses to find the solution.

Like the k_{UA}^{HM}, also for h_{dev}^{HM} the models proposed by the PSO are more scattered than the solutions proposed by the DE when using the same black-box form. In addition, using the same optimiser, the polynomial functions show fewer scattered solutions than the MRF (Table 3). The serial training results in the lower median and variance fitness function in the test set values when the solution is searched with the PSO/MRF. However, this technique is not optimal for predicting the coefficient black-box model where the DE/poly returns the best coefficient search.

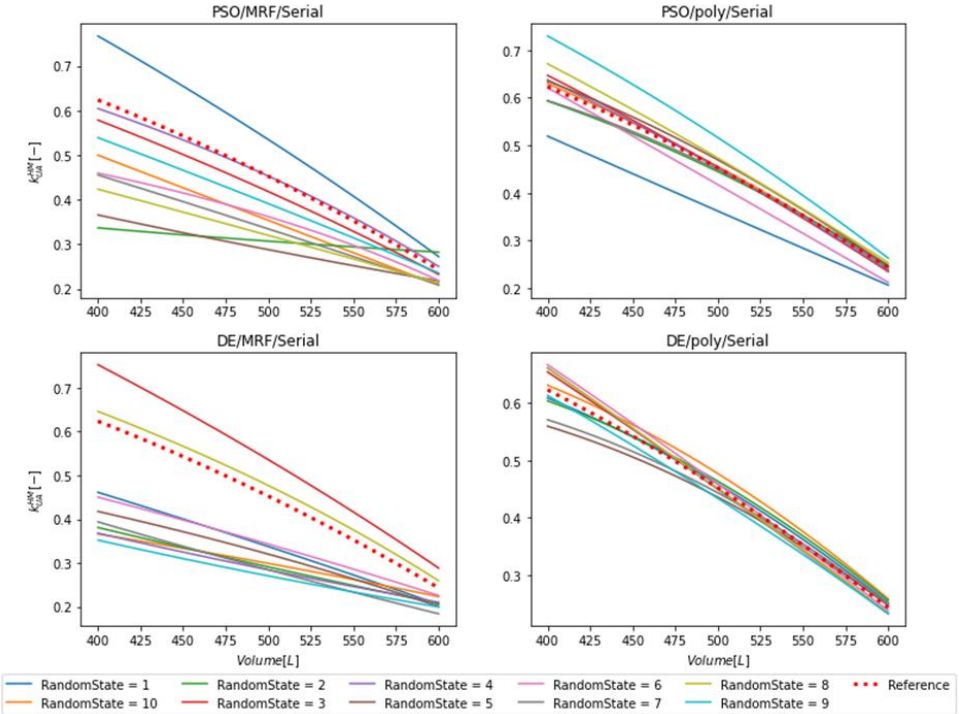

Figure 6: Predicted profile of k_{UA}^{HM} obtained with the serial approach with various combinations of optimisers and black-box structure. All the solutions are within the acceptable range and follow the correct trend of the predicted variables. However,

the polynomial show higher robustness than the MRF since the lower scattering of the solutions. The DE report higher robustness to the initial condition in the optimisation than the PSO.

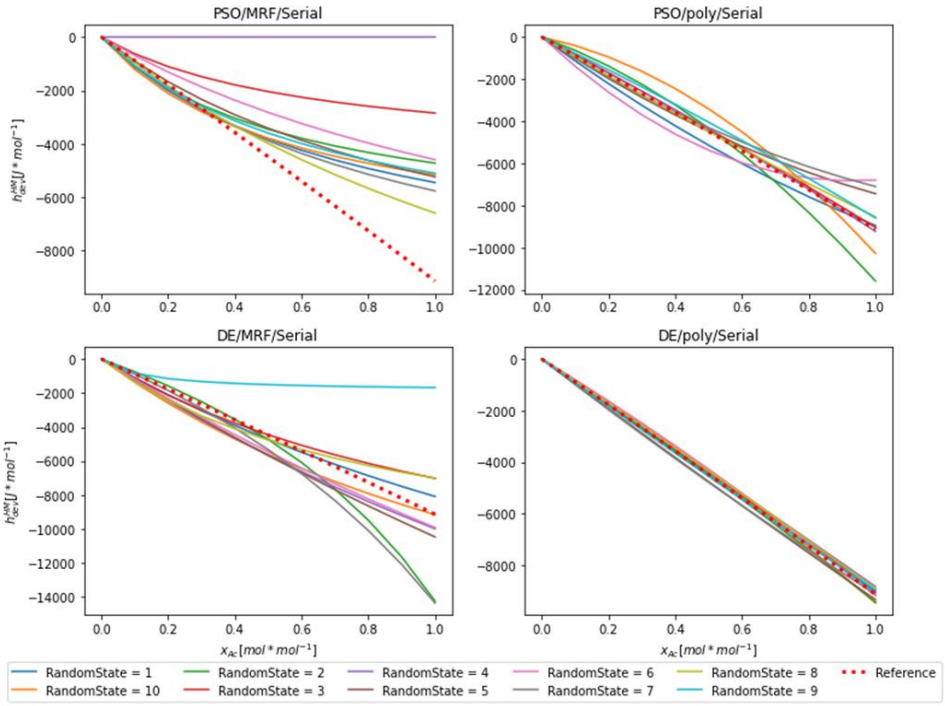

Figure 7: Predicted profiles of h_{dev}^{HM} obtained with the serial approach using various combinations of optimisers and statistical functions. The solutions are calculated at 50°C. All the solutions follow the reference trend and show comparable prediction accuracy. The MRF shows lower robustness with the initial condition than the polynomial function. The DE increase the robustness of the solution. The Poly/DE show the lower scatting degree of this work and the higher robustness with the initial populations.

3.3 Comparison between the two training approaches

In general, serial training shows a lower variance and median of the fitness function for all the techniques (Figure 3). It increases both the accuracy and the mathematical stability of the solutions. In addition, it reduces the influence of the initial population on the model training for the DE and PSO.

The serial training has higher accuracy and robustness than the parallel approach in predicting the behaviour of the overall system and the trend and the values of the black-box model. The primary deficiency observed in the parallel training is the mathematical instabilities in the test set linked to the synergistic effect of the two models in the coefficient search. The parallel training performs the coefficient search in a space with seven degrees of freedom. Moreover, these variables belong to different sub-models (enthalpy and heat transfer coefficient models). This increases the complexity of the system, and the overall fitness function can have more local minima compared to the serial training approach. On the contrary, the serial approach performs the coefficient search at first on a coefficient space with three degrees of freedom during the *first evaporation* (see Section 2.1), then on a coefficient space with four degrees of freedom during the *constant volume* phase (see Section 2.1). In addition, the coefficient search influences only one black-box model in this case.

Given that, the DE performs a better coefficient search compared to the PSO thanks to its augmented exploration capabilities, allowing the search to escape from local minimum easily.

The black-box functional form helps in decreasing the complexity of the cost function shape. The usage of the MRF increases the non-linearity due to the presence of the coefficients at the denominator. On the contrary, the polynomial has a linear coefficient search. Therefore, using the polynomial black-box structure decreases the complexity of the fitness function with 7 degrees of freedom and makes the coefficient search less complex.

A step toward reducing the complexity of the fitness functions is splitting the training of the two statistical models into two different moments. In this case, the coefficient search is more accurate than the one carried out with the parallel training. Here, all the techniques converge to a solution and return models with high accuracy and robustness. This is possible thanks to the reduced complexity of the fitness function that returns a local minimum from which is easier to escape. The technique that better performs in the serial training is the DE/poly. In this case, there is a combination of the high exploration capabilities and robustness of the optimisation algorithm, the reduced non-linearity provided by the polynomial structure and the reduced complexity obtained with the training splitting. However, the performance difference between the serial and the parallel training is negligible in this case. In other words, parallel training performs as well as serial training when DE/poly is employed as the training technique.

3.4 Extrapolation capabilities of the trained models

As discussed in section 2.2, the extrapolation capabilities of the model were evaluated on an extra batch. It was executed utilising half of the volume compared to the batches on which the models have been trained. From Table 4, it can be detected how the hybrid model outperforms a case-optimised first-principle model. In addition, from this table, one can see how serial training has superior results compared to parallel training. However, the difference is highly remarkable when PSO and MRF are employed to train the model. On the contrary, the accuracy of the models obtained by DE and polynomial function has little change when the training approach has changed.

Table 4 proves how the hybrid model developed with the techniques reported in this study can generalise and extrapolate very well on the variable area that was never experienced. This is an essential step when dealing with data that lacks variety.

Table 4: Determination coefficient (R^2) of the profiles within the reboiler predicted for the batch obtained with half of the volume than the batches used for the training.

	Reboiler Temperature		*Reboiler Volume*	
Technique	*Parallel*	*Serial*	*Parallel*	*Serial*
PSO/MRF	0.973	0.976	0.781	0.881
PSO/poly	0.986	0.986	0.894	0.916
DE/MRF	0.974	0.986	0.789	0.926
DE/poly	0.986	0.986	0.931	0.918
FP model	0.975		0.707	

4 Conclusions

This paper introduces a novel approach to developing hybrid models for chemical processes. The methodology aims to predict the behaviour of chemical systems under uncertainty (see SI section 1.5 and SI section 1.2). The proposed technique can work with data that lacks variety and value. Moreover, the resulting model has been shown to have enhanced extrapolation capabilities. The study case utilised in this paper is the constant volume solvent switch; however, the proposed methodology applies also to reactive systems and continuous processes. Besides developing a robust hybrid model, the methodology shows promising capabilities in estimating the physical coefficients of the system and letting their values be within acceptable physical boundaries.

This work investigated the performances of two heuristic optimisers (DE and PSO), two black-box functional structures (MRF and polynomial) and two training approaches (Serial and Parallel). The hybrid models based on the polynomial functions achieved better results than the MRF. The linearity of the polynomial form with the coefficient allows a more accurate coefficient identification than the MRF. The MRF structures report mathematical instabilities and overfitting issues when trained in

parallel. On the contrary, the hybrid models based on the polynomial functions do not show any mathematical instabilities. Although the serial approach usage resolved the mathematical stability issues, the prediction capabilities of the models based on MRF do not increase.
The role of the optimiser is marginal to the hybrid model accuracy. Even though the accuracy of the solutions proposed by the DE is higher than the one obtained with the PSO, the quality of the two models is comparable. On the contrary, the optimisers play a crucial role in estimating the coefficients of the black-box functions. The coefficient searches obtained with the PSO are more scattered and sensitive to the initial population positions than the ones proposed by the DE. For these reasons, the authors suggest using the PSO when there is no need to perform physical parameter estimation and the quality of the hybrid model is evaluated only on its final predictions. By contrast, the DE is suggested when the black-box coefficient estimation drives crucial information about the process.

The training approach has a crucial impact on the quality of the solutions. The serial approach stabilises the black-box solutions making them less sensitive to the initial population of the optimisers compared to the parallel training. This effect is linked to the lower complexity of the benchmark function when training the system in a serial manner.

The hybrid model obtained with the serial approach employing the polynomial function as a statistical function trained with the differential evolution algorithm (DE/Poly/Serial) returns a hybrid model with both the highest prediction capabilities and the most stable black-box function estimation. The high robustness of the optimiser, the linearity of the coefficient search and the serial training reduce the complexity of the cost function. This combination of search algorithm /model/ training scheme 1) reduces the dispersion of the black-box solutions, 2) lets the estimated parameters be within the acceptable physical boundaries and 3) returns a hybrid model able to predict the system behaviour with high accuracy.

Acknowledges

This work was supported by the DAP2CHEM ICON project funded by VLAIO and Catalisti (VLAIO HBC.2020.2455). The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

Competing Interests

KU Leuven submitted a patent application including these results on 16th December 2022 to the European Patent Office. The patent application number is EP22214177.2. Ulderico Di Caprio, Min Wu, Furkan Elmaz, Siegfried Mercelis, Peter Hellinckx and M. Enis Leblebici are included in the inventors list.

References

Aguitoni, M.C., Pavão, L.V., Siqueira, P.H., Jiménez, L., Ravagnani, M.A. da S.S., 2018. Heat exchanger network synthesis using genetic algorithm and differential evolution. Comput. Chem. Eng. 117, 82–96. https://doi.org/10.1016/J.COMPCHEMENG.2018.06.005

Angira, R., Santosh, A., 2007. Optimisation of dynamic systems: A trigonometric differential evolution approach. Comput. Chem. Eng. 31, 1055–1063. https://doi.org/10.1016/J.COMPCHEMENG.2006.09.015

Austin, A.P., Krishnamoorthy, M., Leyffer, S., Mrenna, S., Müller, J., Schulz, H., 2021. Practical algorithms for multivariate rational approximation. Comput. Phys. Commun. 261, 107663. https://doi.org/10.1016/j.cpc.2020.107663

BBVA, 2020. The five V's of big data [WWW Document]. URL https://www.bbva.com/en/five-vs-big-

data/ (accessed 7.15.21).
Biazi, L.E., Martínez-Jimenez, F.D., Bonan, C.I.D.G., Soares, L.B., Morais, E.R., Ienczak, J.L., Costa, A.C., 2020. A differential evolution approach to estimate parameters in a temperature-dependent kinetic model for second generation ethanol production under high cell density with Spathaspora passalidarum. Biochem. Eng. J. 161, 107586. https://doi.org/10.1016/j.bej.2020.107586
Bird, R.B., Stewart, W.E., Lightfoot, E.N., 2006. Transport phenomena. John Wiley & Sons.
Bogojeski, M., Sauer, S., Horn, F., Müller, K.R., 2021. Forecasting industrial aging processes with machine learning methods. Comput. Chem. Eng. 144, 107123. https://doi.org/10.1016/j.compchemeng.2020.107123
Chiang, L., Lu, B., Castillo, I., 2017. Big Data Analytics in Chemical Engineering. Annu. Rev. Chem. Biomol. Eng. 8, 63–85. https://doi.org/10.1146/annurev-chembioeng-060816-101555
Civicioglu, P., Besdok, E., 2013. A conceptual comparison of the Cuckoo-search, particle swarm optimisation, differential evolution and artificial bee colony algorithms. Artif. Intell. Rev. 39, 315–346. https://doi.org/10.1007/s10462-011-9276-0
Demchenko, Y., de Laat, C., Membrey, P., 2014. Defining architecture components of the Big Data Ecosystem, in: 2014 International Conference on Collaboration Technologies and Systems (CTS). pp. 104–112. https://doi.org/10.1109/CTS.2014.6867550
Dragoi, E.N., Curteanu, S., 2016. The use of differential evolution algorithm for solving chemical engineering problems. Rev. Chem. Eng. 32, 149–180. https://doi.org/doi:10.1515/revce-2015-0042
Elgue, S., Prat, L., Cabassud, M., Cézerac, J., 2006. Optimisation of solvent replacement procedures according to economic and environmental criteria. Chem. Eng. J. 117, 169–177. https://doi.org/10.1016/j.cej.2005.11.017
Fisher, O.J., Watson, N.J., Escrig, J.E., Witt, R., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L., 2020. Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. Comput. Chem. Eng. 140, 106881. https://doi.org/10.1016/j.compchemeng.2020.106881
Ge, X., Wang, B., Yang, X., Pan, Y., Liu, Botan, Liu, Botong, 2021. Fault detection and diagnosis for reactive distillation based on convolutional neural network. Comput. Chem. Eng. 145, 107172. https://doi.org/10.1016/J.COMPCHEMENG.2020.107172
Glassey, J., von Stosch, M., 2020. Hybrid Modeling in Process Industries. CRC Press.
Hong, Y., Ma, D., Ye, Z., 2018. Multivariate rational regression and its application in semiconductor device modeling. J. Semicond. 39, 094010. https://doi.org/10.1088/1674-4926/39/9/094010
Kennedy, J., Eberhart, R., 1995. Particle swarm optimisation, in: Proceedings of ICNN'95 - International Conference on Neural Networks. pp. 1942–1948 vol.4. https://doi.org/10.1109/ICNN.1995.488968
Mariano, A.P., Costa, C.B.B., de Toledo, E.C.V., Melo, D.N.C., Filho, R.M., 2011. Analysis of the particle swarm algorithm in the optimisation of a three-phase slurry catalytic reactor. Comput. Chem. Eng. 35, 2741–2749. https://doi.org/10.1016/J.COMPCHEMENG.2011.06.001
McBride, K., Sanchez Medina, E.I., Sundmacher, K., 2020. Hybrid Semi-parametric Modeling in Separation Processes: A Review. Chemie Ing. Tech. 92, 842–855. https://doi.org/https://doi.org/10.1002/cite.202000025
Mohan, P., Nicholas Emery, A., Al-Hassan, T., 1992. Review heat transfer to Newtonian fluids in mechanically agitated vessels. Exp. Therm. Fluid Sci. 5, 861–883. https://doi.org/https://doi.org/10.1016/0894-1777(92)90130-W
Mohd Ali, J., Hussain, M.A., Tade, M.O., Zhang, J., 2015. Artificial Intelligence techniques applied as estimator in chemical process systems - A literature survey. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2015.03.023
Papadakis, E., Tula, A.K., Gani, R., 2016. Solvent selection methodology for pharmaceutical processes: Solvent swap. Chem. Eng. Res. Des. 115, 443–461. https://doi.org/10.1016/j.cherd.2016.09.004
Qin, S.J., Chiang, L.H., 2019. Advances and opportunities in machine learning for process data

analytics. Comput. Chem. Eng. 126, 465–473. https://doi.org/10.1016/j.compchemeng.2019.04.003
Reis, M.S., Braats, R.D., Chiang, L.H., 2016. Big Data: Challenges and Future Research Directions [WWW Document]. AIChe. URL https://www.aiche.org/resources/publications/cep/2016/march/big-data-challenges-and-future-research-directions (accessed 4.14.21).
Sansana, J., Joswiak, M.N., Castillo, I., Wang, Z., Rendall, R., Chiang, L.H., Reis, M.S., 2021. Recent trends on hybrid modeling for Industry 4.0. Comput. Chem. Eng. 151, 107365. https://doi.org/10.1016/j.compchemeng.2021.107365
Schwaab, M., Biscaia, Jr., E.C., Monteiro, J.L., Pinto, J.C., 2008. Non-linear parameter estimation through particle swarm optimisation. Chem. Eng. Sci. 63, 1542–1552. https://doi.org/https://doi.org/10.1016/j.ces.2007.11.024
Stacey, A., Jancic, M., Grundy, I., 2003. Particle swarm optimisation with mutation, in: The 2003 Congress on Evolutionary Computation, 2003. CEC '03. pp. 1425-1430 Vol.2. https://doi.org/10.1109/CEC.2003.1299838
Storn, R., Price, K., 1997. Differential Evolution - A Simple and Efficient Heuristic for Global Optimisation over Continuous Spaces. J. Glob. Optim. 11, 341–359. https://doi.org/10.1023/A:1008202821328
Vesterstrom, J., Thomsen, R., 2004. A comparative study of differential evolution, particle swarm optimisation, and evolutionary algorithms on numerical benchmark problems, in: Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753). pp. 1980-1987 Vol.2. https://doi.org/10.1109/CEC.2004.1331139
von Stosch, M., Oliveira, R., Peres, J., Feyo de Azevedo, S., 2014. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Comput. Chem. Eng. 60, 86–101. https://doi.org/10.1016/j.compchemeng.2013.08.008
Willis, M.J., von Stosch, M., 2017. Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models. Comput. Chem. Eng. 104, 366–376. https://doi.org/10.1016/j.compchemeng.2017.05.005
Zhang, D., Wei, B., 2014. Comparison between differential evolution and particle swarm optimisation algorithms, in: 2014 IEEE International Conference on Mechatronics and Automation. pp. 239–244. https://doi.org/10.1109/ICMA.2014.6885702