

Digital Twins to Address Flowsheeting Limitations

Emma Pajak and Cameron Aldren

Department of Chemical Engineering, Imperial College London, U.K.

Abstract As a rapidly growing field, the flowsheeting industry's fundamental importance to process design is illustrated by its lucrative nature. Flowsheeting software, as with any assumption-based engineering modelling, faces limitations. Digital twins offer potential advancements that could address the limitations of flowsheeting, such as poor modelling accuracy, limited customisation, accumulation of errors, and poor cost estimation. Whilst research has explored unit operation digital twins, there has not been an endeavour to apply them specifically to the limitations of flowsheeting. Therefore, this project aimed to explore the use of digital twins of unit operations to specifically address flowsheeting limitations. In line with achieving this aim, a pump, heat exchanger, and reactor were selected, coded in Python, and subsequently embedded in the open-source flowsheeting software, DWSIM. Data for the digital twins were either sourced from manufacturers or generated in ASPEN, before processing through methods such as neural networks or polynomial regression. The key findings included: the pump library demonstrating a more accurate cost estimation compared to traditional models; the grey box reactor digital twin addressing assumptions of idealised models, improving accuracy; and the heat exchanger's preliminary success in its application to multiple fluid cases, showing potential to reduce the data required by digital twins. It was concluded that with consideration of the limitations around data availability, paired with further engineering theory implementation, unit operation digital twins have the potential to offer improvements to flowsheeting. Looking at the applications of this potential, from a manufacturer's perspective, digital twins of their equipment could offer compatibility validation and real system performance predictions which would improve customer confidence and, in turn, equipment sales.

Keywords: *Digital Twin, Flowsheeting, Unit Operations, Grey Box, Neural Network*

Introduction

Flowsheeting, or process simulation, is a fundamental tool in process design used across the breadth of chemical engineering. It supports material and heat balance modelling, equipment design, sensitivity analysis, and cost analysis [1]. Flowsheeting software can provide first estimations of physical feasibility, equipment sizing, cost and commercial viability for a chemical process that would otherwise be overly cumbersome by hand. A taught module in most chemical engineering degrees, flowsheeting is a crucial skill [2].

As per the principle known as Moore's Law that states the capability of computers doubles every two years, intensive simulations are continually becoming more accessible [3]. This has facilitated process simulation becoming a highly active area of research. By its nature as an advancing field, and fundamental importance in commercial engineering projects, flowsheeting software development has become an increasingly financially lucrative area. The latest market research predicts that the digital manufacturing industry will be worth 120 billion USD by 2030 [4]. Popular proprietary software packages include ASPEN, which mainly focuses on a sequential modular process modelling approach, and gPROMS which, conversely, is an equation-oriented software [5] [6].

As with any engineering modelling based on assumptions, flowsheeting faces limitations – specifically, challenges with assumptions providing a poor representation of 'real system behaviour'. Furthermore, as flowsheeting often models multiple interconnected unit operations, a singular error can accumulate throughout the flowsheet, further

damaging the accuracy of the design. Beyond this crucial downfall, unit operations can be rigid by design, rendering them non-customisable, and, therefore, unsuitable for bespoke processes. Flowsheeting cost estimation can also be a source of uncertainty in process design; costing packages are generally based upon correlations with a typical model CAPEX in the accuracy range of -50 to 100% [7]. Equipment design also holds inaccuracies as it treats sizing as a continuum; this can cause issues as manufacturers, unless specialists, offer a discontinuous set of equipment sizes. Whilst flowsheeting undoubtedly remains a crucial stage of process design, the accuracy of software recommendations and results hinges directly upon the specific assumptions employed - and implicitly adopted - by idealised models of unit operations. Inaccuracies with flowsheeting necessitate pilot-scale process development, which is expensive: to give an example, BP made a 19.5 million USD investment in a pilot-scale plastic recycling plant, in 2019 [8]. From an accessibility perspective, whilst open-source, community software exists, better-developed and integrated proprietary flowsheeting software licences are expensive and therefore less accessible for small-medium enterprises (SME) and start-up companies.

In summary, the flowsheeting field currently faces the following limitations: poor modelling accuracy, accumulation of errors in a flowsheet, limited customisation, and poor cost estimation. These limitations promote an opportunity to consider what improvements and concepts could be applied to further the current technology.

The term digital twin does not have a single discrete definition [9]. Rather, there are many definitions that

encompass similar key themes and ideas, each with slight nuances. Fundamentally, a digital twin is a “virtual copy of a physical asset, process, system or environment” [10] either created using real-world data, or, using real-time data from sensors on the physical version of the entity being simulated [11]. Although digital twins are an active area of research in the chemical industry, it is apparent that their specific application to the limitations of flowsheeting has not been fully explored.

Digital twins offer numerous potential advancements on traditional flowsheeting models; they use real process data as opposed to relying solely upon assumptions, in theory yielding a unit operation that models real system behaviour as opposed to solely approximating it. Further to model accuracy improvements, digital twins have the potential to address poor costing estimation, as exact process equipment data can be provided, introduce more flexibility and customisation of models, and, if based on an open-source platform, could offer great improvements to the process simulation capabilities of SMEs. In addition to bridging the gap on pre-existing flowsheeting limitations, the use of digital twins could further homogenise the flowsheeting stage of process design. With costing data sourced from manufacturers, as opposed to a single CAPEX figure, the output of a system of digital twins could essentially provide a ‘shopping list’ to the user, detailing the make, model, and cost of equipment they must procure.

Background

Many companies, including ASPEN, offer commercial-scale digital twin solutions to aid their clientele. These solutions, however, focus less on process design, and more on process operation applications, including operation scheduling, maintenance, and process control [12]. Such industrial attention has reaped substantial rewards for ASPEN’s clientele; for example, the speciality chemicals manufacturer, Momenive, employed the use of an ASPEN digital twin, facilitating a 25% reduction in site inventory and a 40% decrease in supply lead time [5]. Therefore, with such tangible benefits offered by digital twin technology, attention has turned to employ their use in the earlier stages of a processing plant’s lifecycle.

Within a subsection of ASPEN’s ‘Plant Digital Twin’, is the capacity for the digital twin to be used as a process model. This approach aims to unify the process model with digital twins once the plant has been constructed – using machine learning techniques to ‘calibrate’ the model to align with real operation [5]. Research by Bosch et al. highlights a similar approach to linking digital twin production to the conceptual design phase of process development. This approach translates a flowsheet into a ‘template’ for a digital twin, with a view that it can be filled with sub-models of the equipment as

and when the data is made available through testing [13]. This approach to digital twin development, using an existing chemical process to provide the requisite data to develop the digital twin, appears to have been adopted in further industrial instances. As of yet, there has been limited focus on developing digital twins for use during process design, to influence equipment selection and other significant financial decisions before construction.

Beyond digital twins of entire chemical processes, there have been efforts to produce unit operation digital twins that can be connected to produce a flowsheet model. Some flowsheeting software packages, such as DWSIM, allow users to develop a custom neural network operation, based on a dataset [14]. This would allow a user to develop a simple digital twin of a piece of equipment, should they have a relevant dataset available. Such a unit operation conforms to the ‘black box’ approach from machine learning, wherein complex neural networks are used to model a process. Although very accurate at modelling the nuances of the specific system, black boxes have no capacity for a user to interpret how the model translates the inputs onto the outputs [15]. As a result, such an operation can be rigid and does not allow for user insight to modify how the unit operation is calculated.

On the other hand, white box models, based entirely upon the physics of the process, have flexibility as the parameters of the model can be easily manipulated. However, these techniques become very complex, especially in 3D flow scenarios [16] and, as such, in the context of flowsheeting, require significant simplifications to make the flowsheet solvable, due to computational power limitations. Therefore, it is proposed that hybrid ‘grey-box’ models, using both data modelling methods, such as neural networks, and information based on the physics of the process, would allow the model to be solved in a reasonable time frame, whilst achieving a higher level of accuracy than an over-simplified white box model [17].

Therefore, this project intends to explore the research gap of using digital twins of unit operations to specifically address flowsheeting limitations. The following objectives have been constructed to support the achievement of the project aim:

- Identify three suitable unit operations, and the limitations of their traditional models, to produce digital twins that address these shortfalls.
- Implement digital twins in open-source process simulation software DWSIM using Python code.
- Conduct analysis to compare digital twins to traditional flowsheeting models to determine benefits and drawbacks.
- Explore and discuss the future of grey box digital twins within flowsheeting and highlight any obstacles to their production.

Methodology

In line with the project's aim, three unit operations were selected, created, and tested to best investigate the potential of digital twins in addressing the limitations of flowsheeting. A complete flowsheet comprises numerous different operations and therefore a well-balanced investigation of their digital twin implementations is required. A pump, heat exchanger, and reactor were chosen as, owing to their governing engineering relationships, they range in complexity.

The digital twins were coded in Python to then be embedded in a flowsheeting software. Accordant with the recognition that proprietary software is somewhat more developed than its open-source counterparts, and with the view of addressing the limitations of flowsheeting, the open-source software DWSIM (version 8) was selected [18]. DWSIM was specifically chosen as it has the infrastructure to easily integrate custom scripts directly into the software. Additionally, it is CAPE-OPEN compliant – CAPE-OPEN is a universal set of standards for process modelling software – meaning it supports common thermodynamic property packages. Also, CAPE-OPEN unit operations are more easily translated to other flowsheeting software [19].

Pump

To model exact pump behaviour, pump curves were sourced from Grundfos, a global pump manufacturer, as the datasets were readily available in an XLSX format for the entirety of their product range [20]. The datasets contained the following discrete data: head, power, efficiency and NPSH, all against flowrate. Polynomial regression models were applied to the datasets such that any desired pressure change within the given operating range could be achieved. An assumption was made that the pump would be operated with an inverter.

A python script defining the digital twin was implemented into the DWSIM user interface, which allows for a user-defined input stream and desired pressure change. The outlet pressure is calculated and compared to the pump curve to ensure that operation was feasible, i.e., the desired outlet pressure is within the operating range of the pump. Then, the outlet stream fluid enthalpy is calculated using **Equation 1**, where m is mass flowrate [kg s^{-1}], h is enthalpy [kJ kg^{-1}], P is power [kW], and η is efficiency [-]:

$$m_{out}h_{out} = m_{in}h_{in} + P\eta \quad [1]$$

The outlet pressure and enthalpy sufficiently define the pure, single-phase stream as per Gibbs' phase rule. A layer of validation was implemented to ensure the pump digital twin was operating within the range of its physical counterpart. For example, checking the temperature to ensure the fluid is

entirely in the liquid phase, and ensuring the difference between the inlet pressure (P_{in}) and NPSH, taken from the pump curve, is greater than the bubble pressure (P_{bubble}^*) for the liquid, i.e., the real pump will not cavitate (**Equation 2**).

$$P_{in} - NPSH > P_{bubble}^* \quad [2]$$

Pump Library

With the pump successfully addressing the accuracy limitation, it was further developed to explore its applicability in tackling the additional flowsheeting limitations around process equipment selection and costing. For the pump digital twin, a single pump dataset is inputted to DWSIM, whereas the pump library can ingest multiple pump datasets. The workflow of the pump library is, given a user-defined inlet stream and desired pressure change, the code simulates the digital twin for each pump dataset provided, selecting the optimal pump for the given scenario, whilst still outputting the outlet stream data. The optimal pump is determined by a simple optimisation model, where the user can select to optimise by price or power – i.e., if optimising based on price, DWSIM will output the least expensive pump that is still suitable for the given instance. This approach also used the aforementioned layer of validation to reject any non-viable pumps before the optimisation selection step. To assess the improved accuracy offered by the pump library, its cost and energy usage estimations for a set of scenarios were subsequently compared to an ASPEN pump model.

Heat Exchanger

The heat exchanger digital twin aimed to accurately predict the outlet temperature of the two streams, given two fully defined inlet streams. As the complexity of the heat exchanger is greater than that of the pump, its behaviour is dependent upon multiple different variables; therefore, a neural network approach was adopted, as it was apparent that a neural network would best represent the complexities of real system behaviour and non-idealities. To gain further insight, and develop the heat exchanger beyond the black box approach, the neural network was informed by heat transfer theory to explore whether a digital twin built on a water-water (water cold and water hot stream) dataset could accurately predict heat exchanger behaviour for other fluids.

The dataset fed to the neural network comprised four inlet and two outlet variables: the streams' mass flowrate, m_{in}^C , m_{in}^H ; the inlet temperature of the cool and hot streams, T_{in}^C , T_{in}^H , and the temperature of the outlet streams, T_{out}^C , T_{out}^H . Whilst it is plausible that manufacturers have testing rigs capable of generating this dataset, a synthetic version was used as, currently, these datasets are not readily available to consumers.

The synthetic water-water dataset was simulated using an ASPEN heat exchanger model by performing a four-variable sensitivity analysis on the inlet operating parameters to determine the hot and cold outlet stream temperatures. Furthermore, the identical model was used to simulate testing data for different fluid cases. It was ensured that the model operated solely in the liquid phase, to avoid further complexities during this initial investigation.

Instead of training the neural network on mass flowrate and temperature, it was trained on a pre-processed dataset, containing the variables $(mC_p)^H$, $(mC_p)^C$, T_{in}^H and T_{in}^C , capturing the thermal inertia of the fluid – enabling the digital twin to be tested for its applicability to other fluids. The neural network is then able to predict T_{out}^H and T_{out}^C for a given instance. The DWSIM implementation of the heat exchanger calls the pre-trained neural network and feeds the inlet temperature and heat capacity flowrate terms (the product of the mass flowrate and fluid specific heat capacity) for both streams and predicts the temperature of both outlet streams.

Testing was conducted to verify if the heat exchanger digital twin yielded accurate predictions of the temperature of the outlet streams. A single test set involved randomly selecting 100 different permeations of inlet conditions from the testing dataset, collecting the digital twin output predictions for each permeation, and calculating a single R^2 value to compare the predictions against the actual values. It was found that running a single test set multiple times caused the R^2 value to vary considerably, hence it was decided to run each test set 25 times to calculate the mean and standard deviation of the R^2 value to account for sampling bias. It was desired to understand the impact of the number of water-water training data points on the accuracy of prediction for each case. Hence, this testing procedure was repeated for each case using neural networks of varying training data points, n .

Reactor

To investigate the potential of a reactor digital twin, a simple, acid-catalysed esterification reaction of ethanol and acetic acid to form ethyl acetate was chosen. Due to the complex nature of a reactor, two different approaches were considered to produce a reactor digital twin.

The first approach was informed by reaction engineering to build upon the pre-existing, traditional reactor models by addressing their limitations by using real data in lieu of a poor ideal assumption. More specifically, this digital twin incorporated a real residence time distribution (RTD) to remove the assumption of perfect plug flow and, therefore, conformed to the more desirable grey box modelling approach. The second approach, like the heat exchanger, employed a neural network fit to predict the behaviour of the reactor. This

methodology will be referred to as a black box approach as the inlet variables are fed to the neural network – which is essentially a black box that outputs the result of the reactor in the form of fully defined outlet streams.

Approach 1: Grey Box

An experimental RTD was sourced from a research paper [21]. As an RTD is specific to a flowrate, reactor geometry, and volume, the digital twin built is specific to the reactor from which the RTD was captured. **Equations 3 & 4**, sourced from H. S. Fogler, were applied such that conversion could be obtained from the digital twin [22].

$$E(t) = \frac{C(t)}{\int_0^\infty C(t) dt} \quad [3]$$

$$\bar{x} = 1 - \int_0^\infty e^{-kt} E(t) dt \quad [4]$$

$C(t)$ is the RTD for a pulse tracer test, $E(t)$ is the probability density function, k is the first order rate constant, t is time, and $\bar{x}$ is the mean conversion. To understand if the digital twin offered improvements to the ideal model, the testing approach involved varying the rate constant across a broad range, to highlight whether conversion predictions differed between the two.

Approach 2: Black Box

The black box reactor's neural network ingested data that informed it on the impact of reactor temperature, feed flowrate, feed ratio, and reactor volume on reactant conversion. For simplicity, the reaction was defined as isothermal. Whilst it was appreciated that this is a cumbersome dataset that would not be available from a manufacturer, this approach was followed to gain a greater understanding of the use of neural networks to model digital twins of unit operations with multiple dependent variables. Hence, similarly to the heat exchanger, sensitivity analysis was conducted in ASPEN to synthesise a large dataset both for training the neural network and testing its accuracy.

Kinetic data were not required for the neural network as it is a black box which has an inherent appreciation of the complex relationships between inlet and outlet variables. However, the synthetic model, in ASPEN, used simple Arrhenius kinetics to simulate the dataset [23]. Unlike the grey box, this approach did not require comparison to the ideal model; since the neural network is built using a 'real system' dataset, if its predictions are accurate, it is inherently, by definition, a true representation of the real system. Hence, testing was rather carried out to determine the accuracy of the neural network's fit to the dataset. Importantly, statistical analyses were performed to better understand the impact of the number of training data points, n , on the accuracy of predictions.

Results

In line with the third project objective, the relevant outcomes of the three-unit operation analyses are shown to establish their capabilities in serving as improvements on traditional models.

Pump

Building upon traditional flowsheeting software, the output of the pump library, as well as a complete stream table, includes a pump curve displaying where the user's pump is operating, as shown in **Figure 1**.

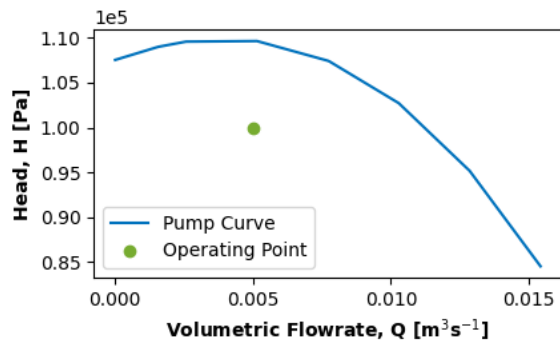


Figure 1 Pump curve of volumetric flowrate against head with operating point displayed from DWSIM pump library simulation. Data for this curve was sourced from Grundfos [20].

For the pump library, if an optimal pump is selected, a pop-up occurs, notifying the user of the model of the optimal pump, as well as its power or price requirement – depending on whether power or price optimisation was selected upon running the pump library. In the eventuality where an optimal pump could not be provided, as none of the pumps in the library had suitable operating ranges for the user's requirements, or the validation parameters (e.g., temperature and NPSH) were violated for all pumps, an error is passed to the command line.

Table 1 Table showing results of costing and power requirement comparison between ASPEN model and DWSIM digital twin.

Scenario		1	2	3
Pressure Change [bar]		10.0	5.0	1.0
Flowrate [kg s ⁻¹]		5.0	1.0	5.0
Cost [£]	DWSIM	5710	5370	2090
	ASPEN	5810	3900	4320
Power [kW]	DWSIM	9.01	3.92	0.92
	ASPEN	8.46	1.70	0.99

The output of the pump library is dependent on the data supplied and, therefore, the performance of the pump library, relative to the optimisation functions, would improve monotonically as more pumps are included. For this analysis, 20 Grundfos pumps were provided to the digital twin. As this is a relatively small number of pumps, compared to the product range offered by Grundfos, operating points were carefully selected such that these results highlight the potential benefits of the pump library. Therefore, simulations were run for pressure increases from 1

to 10 [bar] and flowrates from 1 to 10 [kg s⁻¹], with key results summarised in **Table 1**.

For most simulations, the DWSIM pump cost was lower than ASPEN's, e.g., scenario 3 - £2090 compared to £4320, supporting the hypothesis posited by Walter et. al, that ASPEN costing accuracy lies within a -50 to 100% of the actual value [7]. In contrast, the ASPEN energy usage was generally lower than DWSIM, e.g., scenario 2 – 3.92 [kW] compared to 1.70 [kW]. A 130% difference in energy usage is surprising, as ASPEN does account for a pumping efficiency parameter, but the exact pump curve efficiency taken in the DWSIM module appears to provide a markedly different value to ASPEN's estimation. These inconsistencies justify the need for a digital twin as they can offer an accurate cost estimation.

Heat Exchanger

For the case of the heat exchanger, having verified the digital twin was successfully determining the outlet stream temperatures for the water-water fluid case, it was further tested on different fluid cases, **Table 2**.

Table 2 Table of heat exchanger digital twin fluid test cases.

	Case 1	Case 2	Case 3
Cool Side	Ethanol	Wtr-Eth*	Heptane
Hot Side	Ethanol	Water	Heptane

* Water-Ethanol mixture 1:1 molar ratio

As the R² value of the testing predictions is the key parameter for this digital twin's performance, its mean and standard deviation are plotted in **Figure 2**.

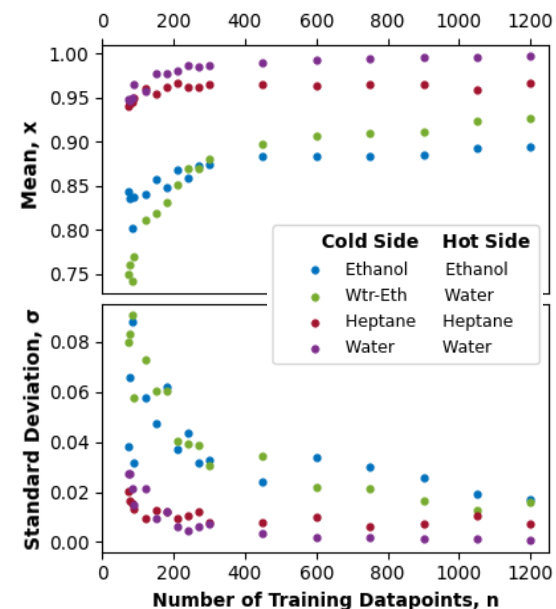


Figure 2 Graphs showing statistical analysis of the impact of the number of neural network training points, n, on the R² of heat exchanger predictions. Top graph: mean R² vs n. Bottom graph: standard deviation of R² vs n.

It is apparent from the water-water curves (purple data points in **Figure 2**) that the heat exchanger digital twin provides accurate predictions of the

outlet stream temperatures, even when trained with relatively few data points. For example, at $n = 120$ where, the mean R^2 value was 0.982, and the standard deviation was 0.006, meaning the neural network consistently gave accurate estimations of the testing data. The three remaining curves for cases 1-3 follow similar trends for both the mean and standard deviation: as the number of training data points increases, so does the mean of the R^2 values, and the standard deviation decreases. This analysis verifies that the digital twin, although built using water-water heat exchanger data, serves as a somewhat good predictor of heat exchanger behaviour for the other fluid cases. Thus, reducing the number of physical tests required from a manufacturer as one dataset can accurately simulate predictions for other fluids. Although promising, there is a need for further investigation into this approach as its potential, and limitations, need full and careful consideration. Such further investigation could involve the use of a real heat exchanger, rather than a synthetic dataset.

Reactor

Approach 1: Grey Box

To assess the potential benefits of the grey box approach over the traditional ideal model, a plot of conversion against the kinetic constant for both the digital twin (blue) and ideal plug flow reactor model (green) was developed, as shown in **Figure 3**.

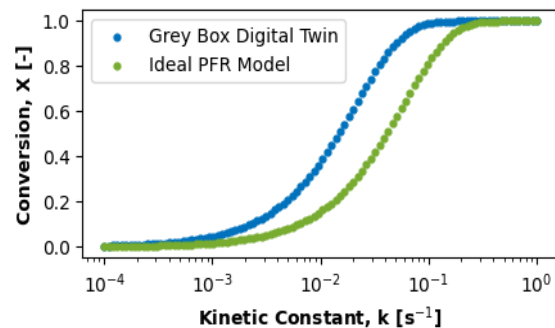


Figure 3 Graph of conversion against varying kinetic constant for an idealised reactor model (green) and the grey box reactor digital twin, incorporating a residence time distribution (blue) to better model real system behaviour.

Between a kinetic constant of $0.001 \text{ [s}^{-1}\text{]}$ and $0.1 \text{ [s}^{-1}\text{]}$, there is a marked difference in conversion, with the greatest difference at $k = 0.032 \text{ [s}^{-1}\text{]}$. At which point, the ideal plug flow reactor (PFR) model predicts a conversion of 0.42, compared to 0.78 predicted by the digital twin. The difference in conversion verifies the need for a digital twin as an improvement on the idealised PFR model. This significant difference in conversion is present because of axial dispersion within the reactor, which yields an asymmetrical residence time distribution.

The substantial benefit provided using an RTD is its incorporation of the impact of axial dispersion. In their study, Obonukut and Bassey, [21], highlighted

that, despite having a high Reynolds and Peclet number, their real reactor experienced substantial axial dispersion. It is in instances such as these, where the idealised model yields such an under prediction of conversion, that the fundamental insight offered by a digital twin can significantly improve the accuracy of predictions.

Approach 2: Black Box

As aforementioned, the ‘black box’ neural network inherently contains an appreciation for the exact nature of any non-idealities present in the system. Therefore, analysis was not required to determine the accuracy of the prediction as inherently it incorporates all non-idealities as based upon a real dataset. Instead, the KPI of this digital twin was not based on a comparison to an ideal model, but rather, the fit of the neural network to the parent dataset. Like the heat exchanger digital twin, the accuracy of the black box digital twin’s prediction was measured through the R^2 value. **Figure 4** shows a graph of the standard deviation and mean of the R^2 value against the number of data points used to train the reactor neural network.

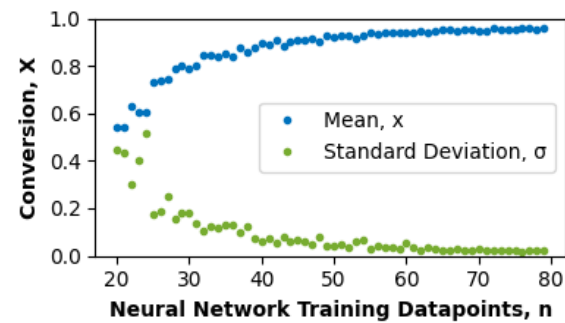


Figure 4 Graph of the relationship between the mean and standard deviation of the R^2 value for the black box reactor digital twin, against the number of neural network training data points, n , for reactor approach 2.

At 43 training data points, n , an R^2 of 0.9 is achieved, and, at 66 data points, an R^2 of 0.95 is achieved. Equally, once the training data points surpass 40, the standard deviation is lower than 0.065. This illustrates that the digital twin can both accurately and precisely predict the outlet conditions of the reactor effluent, given a set of initial conditions.

From a computational standpoint, it would be feasible to run a neural network trained on a number of data points in the hundreds, even thousands, to achieve an even stronger R^2 value. However, the limitations stand with the acquisition of such substantial amounts of data, specific to the reactor being modelled. Whilst these results show the black box reactor approach can be highly accurate in modelling the real behaviour of a reactor, the required data comes at a cost. It is unrealistic to expect manufacturers to be able to provide such a detailed dataset that essentially documents a full-variable sensitivity analysis of their equipment for all conceivable reactions.

Discussion

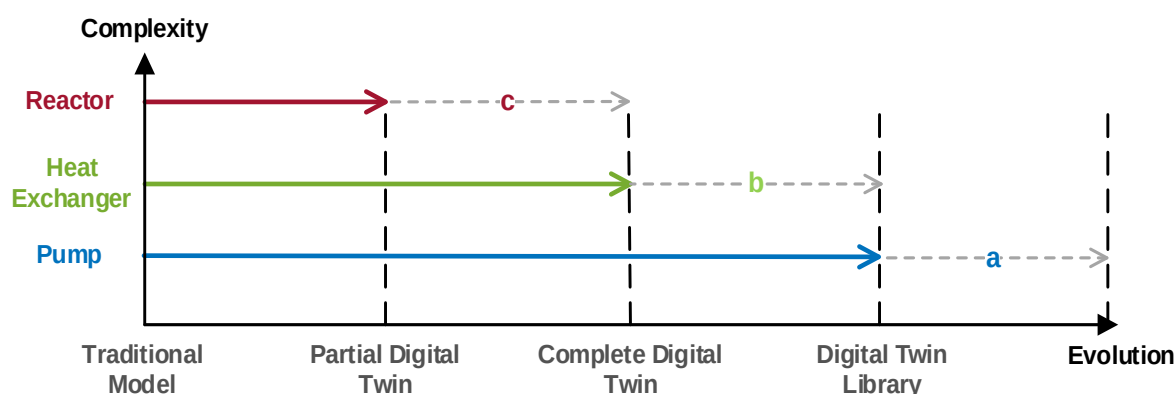


Figure 5 Schematic of digital twin ‘roadmap’, demonstrating the evolution of the unit operation digital twins developed within this project; labels a, b and c represent the next step of each digital twin. Whilst the figure plots the three discrete digital twins produced, as these twins are vehicles to represent varying complexity levels, all pieces of process equipment, with their corresponding complexity, exist on this evolutionary axis.

The varying complexity of the three-unit operation digital twins resulted in three digital twins at different stages of evolution. These different stages are plotted on an evolutionary axis, as shown in **Figure 5**, with the next steps for each digital twin highlighted by arrows, a, b, and c. In line with evaluating the potential of unit operation digital twins to address the limitations of traditional flowsheeting models, it is necessary to discuss their adherence to the grey box model definition, highlight any obstacles to their production, and discuss the potential uses of the finished products.

Pump Library

As the production of the pump library utilises both real data models and fluid mechanics theory, it adhered well to the definition of a grey box model. It showed success in addressing the flowsheeting limitation of inaccurate cost estimation and equipment selection. To build upon this success and take the next step in the development of the pump library digital twin, as shown in **Figure 5** arrow a, key challenges on data availability and collection need to be overcome.

Whilst pump curve graphs are readily available from manufacturers, currently, Grundfos is one of the few manufacturers where the raw data can be downloaded, as opposed to just being able to view graphs. The availability of the raw data in an appropriate format is fundamental to building and embedding the pump digital twin in a flowsheeting environment. There is potential for online tools to be employed to extract the data, however, this itself has its own limitations and is dependent on the graph format. Even with data being available in a convenient format, to fully realise the potential of the pump library, collecting this data needs to become a more automated process. As discussed in the results, the pump library’s prediction monotonically improves with a larger pump library; hence, greater automation of the data collection

process is important. Furthermore, whilst a primitive optimisation algorithm has been implemented in this project, further research into the multi-objective optimisation of pump choice would further facilitate a more flexible, user-customisable, selection.

Heat Exchanger

As the heat exchanger digital twin is built using a neural network informed by heat transfer theory, it too adheres to a grey box model definition. Within this category there are some digital twins, like the heat exchanger, that have a significant reliance on data modelling techniques, thus resulting in some black box-related restrictions, e.g., limitations in capacity for a user to understand the translation from input to output. The heat exchanger did show preliminary success through its potential to improve upon the poor accuracy of traditional flowsheeting models. Furthermore, it demonstrated potential in using a single heat exchanger case to predict the behaviour of different fluid cases. To further develop the heat exchanger digital twin, as shown in **Figure 5** by arrow b, the following challenges on data availability and application of heat transfer theory need to be overcome.

Whilst manufacturers do have suitable datasets of heat exchanger operation, it is something not currently readily available to consumers. To verify the findings from this project, further heat exchanger analysis must be conducted with real data. Additionally, in this early stage of investigation, an assumption of fully liquid operation was made – further considerations of how this digital twin can be applied to all heat exchanger instances, e.g., boilers and condensers, are required. Further validation of the potential to generalise the heat exchanger for use on other fluids by training the neural network on heat capacity flowrates is also recommended.

Reactor

The first approach to developing a reactor digital twin, as it applied real data to reaction engineering

theory, conformed well to the definition of a grey box model. In contrast, approach 2 was a paragon of a black box model as it was based solely on a neural network. Whilst approach 2 resulted in an accurate digital twin of a complex unit operation, it bears significant resemblance to the approach demonstrated by Bosch et al. [13], as this digital twin can only be developed once the equipment has been built and operational data collected. It is due to this, and the fact a manufacturer would not be able to provide such a dataset, that approach 2 (black box) will not be considered a viable approach for future reactor digital twins.

The grey box approach to producing a reactor digital twin showed significant success in addressing the flowsheeting limitation of poorly approximating real system behaviour. This was achieved through the removal of the perfect plug flow assumption by incorporating a real RTD. A natural next step in the advancement of this digital twin, as shown by arrow c, would be in addressing limitations present in the availability of the data required to produce the digital twin. RTDs are specific to a reactor's geometry and inlet flowrate and, therefore, for an unrestricted application to flowsheeting, multiple RTDs would be needed to build a digital twin of a single reactor. It is apparent that most reactor manufacturers have RTD data internally, but this data is not easily accessible to a consumer. A further consideration would be generalising this modelling for more complex kinetics and different types of chemical reactors.

Summary of Unit Operations

Whilst the three unit operations provide a demonstration of the potential of digital twins to address the limitations of flowsheeting, due to their differing complexities, each holds its own challenges in realising said potential. A shared theme between these digital twins is in the varying availability of consistently formatted, relevant, and complete data. Naturally, as systems of higher complexity are governed by an increasing number of variables, sourcing data that matches these requirements becomes increasingly difficult.

As per their definition, grey box models hinge upon two key pillars, models informed by real data, and incorporation of the process' physics. As such, is foreseen that with improvements in data availability and further focus on the implementation of engineering theory, the potential of digital twins of unit operations within flowsheeting could be further realised.

Flowsheeting Applications of Digital Twins

With the results validating the improvements made by unit operation digital twins, it is equally important to consider where their potential can be applied within flowsheeting and in a broader chemical engineering context. A key benefit

provided by these digital twins is that there is a substantial incentive for manufacturers to produce digital twins of their process equipment to further incentivise customers. For instance, a digital twin of a physical piece of equipment could be distributed with every purchase, thus enabling customers to run operational tests virtually, saving time and resources. For this to be feasible, ensuring the digital twins are CAPE-OPEN compliant is important as this would facilitate customers importing them into a CAPE-OPEN compliant flowsheeting software of their choosing. Equally, before the acquisition of the equipment, digital twin application programming interfaced (APIs), similar to the pump library, could be implemented on a manufacturer's website, such that customers receive recommendations of the most suitable piece of equipment for their requirements.

Alternatively, from a process design perspective, the digital twins could be further developed and implemented by flowsheet software companies. This would incentivise manufacturers to provide datasets for their product range, to promote their process equipment to customers. As demonstrated with the pump library this would ensure that the best unit operation is put forward to customers, encouraging competitive pricing from the manufacturers.

Digital Twin Applications Beyond Flowsheeting

Although the application of digital twins in flowsheeting is an area of significant potential, it is important to recognise that the current research areas of operation scheduling, maintenance, and process control, are still highly relevant. Therefore, the prospect of integrating these flowsheeting digital twins with these operations digital twins could be an area of further exploration. A key consideration is that the flowsheeting digital twins, produced in this project, are steady state models, not dynamic models which are more suitable for process control applications.

As such, further exploration into the production of dynamic unit operation digital twins could facilitate the implementation of an operations digital twin. Furthermore, in the context of process control optimisation, a highly specific and accurate digital twin is required to model the behaviour of the system in response to disturbances with a sufficient degree of accuracy. Therefore, the dynamic digital twin would need further tuning, once construction and testing have been completed, to give a better emulation of the real system. This calibration of a dynamic model to real system behaviour adopts a similar approach to that highlighted by R. Beck in ASPEN's white paper on the future of digital twins [5].

Conclusion

In conclusion, this project aimed to investigate the potential of employing unit operation digital twins to address flowsheeting limitations. This overarching aim was supported by four key objectives: building digital twins that serve as an improvement on idealised models, implementing the digital twins into DWSIM, conducting analysis to determine their benefits and challenges, and, finally, exploring where the potential of the digital twins could be applied within the flowsheeting industry.

The methodology was underpinned by the selection of three suitable unit operations – a pump, heat exchanger and reactor. These unit operations offered varying levels of complexity, number of dependent variables, and data requirements, hence serving as appropriate focal points of the investigation.

After constructing the digital twins' Python code, the DWSIM implementations enabled analysis that determined their potential as improvements upon current models. The pump library showed strong potential as it saw up to a 50% improvement in cost estimation, as well as facilitating the selection of process equipment. In the instance of the reactor digital twin, in addressing the ideal model assumption of perfect plug flow through the application of a real RTD dataset, its output gave a closer approximation of the real system's behaviour, with up to a 36% improvement made in conversion estimation. The heat exchanger also showed similarly promising accuracy, with R^2 values in the range of 0.85 to 0.98 for different fluids, when trained on over 300 data points.

There are two fundamental aspects to a grey box digital twin: the use of models built upon real data, and an incorporation of engineering theory. As well as recognising the benefits of the digital twins, consideration was given to identify the challenges they currently face. This project observed that improvements in data availability alongside a continued focus on implementing the engineering theory would alleviate these challenges.

It has been concluded that, with consideration of the identified limitations, the potential of unit operation digital twins could serve as improvements to traditional flowsheeting models. As such, it is conceivable that there is sufficient incentive for equipment manufacturers and flowsheeting software providers to develop this technology further. Furthermore, there is future potential for these flowsheeting digital twins to be homogenised with pre-existing operations digital twins, facilitating the production of a singular digital twin to support flowsheeting, dynamic modelling, and control system optimisation.

Outlook

Having established digital twins have the potential to address flowsheeting limitations, this outlook aims to outline possible avenues for development upon this project. Although recommendations will be made for each unit operation, focus will be given to the heat exchanger and reactor digital twins because, as shown in **Figure 5**, they are not as developed on the evolution scale relative to the pump library.

Concerning the pump library, the next steps could consider how to integrate multiple manufacturers into one digital twin. This would involve considering how to incorporate pump curve datasets of different formats as, naturally, the format will differ by manufacturer.

The grey box reactor approach successfully removed the assumption of perfect plug flow; however, further traditional, idealised assumptions remain. More specifically, the assumption of isothermal operation is a severely limiting one; temperature has a significant impact on reaction kinetics, resulting in poor approximations of reactor behaviour. For example, Arrhenius kinetics observes an exponential relationship with respect to temperature. Building upon the current grey box model, an investigation on how temperature could be more accurately captured by the digital twin would offer a significant improvement of real system behaviour predictions. Additionally, pairing the RTD with a specific set of reaction kinetics would allow inference on how the digital twins could be applied realistically in flowsheeting, as the grey box model was applied across a wide range of kinetic constants to verify the difference between its conversion and that of an ideal PFR.

Furthermore, if digital twins are to be used to inform equipment design and costing, credibility on their accuracy is a necessity. Taking the heat exchanger as an example, verifying the potential demonstrated in this project using real heat exchanger rig data would be the first step in fostering this credibility. Moreover, a similar approach can be taken with the reactor digital twin as its results are based upon the synthetic ASPEN model dataset.

Due to the nature of flowsheeting encompassing a wide range of unit operations, it is pertinent to consider others beyond the three explored in this project and analyse any further complications and challenges they may bring.

References

- [1] L. Evans, J. Boston, H. Britt, P. Gallier, P. Gupta, B. Joseph, V. Mahalec, E. Ng, W. Seider and H. Yagi, "ASPEN: An Advanced System for Process Engineering," *Computers & Chemical Engineering*, vol. Volume 3, no. Issues 1-4, pp. Pages 319-327, 1979.
- [2] Z. N. Pintarič and Z. Kravanja, "Towards Outcomes-Based Education of Computer-Aided Chemical Engineering," *Computer Aided Chemical Engineering*, vol. 38, pp. 2367-2372, 2016.
- [3] R. R. Schaller, "Moore's law: past, present and future," *IEEE Spectrum*, vol. 34, no. 10.1109/6.591665, pp. 52-59, 1997.
- [4] Oliver Wyman, "Projected size of the global market for digital manufacturing in 2030, by industry segment (in billion U.S. dollars) [Graph]," Statista, 2016.
- [5] R. Beck, "The Digital Twin and the Smart Enterprise," Aspen Technology, Massachusetts, 2019.
- [6] Siemens PSE, "Siemens PSE," 2022. [Online]. Available: <https://www.psenterprise.com/products/gpro.ms>. [Accessed 05/12/2022].
- [7] O. Walter, A. Tremel, M. Prenzel, S. Becker and J. Schaefer, "Techno-economic analysis of hybrid energy storage concepts via flowsheet simulations, cost modeling and energy system design," *Energy Conversion and Management*, vol. 218, no. 112955, 2020.
- [8] Edie Newsroom, "BP invests £19.5m in innovative pilot plant for 'unrecyclable' plastics," Faversham House Ltd, 2019.
- [9] D. Jones, C. Snider, A. Nassehi, J. Yon and B. Hicks, "Characterising the Digital Twin: A systematic literature review," *CIRP Journal of Manufacturing Science and Technology*, Vols. 29, Part A, no. 1755-581, pp. 36-52, 2020.
- [10] Unity, "What is a Digital Twin?," 2022. [Online]. Available: <https://unity.com/solutions/digital-twin-definition>. [Accessed 05/12/2022].
- [11] AWS Amazon, "What Is Digital Twin Technology?," 2022. [Online]. Available: <https://aws.amazon.com/what-is/digital-twin/>. [Accessed 05/12/2022].
- [12] A. M. Madni, C. C. Madni and S. Lucero, "Leveraging Digital Twin Technology in Model-Based Systems Engineering," *Systems*, vol. 7, no. 1, 2019.
- [13] S. Boschert, C. Heinrich and R. Rosen, "Next Generation Digital Twin," in *TMCE 2018*, Las Palmas de Gran Canaria, 2018.
- [14] D. Medeiros, "LinkedIn," 24 April 2018. [Online]. Available: <https://www.linkedin.com/pulse/integrating-chemical-process-simulator-tensorflow-daniel-medeiros/>. [Accessed 08/12/2022].
- [15] J. Benitez, J. Castro and I. Requena, "Are artificial neural networks black boxes?," *IEEE Transactions on Neural Networks*, vol. 8, no. 5, pp. 1156-1164, 1997.
- [16] Q. Meng, Y. Wang, X. Yan and Z. Li, "CFD assisted modeling for control system design: A case study," *Simulation Modelling Practice and Theory*, vol. 17, no. 4, pp. 730-742, 2009.
- [17] P. Aivaliotis, K. Georgoulas, Z. Arkouli and S. Makris, "Methodology for enabling Digital Twin using advanced physics-based modelling in predictive maintenance," *Procedia CIRP*, vol. 81, pp. 417-422, 2019.
- [18] D. Medeiros, "DWSIM Process Simulation, Modelling and Optimization Technical Manual," 2018. [Online]. Available: https://dwsim.inforside.com.br/docs/mobile/tech_manual.pdf. [Accessed 08/12/2022].
- [19] D. Piñol, J. C. Rodriguez, M. Halloran, W. Drewitz, I. Richard Szczepanski, M. Pons, M. Woodman, P. Banks and J. v. Baten, "Thermodynamic and Physical Properties v1.0 CAPE-OPEN," *COLAN*, no. 1.08.008.DOC, pp. 1-87, 2011.
- [20] Grundfos, "Grundfos," 2022. [Online]. Available: <https://www.grundfos.com/uk/about-us>. [Accessed 05/12/2022].
- [21] M. Obonukut and P. Bassey, "Residence Time Distribution of A Tubular Reactor," *International Journal Of Scientific Research And Education*, vol. 4, no. 1, pp. 4767-4777, 2016.
- [22] H. Fogler, *Elements of Chemical Reaction Engineering* 4th Edition, Philadelphia, PA: Prentice Hall, 2005.
- [23] W. D. Seider, D. R. Lewin, J. Seader, S. Widago, R. Gani and K. M. Ng, *Product and Process Design Principles - Synthesis, Analysis, and Evaluation* (4th Edition), John Wiley & Sons, 2017.

Appendices

Appendix A: Pump

Pump Digital Twin Implementation Diagram

For the pump digital twin implementation in DWSIM, the user-defined inlet stream and operating conditions are taken by the code and then compared to operating points on the pump curves, sourced from Grundfos. This generates a list of pumps which can perform the user's desired operation. This list is subsequently passed to the optimisation algorithm, which takes the user input optimisation parameter, and optimises the pump selection using a list of key performance indicators (KPIs) for each pump on the list, taken from the manufacturer. Subsequently, if an optimal pump has been selected, the model of the pump is outputted to the user, along with the exact outlet properties given by this pump model. This process is summarised by **Figure 6**.

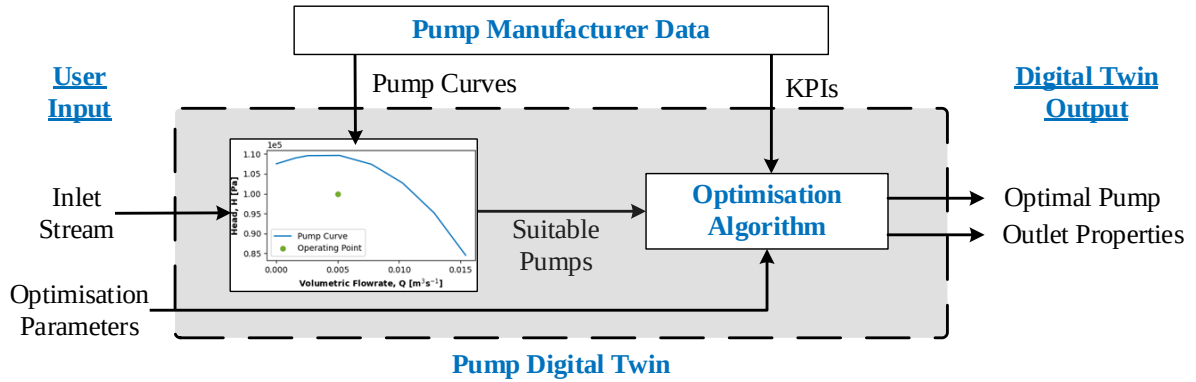


Figure 6 Diagram demonstrating the grey box pump library digital twin approach.

Pump Curve Dataset

The pump curve datasets sourced from Grundfos were excel files of a consistent format. These datasets were used in the implementation of the pump digital twin library, as highlighted above. **Table 3** is an example excel file.

Table 3 Table showing Grundfos example pump dataset as downloaded from the website [20].

Q [m ³ hr ⁻¹]	H [m]	P_2 [kW]	P_1 [kW]	η_2 [-]	η_1 [-]	NPSH [m]
0.000	16.229	0.535	0.637	0.000	0.000	0.816
1.823	16.105	0.581	0.688	13.751	11.601	0.895
2.632	16.030	0.602	0.712	19.071	16.118	0.925
5.264	15.694	0.672	0.792	33.445	28.373	1.007
7.896	15.206	0.742	0.874	43.978	37.350	1.078
10.528	14.557	0.812	0.956	51.350	43.569	1.155
13.160	13.733	0.877	1.036	56.039	47.431	1.257
15.792	12.725	0.937	1.110	58.332	49.206	1.401
18.424	11.520	0.989	1.177	58.346	49.035	1.602

Appendix B: Heat Exchanger

Heat Exchanger Digital Twin Implementation Diagram

For the heat exchanger, the data taken from the synthetic ASPEN testing rig were pre-processed, resulting in the list of data shown in **Table 4**. These data were subsequently used to train a neural network, which uses the user input operating, and inlet, conditions to provide an estimation of the hot and cold outlet stream properties.

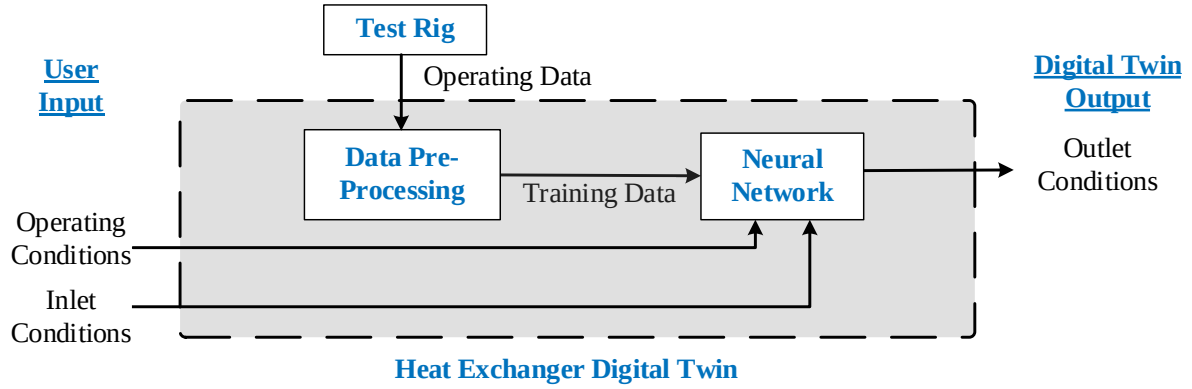


Figure 7 Diagram showing the grey box modelling approach for the heat exchanger digital twin.

Heat Exchanger Neural Network Training Data

The data shown in **Table 4** shows a sample of the ASPEN model data, after having been pre-processed, used to train the neural network. A similar dataset was produced for the different fluid combinations, as highlighted in **Table 2**, to test the model. It was ensured that the testing data were in the range of the training data, such that the neural network was not extrapolating its predictions.

Table 4 Table of example ASPEN simulated heat exchanger dataset for water cool side, water hot side.

mC_p^C [J K ⁻¹]	mC_p^H [J K ⁻¹]	T_{in}^C [K]	T_{in}^H [K]	T_{out}^C [K]	T_{out}^H [K]
4184	4184	335	275	287	323
4184	4184	335	280	291	324
4184	4184	335	285	295	325
4184	4184	335	290	299	326
4184	4184	335	295	303	327
4184	4184	335	300	307	328
4184	4184	335	305	311	329
4184	4184	335	310	315	330
4184	4184	335	315	319	331
4184	4184	340	275	287	327

Heat Exchanger Statistical Analyses

Table 5 highlights the results of the statistical analysis used to produce **Figure 2**. The mean and standard deviation of the R^2 values were calculated as each random sample of n_{train} and n_{test} was repeated 25 times. This allows for the implications of sampling bias to be mitigated; the mean gives key insight into the accuracy of the predictions and the standard deviation insight into the precision of the predictions.

Table 5. Table of statistical analysis of R^2 values of testing data for heat exchanger neural network of different fluids.

n_{train}	n_{test}	Case 1		Case 2		Case 3		Water	
		$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
60	100	0.778	0.123	0.709	0.076	0.934	0.020	0.944	0.027
66	100	0.821	0.050	0.770	0.067	0.940	0.015	0.946	0.021
72	100	0.843	0.038	0.749	0.080	0.940	0.020	0.948	0.028
78	100	0.835	0.066	0.760	0.083	0.943	0.017	0.946	0.028
84	100	0.801	0.088	0.741	0.091	0.945	0.015	0.950	0.021
90	100	0.837	0.032	0.769	0.058	0.949	0.013	0.964	0.015
120	100	0.839	0.058	0.811	0.073	0.961	0.010	0.957	0.021
150	100	0.857	0.048	0.819	0.061	0.954	0.013	0.977	0.010
180	100	0.848	0.062	0.832	0.060	0.961	0.012	0.978	0.012
210	100	0.869	0.037	0.851	0.041	0.966	0.009	0.981	0.006
240	100	0.858	0.044	0.869	0.039	0.962	0.011	0.986	0.004
270	100	0.873	0.031	0.870	0.039	0.962	0.012	0.984	0.006
300	100	0.874	0.033	0.880	0.031	0.966	0.008	0.986	0.007
450	100	0.883	0.024	0.898	0.034	0.965	0.008	0.990	0.004
600	100	0.883	0.034	0.906	0.022	0.963	0.010	0.993	0.002
750	100	0.883	0.030	0.910	0.021	0.965	0.006	0.995	0.002
900	100	0.885	0.026	0.912	0.017	0.965	0.007	0.995	0.001
1050	100	0.892	0.019	0.924	0.013	0.959	0.011	0.996	0.001
1200	100	0.895	0.017	0.926	0.016	0.966	0.007	0.997	0.001

Appendix C: Reactor

Grey Box Reactor Digital Twin Implementation Diagram

For the grey box reactor digital twin, the user is prompted to fully define the reactor inlet stream – as the reactor operation is isothermal, this implicitly defines the reactor conditions. From here, kinetic considerations are incorporated; for this initial investigation into the potential of the digital twin, as opposed to using a set of reaction kinetics, the kinetic constant, k , was varied across the range 10^{-4} to 10^0 $[s^{-1}]$ as it was desired to determine if the digital twin's conversion prediction differed to that of the ideal model. Next, using **Equations 3 & 4** the conversion for the reactor is calculated, and subsequently, a mass balance is employed to fully define the outlet stream parameters.

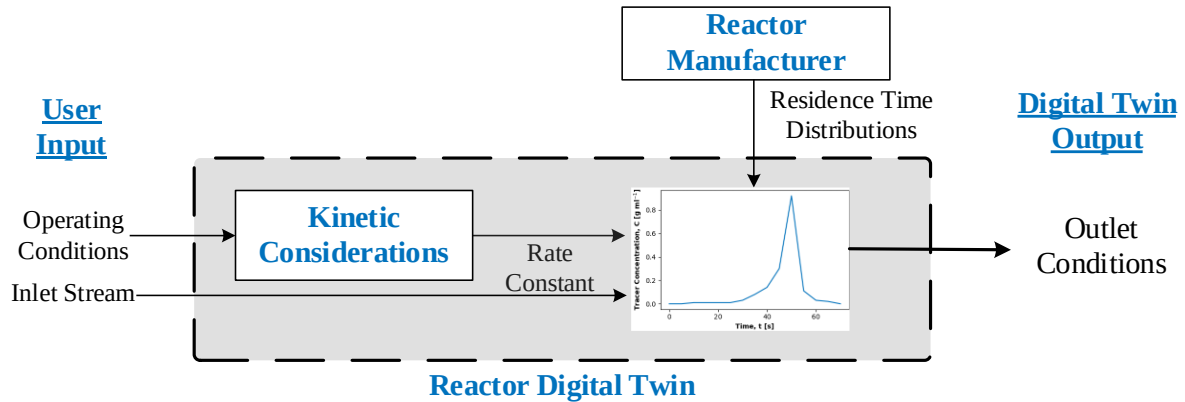


Figure 8 Diagram showing the grey box digital twin implementation of the reactor.

Grey Box Reactor Residence Time Distribution

Figure 9 shows the residence time distribution implemented in the grey box model.

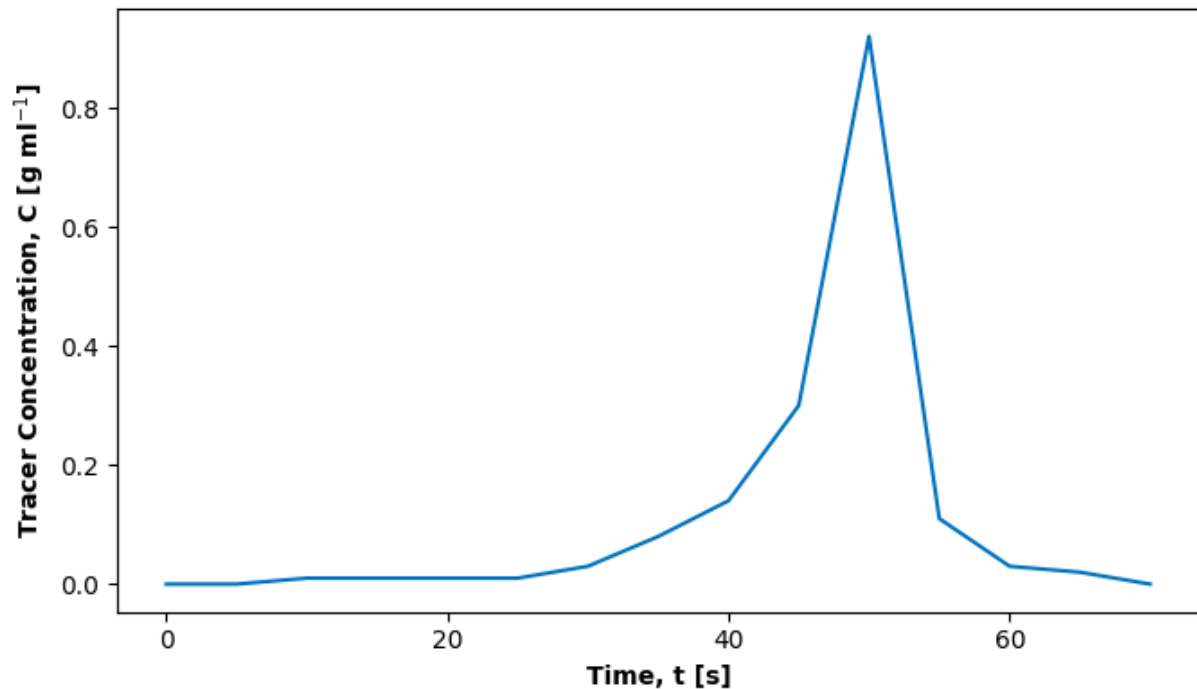


Figure 9 Graph of RTD (concentration against time), adapted from Obonukut and Bassey [21].

Black Box Reactor Digital Twin Implementation Diagram

The black box reactor used a synthetic testing rig dataset created in ASPEN, a sample from which is shown in **Table 6**. This reactor predicts the conversion of ethanol for a given feed flowrate, reactor temperature, ethanol feed fraction and reactor volume. Subsequently, a mass balance is performed in the DWSIM python script to specify the outlet stream properties. As this reactor is isothermal, the inlet temperature (and pressure) are simply transcribed from the inlet to the outlet.

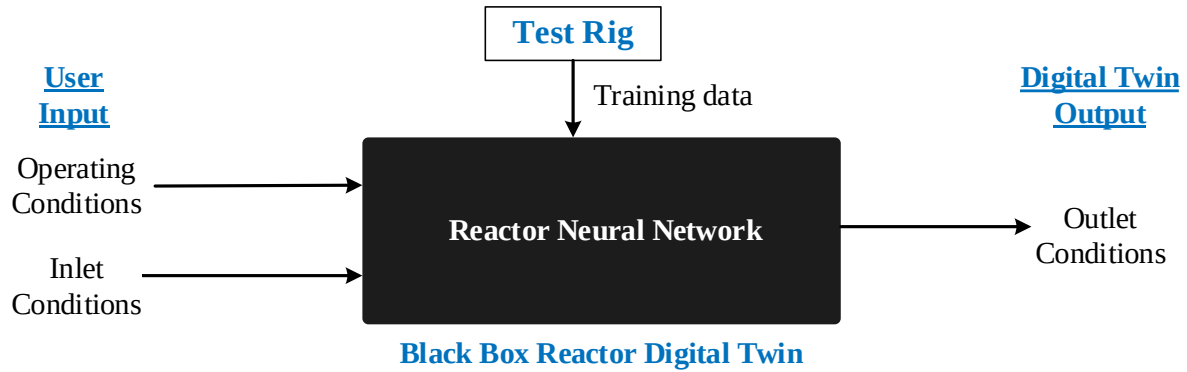


Figure 10 Diagram showing the black box reactor implementation. The neural network was trained on a synthetic testing rig dataset taken from ASPEN.

Black Box Reactor Training Data

Table 6 Table showing ASPEN simulated dataset for reactor digital twin black box approach.

Feed Flowrate [mol s ⁻¹]	Reactor Temperature [K]	Ethanol Feed Fraction [-]	Reactor Volume [m ³]	Ethanol Conversion
27.8	303	0.000124	0.00314	0.752
27.8	308	0.000124	0.00314	0.752
27.8	313	0.000124	0.00314	0.752
27.8	318	0.000124	0.00314	0.752
27.8	324	0.000124	0.00314	0.752
27.8	329	0.000124	0.00314	0.752
27.8	334	0.000124	0.00314	0.752
27.8	339	0.000124	0.00314	0.752
27.8	344	0.000124	0.00314	0.752
27.8	349	0.000124	0.00314	0.752