Review

Modelling for Digital Twins—Potential Role of Surrogate Models

Ágnes Bárkányi, Tibor Chován [ID], Sándor Németh and János Abonyi *[ID]

MTA-PE "Lendület" Complex Systems Monitoring Research Group, Department of Process Engineering, University of Pannonia, Egyetem Street 10, H-8200 Veszprém, Hungary; barkanyia@fmt.uni-pannon.hu (Á.B.); chovan@fmt.uni-pannon.hu (T.C.); nemeth@fmt.uni-pannon.hu (S.N.)
* Correspondence: janos@abonyilab.com

Abstract: The application of white box models in digital twins is often hindered by missing knowledge, uncertain information and computational difficulties. Our aim was to overview the difficulties and challenges regarding the modelling aspects of digital twin applications and to explore the fields where surrogate models can be utilised advantageously. In this sense, the paper discusses what types of surrogate models are suitable for different practical problems as well as introduces the appropriate techniques for building and using these models. A number of examples of digital twin applications from both continuous processes and discrete manufacturing are presented to underline the potentials of utilising surrogate models. The surrogate models and model-building methods are categorised according to the area of applications. The importance of keeping these models up to date through their whole model life cycle is also highlighted. An industrial case study is also presented to demonstrate the applicability of the concept.

Keywords: digital twin; surrogate model; model life cycle; model maintenance

1. Introduction

Digital twins can play a key role in combining the related physical and virtual entities into an efficient Cyber–Physical Production System (CPPS) [1]. More comprehensively, they can be considered as virtual representations of physical entities as well as their functions, data and capabilities provided that adequate synchronization from the physical world is available [2]. Diverse applications of these concepts have led to a wide range of interpretations, therefore, digital twins can be defined in many ways that are principally based on the purpose of their application. Product-related virtual prototyping and the digital twin of a physical entity were differentiated in [3]. The latter is explained as a synchronized representation of relevant information, e.g., structure, function and behaviour related to the physical entity. There are several commercially available engineering tools that support CPPS technology and the construction of certain types of models for digital twins mainly to mimic the operation of manufacturing systems [4] and their interactions with human operators [5].

The most important benefits expected from the application of the digital twin concept are the following [6]:

- Real-time monitoring and control could be extended in depth to large systems.
- Greater levels of efficiency and safety could be reached.
- Predictive maintenance scheduling supported by early fault detection.
- Scenarios and risk assessment as well as efficient and well-informed decision-making offer further benefits.

Digital twins often use detailed mathematical models, which have high computational demands that hinder the implementation of effective solutions of optimisation [7], including multi-objective particle swarm optimisation [8], and control and scheduling tasks [9]. In the case of chemical engineering systems, the simulation tasks would be the modelling of chemical manufacturing processes [10], scheduling scenarios [11] or complex

thermodynamics [12]. Most of these simulations require high computational demand and time to evaluate the different unknown model functions used in the simulator. The application of finite difference methods is hindered in the case of models that are noisy or discontinuous [13]. Therefore, the application of simulations with high computational demands in performing domain exploration, optimisation or sensitivity analysis becomes difficult because of the high number of the required function evaluations. An applicable solution to the aforementioned problems is the use of surrogate models (also known as metamodels, regression models or emulators), which are mathematically simple models that map or regress the input–output relationships of a more complex, computationally intensive model.

Surrogate models are used to substitute black-box of first-principle models that are either computationally expensive to evaluate or do not supply gradients [14]. In these situations, and when inaccuracies may occur due to the stochastic nature of the processes or due to unknown model parameters, it is beneficial to approximate the objective function and the corresponding gradients by surrogate models that lend themselves more easily to optimisation algorithms [15]. From this viewpoint simple linearization of complex models can also serve as surrogate models if the accuracy of the extracted linear model is adequate for the intended use of the model. Unfortunately, in most of the cases, surrogate models should describe more complex dependency of the variables and should approximate gradients that are difficult to evaluate, e.g., due to the stochastic nature of the system. In case of the simulators of discrete manufacturing systems, the discrete event and stochastic models do not lend themselves for simple linearization either. These problems will be presented in the case study of this work. Therefore, there is a need for surrogate models that

1. can be easily extracted from the complex simulators of digital twins,
2. handle the uncertainty and complex behaviour of the systems,
3. can be easily utilised in optimisation and control algorithms.

Another recently emerging solution for the high computational demand coming from the complexity could be the application of parallel computing. Although, there are some application areas, as well as optimisation problems with time expensive objective functions, in which parallel computing could be not enough. In those cases, their cost should be reduced using data-driven approximations or surrogates [16].

The main objective of this paper was to overview modelling methodologies for digital twins and their application scope, in addition to assessing surrogate models and exploring their application potentials. In this work we emphasize that the maintenance of the models applied in digital twins is extremely important through the whole life cycle of the model [17].

This review is based on the examination of the literature in Scopus by following the PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) [18]. The PRISMA-P workflow consists of a 17 tem checklist intended to facilitate the preparation and reporting of a robust protocol for a systematic review. The search protocol was built over the following steps. A search query was constructed based on the objectives of the research as listed in Table 1 (date of search: 13 October 2020). The following eligibility criteria were applied: (1) Date covered: the range of the search period was unlimited; (2) Search fields: title, abstract or keywords of articles in the data sources; (3) Document types: all types of documents were considered; (4) Language: only studies published in English were considered. The search strategy was defined, and after that, the articles were identified, screened and assessed for eligibility to develop the most relevant publications.

Table 1. Results of publication search in the fields of engineering.

Search Strings/String Pairs	Total Number of Publications
"surrogate model"	5702
"digital twin"	1577
"surrogate model" AND "optimisation"	3411
"surrogate model" AND "control"	754
"digital twin" AND "optimisation"	241
"digital twin" AND "control"	461
"surrogate model" AND "digital twin"	10

The collected bibliographic data were also applied to generate a network of keywords found in the publications (see Figure 1) to highlight the thematic groups how surrogate models are utilised in control or optimisation. Based on Figure 1, it can be concluded that the most relevant keywords reflect the related applications in multiobjective optimisation, structural design, computational fluid dynamics, shape optimisation and design of experiments.
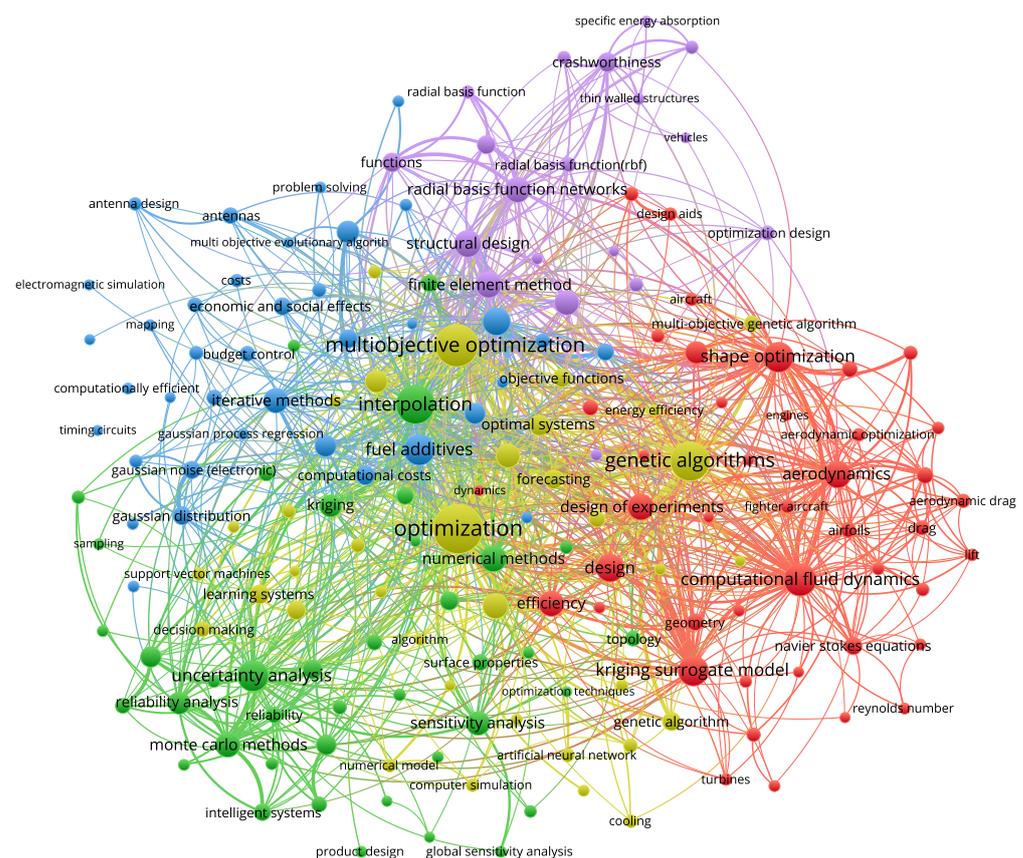
Figure 1. The network of keywords found in publications defines thematic groups connected to the surrogate models and optimisation or control in engineering (between 2017 and 2020).

Surrogate models are successfully and efficiently used in a wide range of engineering problems areas as it is presented in details in Table 2 summarising the main application classes, the applied surrogate model or models and the types of problems. In this paper, we tried to summarise the benefits and the drawbacks of the incorporation of surrogate models into digital twins. Consequently, we suggest that surrogate models can serve as

effective tools in the development of digital twins by replacing computationally prohibitive physics-based models while almost preserving their high fidelity.

In our literature review, an attempt was made to explore the applications as widely as possible in terms of operation optimisation, determination of optimal performance, feasibility analysis, design of process variables as well as design and control of processes. The use of surrogate models has many advantages including simplifying the object of study, reducing the computational demand of optimisation, accelerating the parameter estimation and analysis, simplifying sensitivity analysis to be performed, studying unknown black-box processes, etc.

Table 2. The applied surrogate model types for different application problems.

Problem Type	Surrogate Model Type	Ref.
Direct and Global Optimization		
Operation optimisation of a Cryogenics Natural Gas Liquids recovery unit	Neural Network	[19]
Optimization of catalytic reforming and light naphtha isomerization	Polynomial Function	[20]
Global optimisation of membrane processes	Neural Network	[21]
Determine the optimal structure and operating parameters for a process to minimize the sum of operating and capital costs	Kriging and Neural Network	[22]
Determine the optimal performance of a process at a natural-gas liquefaction plant based on a single mixed refrigerant	Radial Basis Function	[23]
Prediction and optimisation of the reaction and separation performance of a chemical process plant	Support Vector Machines, Kriging and Neural Network	[7]
Optimize operating conditions of a hydrocracking process	Kriging	[24]
Optimization of pumping rates in a coastal aquifer	Radial Basis Function and Kriging	[25]
Feasibility determination via Machine Learning	Support Vector Machine	[26]
Global optimisation of distillation columns	Kriging	[27]
Multi-objective Optimization		
A multi-objective optimisation of guide vanes	Support Vector Machine	[28]
An energy market design problem for a commercial building	Polynomial Function	[29]
Maximise the percentage of a scaffold filled with neotissue	Kriging	[30]
Multi-objective optimisation of management options for agricultural landscapes	Neural Network	[31]
Optimization of a sour water stripping plant	Kriging	[32]

Table 2. *Cont.*

Problem Type	Surrogate Model Type	Ref.
Synthesis and Design		
Design of a reusable launch vehicle for multi-mission	Kriging	[33]
Optimise the process conditions of hydroformylation process	Neural Network	[34]
Optimisation of vinyl chloride monomer production process	Kriging	[35]
Design of microfluidic concentration gradient generators	Kriging	[36]
Scheduling and Planning		
Integrated optimisation of scheduling and dynamic optimisation problems for a sequential batch process	Piecewise Linear Regression	[37]
Integration of planning, scheduling and control; Optimisation of an enterprise of air separation plants	Linear Regression and Neural Network	[38]
Reliability analysis of unidirectional fibre-reinforced plastic composites	Polynomial Function	[39]
Design and Control		
Simultaneous design and control of the Tennessee Eastman (TE) process	Power Series Expansion	[40]
Integration of design and control under uncertainty is developed for multiple steady-state processes	Fuzzy Model	[41]
Integration of process design, control and scheduling illustrated by two example problems, a system of two continuous stirred tank reactors and a small residential combined heat and power (CHP) network	State Space Model	[9]

The main application fields of successfully applied surrogate models in engineering practices are optimisation, namely direct, global and multi-objective optimisation; synthesis and design; scheduling and planning; as well as design and control. Therefore, it can definitely be assumed that the surrogate models can be used successfully in digital twins.

Despite the fact that the application fields are very similar for both the digital twin and surrogate model, only a few publications were found where the surrogate model was applied in the digital twin.

Therefore, the aim of our work was to explore the applicability of surrogate models in digital twins. By analysing surrogate model applications, it can be stated that neither the type of application nor the studied system determines the suitable type of surrogate model clearly. Hence, the paper also tried to present the details of the applicable model types.

According to this aim, the main contributions of this work are as follow:

- In Section 2 we provide a detailed analysis of the modelling approaches that can be used to build models used in digital twins.
- We provide an overview of the major steps of the identification of surrogate models in Section 3.
- In Section 4, the existing applications of surrogate models in digital twins are overviewed.
- The applicability of surrogate models is demonstrated by an industrial case study in Section 5.

- In Section 6, the proposed guideline for the incorporation of surrogate models in digital twins is discussed.

2. Model Building for Digital Twins

The new simulation paradigm for Industry 4.0 is best described by the digital twin concept [42]. The model of an existing or planned system can greatly help to understand or predict the behaviour of a production system or process. It can help reduce different costs, shorter periods of development, improve product quality as well as facilitate knowledge management. The integration of modelling solutions throughout the product life cycle management requires the use of the virtual factory concept. The digital twin is supposed to provide modelling solutions in all phases of the life cycle that support product development and testing in a virtual environment, as well as in the following phases while also gathering and using information from the previous phases. The digital twin contains a digital shadow representing a structured collection of operational, conditional and process data as well as a digital master, i.e., a universal model of the assets and their relations allowing the accumulation of knowledge through the product life cycle [42]. A digital twin of a Cyber–Physical Production System (CPPS) necessarily contains a digital twin of the business process to provide a detailed digital representation of the manufacturing plant that supports decision tools [43]. A possible application framework for digital twins is depicted in Figure 2.
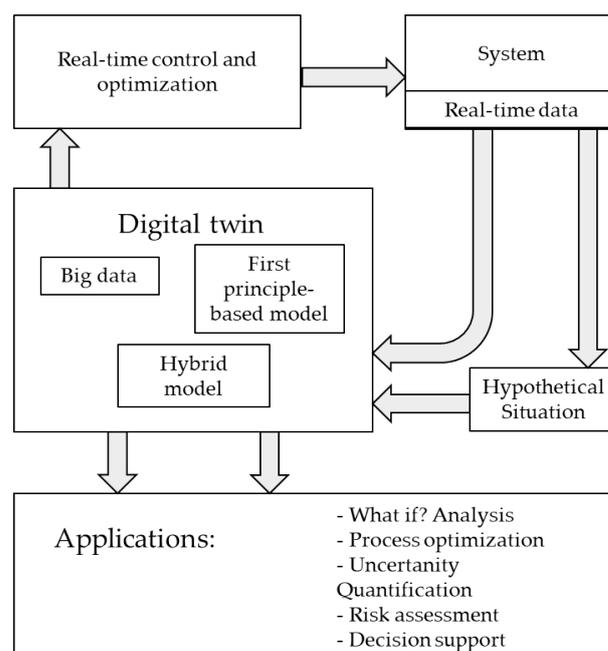

Figure 2. Application framework for digital twins (based on [44]).

It is clear that digital twins have the potential to be applied in almost every domain of our world providing that suitable methods are available to build such digital representations [45]. By definition, digital twins should ultimately be indistinguishable from physical bodies. This requirement sets huge challenges, e.g., dependability, sustainability, reliability and predictability. As depicted in Figure 3, three main approaches for building digital twins were considered: physics-based, data-driven and big data-based hybrid modelling [6].

Physics-based modelling is based on observing the behaviour of the physical entity and developing a partial understanding, which is then expressed in mathematical equations that are finally solved. Since the understanding is only partial and several assumptions are made in the process, a significant proportion of the real physical phenomena is ignored [6,46].

Data-driven modelling in digital twins is becoming more and more popular due to the increasing amount of process data, relatively low- high-performance computing solutions and efficient training methods [47]. Process data not only represents known physics but unknown parts too. Therefore, based on these data, the full physics can be mapped. Physics-based modelling is usually less biased than data-driven methods due to the fact that they use natural laws, which can be easily interpreted and generalized. However the expert building the model can still be a source of bias. The main disadvantages of this approach are possible numerical instability, sometimes still too high computational demand and the likelihood of uncertainties in the models. Nevertheless, high-fidelity simulations can support the development of Reduced Order Models (ROMs), also referred to as surrogate or metamodels, well, which are much better suited for digital twin applications. ROMs are supposed to provide a balance between the level of accuracy and the computational power necessary [48]. This might contribute to our surrogate modelling approach too and can justify our proposal to use existing simulations to develop suitable surrogate models.

Hybrid modelling combines physics-based modelling and data-driven modelling with big data approaches and allows the inclusion of more physics by increasing model complexity. On the other hand, the big data approach can provide better estimates of the related quantities. Approaches that extract data-driven surrogate models from physics-based models by simulations (i.e. model-based surrogates) can be located at the intersection of physics-based modelling and big data solutions, these are known as physics-driven surrogate models [6].
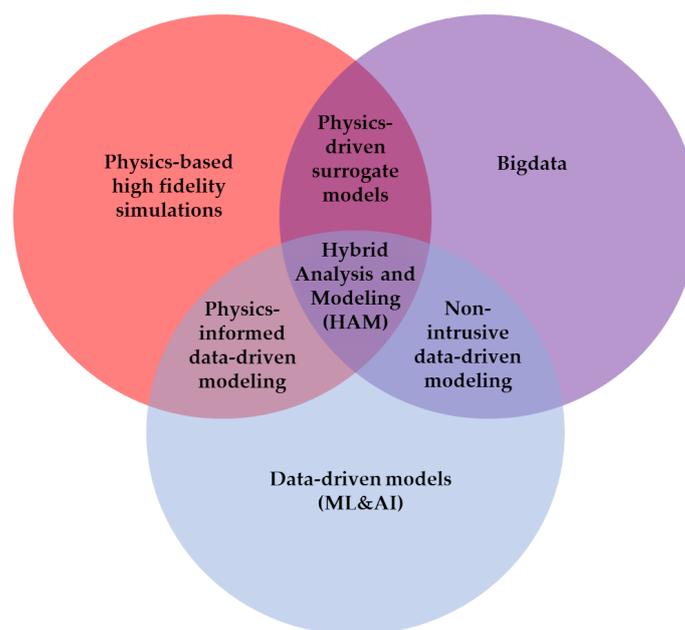

Figure 3. Combination of physics-based and data-driven modelling [6].

The advantages and disadvantages of physics-based modelling, data-driven modelling as well as physics-based surrogate models are summarized in Table 3. Most of the discussed properties are inherited from the modelling approaches independent of digital twins and surrogate models. However, these properties are important to discuss as they determine the applicability of the modelling approaches. Physics-based surrogate models have many inherent advantageous properties of physics-based and data-driven solutions. The main gains originate from the fact that they can describe system behaviour that is closely related to the model they were extracted from and consequently are stable. At the same time, errors and uncertainties can also be related to the original model. The surrogate model uncertainty depends on the type of surrogate model and the number of simulations that are used to train it [49]. Two surrogate techniques, Kriging and polynomial chaos expansion, were employed for modelling wind turbines in [50]. In this work, the general-purpose

uncertainty quantification framework UQLab was applied for the implemented surrogate models. The recommended Eurocode standard approach was utilized to calculate model uncertainty of both surrogate techniques, including all sensors. The application of the surrogate model uncertainty in reliability analysis based on the equivalent reliability index (ERI) and a new smooth sensitivity analysis approach to support the surrogate-based design process were presented in [51]. It was illustrated by three different case studies that the ERI approach could be utilized for surrogate model-based design problems with a low number of training data.

Table 3. Comparison of physics-based modelling, data-driven modelling and physics-based surrogate models (based on [6]).

	Physics-Based Models	**Data-Driven Models**	**Physics-Based Surrogate Models**
Benefits	Solid foundation based on physics and reasoning	Once the model has been trained, it is very stable for making predictionsinferences	Once the model has been trained, it is stable for making predictionsinferences
	Errorsuncertainties can be bounded and estimated	Takes into account long-term historical data and experiences	Errorsuncertainties can be bounded and estimated
	Less susceptible to bias		Less susceptible to bias
	Generalizable to problems with similar physics		Although blackbox-type, it reflects some of the physics
Drawbacks	Difficult to assimilate very long-term historical data	So far, most of the advanced algorithms function like black boxes	Difficult to assimilate very long-term historical data
	Sensitive and susceptible to numerical instability	Not possible to bound errors/ uncertainties	Limited generalization of unforeseen problems
		Bias in data is reflected in the model prediction	
		Poor generalization of unforeseen problems	

The aim of our work was to identify models that are easy to implement and evaluate as well as sufficiently flexible. Such models are, for example, linear regression, neural networks, Kriging and radial basis functions. A detailed discussion of these models can be found in [52,53]. Many engineering problems involve complex computer simulations, which allow more accurate as well as high-fidelity information about complex, multi scale, multi-phase and/or distributed computing systems to be obtained. However, these often contain proprietary codes, if–then operators or numerical integrators in order to describe phenomena that cannot be explicitly described by physics-based algebraic equations. Consequently, the algebraic model of the system and its derivatives are either absent or too complicated to obtain [54]. Surrogate models have the potential to speed up complex modelling without sacrificing accuracy or detail. Also known as metamodels, reduced-order models, model emulators, proxy models, lower fidelity models and response surface metamodels, surrogate models are computationally cheaper and designed to approximate the dominant features of a complex model [55].

The performance of surrogate models relies heavily on the quality and amount of samples. Therefore, it is critical what type of sampling methods are used to obtain samples from the detailed model. Such methods are, for examples, Monte Carlo simulations, genetic algorithms and Gaussian random samplers ([52,56]).

It is vital to decide what purpose these models are used for. One of the most important applications of surrogate modelling is optimisation, particularly robust design optimisation [57] or multi-objective optimisation [58]. Furthermore, surrogates are used for the online control of dynamic processes, as well as in feasibility evaluation, parameter

identification, sensitivity studies and scheduling [53]. From these applications, it can be seen that it is crucial to know the limitations of surrogate models, i.e., to know under what conditions they can be used. These issues are demonstrated in more detail in [52,55,56].

One of the most important tasks in surrogate modelling is to assess the reliability of surrogate models, since a less adequate surrogate model can lead to the loss of resources and have a negative influence on optimisation, prediction or feasibility evaluation. The validation of surrogate model is the process of assessing its reliability. Therefore, the validation of the model is an inherently important task [52].

However, the aforementioned studies exclusively focused either on optimisation or given modelling problems such as the availability of water resources ([55,56]). Therefore, it is important to examine which models, types of sampling methods and validation techniques are suitable for the given modelling problem. In the literature, unambiguous or straight guidelines cannot be found for the selection of surrogate models.

This fact also supports the aim of our work, that is, to explore what types of surrogate models and sampling methods are suitable in digital twins.

Typical tasks that may occur in model-based process operations and development, the so-called Computer-Aided Process Engineering (CAPE) activities, e.g., optimisation, synthesis and design, scheduling and planning as well as control, are illustrated in Figure 4:
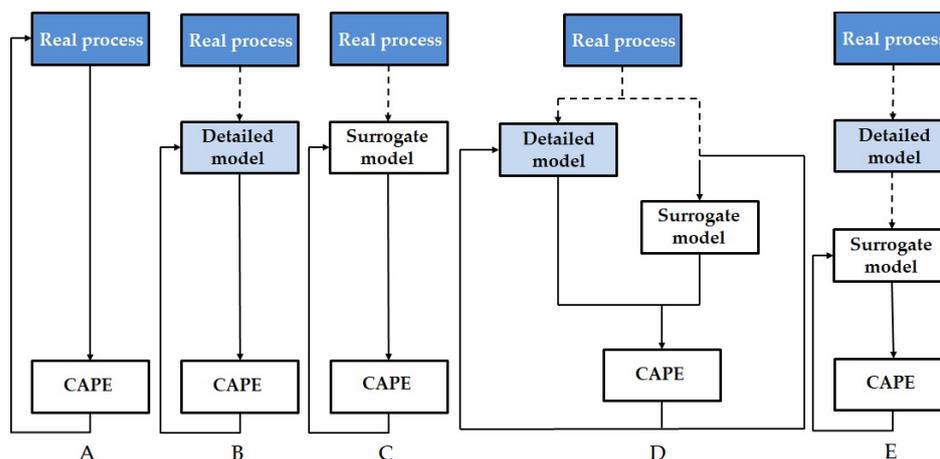

Figure 4. Surrogate-based Computer-Aided Process Engineering (CAPE) activities [59] (A: approach without model, B: approach with detailed model, C: approach with surrogate model, D: multi-fidelity approach, E: surrogate model is used to accelerate the detailed model.

In Figure 4, dotted lines denote the data flow, which is applied for model building. Solid lines represent the optimisation loop. Case A represents an optimisation loop without any models. In Case B, a detailed model is used to speed up the evaluation of the complex real-world function. In Case C, a surrogate model is applied to speed up this evaluation. Case D represents a multi-fidelity approach in which a detailed model of a surrogate model is additionally used. Models can be stacked as shown in Case E, where a surrogate model is used to accelerate the detailed model as well as the detailed simulation.

In a recent review paper [44], several references for using surrogate models in digital twins were included. The physics-driven surrogate models are defined as the intersection of big data and physics-based high fidelity simulations in Figure 3. The combination of principal component analysis (PCA) with Kriging was used to identify accurate low-order models for the development of digital twins of reacting flow applications in [60]. Furthermore, a surrogate model-based method for individualised spot welding sequence optimisation with regard to geometrical quality was introduced in [61].

In spite of the tremendous number of papers dealing with digital twins or surrogate modelling, no detailed and systematic analysis of applying surrogate modelling methods

in digital twins can be found, nor any reproducible case studies that could be the basis of the detailed analysis of the field.

In the next section, the most commonly used models are explained along with brief descriptions of their mathematical formulations as well as applicability limits. Furthermore, some sampling strategies, validation techniques and the main applications of surrogate modelling are presented.

3. Methodology of Surrogate Modelling

Building surrogate models in general requires a step wise development strategy as discussed in the following. Then, a short overview of surrogate model applications clearly suggests their promising capabilities in terms of digital twins.

The scientific challenge of surrogate modelling is the development of a surrogate that is as accurate as possible, using as few simulation evaluations as possible. The process is comprised of three major steps which may be interleaved iteratively (these steps are depicted in Figure 5).

1. Design of experiments and sampling for surrogate modelling (Section 3.1).
2. Model selection and fitting the model parameters based on simulation results using a detailed model (Section 3.2).
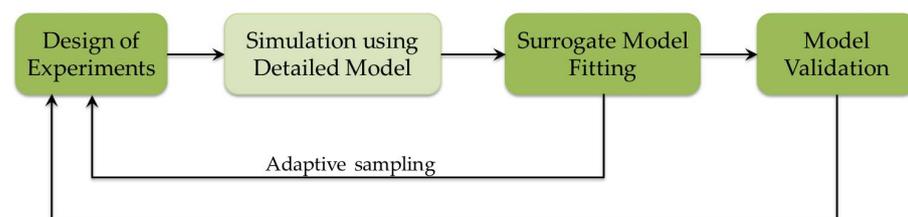3. Surrogate model validation (Section 3.3).

Figure 5. The major steps of building surrogate models.

In the next subsections, the major steps of surrogate model development are described in detail.

3.1. Design of Experiments and Sampling for Surrogate Modelling

The quality of surrogate models are significantly affected by the quality and the sufficiency of sample data [53].

Sampling is the step of generating data points that can be used in surrogate model building. Basically, two types of sampling strategies are differentiated that can be applied to surrogate design, namely adaptive sampling and stationary sampling. Figure 6 represents the comparison of the main steps of stationary and adaptive sampling processes. Stationary sampling methods rely on geometry or patterns. Frequently applied stationary sampling methods are Latin Hypercube Sampling (LHS) as well as the Sobol and Halton sequences [52]. LHS is a statistical method for generating a near-random sample of parameter values from a multidimensional distribution. Sobol and Halton sequences are quasi-random approaches. In these methods, Sobol and Halton low-discrepancy sequences are used to draw the samples. In the case of adaptive methods, new sample locations are determined serially. To start with, a lower number of samples are generated usually using stationary methods. The aim of the adaptive sampling strategy is to decrease sampling requirements by obtaining more samples that improve the quality of the surrogate. Different sampling methods are described in detail in [52]. In Table 4, publications from the last three years were collected. It appears that both stationary and adaptive methods are in use in engineering practices.

Table 4. The applied sampling methods and surrogate model types for different application problems.

| Sampling Method	Problem Type	Surrogate Model Type	Ref.
Stationary sampling methods			
Latin Hypercube	Design optimisation of a reversible axial-flow pump based on an ordinary one-way pump	Neural Network	[62]
	Multi-objective optimisation to obtain the optimal design of a cold plate structure	Polynomial Function	[63]
	Multi-objective optimisation and conjugate heat transfer calculation to obtain optimal cooling layouts on a transonic high pressure guide vane	Kriging	[64]
	Application of a surrogate model in the vibration analysis of graphene sheets	Kriging	[65]
Halton	Study of surrogate approaches to the nonlinear and multi-scale problem of turbulence and fire-spotting in wildland fire modelling	Radial Basis Function	[66]
	Design optimisation of a Pelton turbine runner	Radial Basis Function	[67]
Monte Carlo, Halton-based quasi-Monte Carlo	Natural gas pipeline design	Radial Basis Function	[68]
Adaptive sampling methods			
Adaptive	Predict the transitional flow past rough flat plates	Radial Basis Function	[69]
	Process optimisation of hydrofor-mylation of the 1-dodecene process	Support Vector Machine and Kriging	[70]
	Mixed-integer nonlinear benchmark problems and a chemical process synthesis case study	Kriging and Neural Network	[22]
	Design of microfluidic concentration gradient generators	Kriging	[36]
	Feasibility analysis of the continuous manufacturing of pharmaceutical tablets	Kriging and Radial Basis Function	[71]
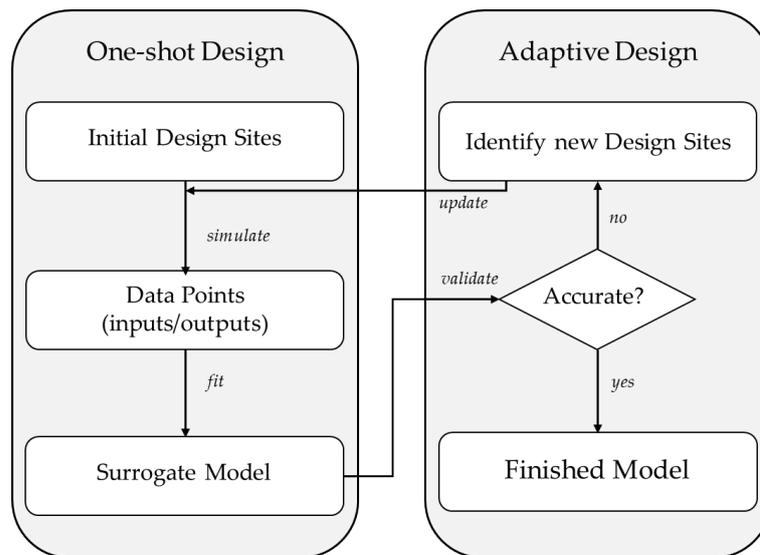
Figure 6. Comparison of the main steps of stationary and adaptive sampling processes based on [53].

3.2. Model Selection and Surrogate Model Structures

The most commonly used surrogate model types are the polynomials, Kriging models or nonlinear regression models of machine learning, like radial basis functions, artificial neural networks and support vector machines as seen in Figure 7. Some examples are presented in Tables 4 and 5.
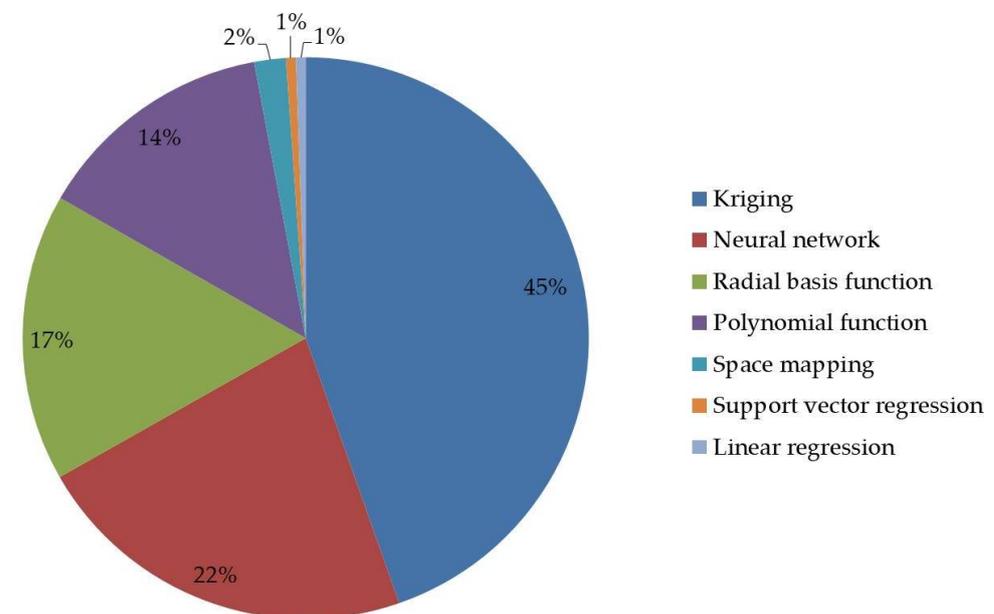

Figure 7. The most often applied surrogate model types in the field of chemical engineering (based on a Scopus search, number of articles: 720).

Linear Regression is the simplest surrogate model. Thanks to its simplicity, its computational requirements are small; therefore, it is often employed in engineering practices. It is well applicable for surrogate-based optimisation, where the number of calls to the function can be very large. In this approach, the surrogate is represented as a linear combination of the input variables as described by Equation (1) [52]:

$$y(\mathbf{x}) = \mathbf{w}_0 + \sum_{i=1}^{d} \mathbf{w}_i \mathbf{x}_i \tag{1}$$

where $\mathbf{x}$ denotes a vector of size d; d stands for the number of variables and $\mathbf{w}$ represents a vector of length $d + 1$.

Polynomial Functions are one of the most often used surrogate models in engineering practices. For regression purposes, computationally, these are the simplest models. Furthermore, they should be used in case of less complex underlying models. They are usually taken into consideration only main effects and first-order interactions, an example of this is shown in Equation (2). Higher-order interactions often lack significance and require more data to fit the additional parameters [53].

$$y(\mathbf{x}) = \mathbf{w}_0 + \sum_{i=1}^{d} \mathbf{w}_i \mathbf{x}_i + \sum_{i=1}^{d} \mathbf{w}_{ii} \mathbf{x}_i^2 + \sum_{i=1}^{d} \sum_{j \leq i}^{d} \mathbf{w}_{ij} \mathbf{x}_i \mathbf{x}_j \tag{2}$$

It follows from the above description that polynomials work well for low-dimensional problems. However, in engineering practices the high dimensional and highly non-linear systems to which they are not applicable are very common [53].

Kriging, also known as Gaussian process modelling [22], is one of the most commonly used surrogate models in the literature. Its mathematical basis is a Gaussian process model; therefore, it does not require a large number of fitted parameters, and at the same time, it is really flexible to describe many different functions and interpolate the data.

A Kriging surrogate model can be formulated as:

$$y(\mathbf{x}) = \sum_{i=1}^{m} \mathbf{w}_i y_i(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{3}$$

where $y_i(\mathbf{x})$ denotes m known independent basis functions that define the trend of mean prediction at location $\mathbf{x}$; $\mathbf{w}_i$ stands for unknown parameters and $\varepsilon(\mathbf{x})$ represents a random error at location $\mathbf{x}$ that is normally distributed with a mean of zero. Kriging is well applicable for problems where the dimensionality is lower than 20, the variables are continuous and the underlying function is smooth. If there are variables, which are discontinuous, the assumption of co-variance stationary of the correlation is not fulfilled and this will lead to low performance [53].

Radial basis functions are a weighted linear combination of local univariate functions applying selected measures of the distance from a point to an origin or a specified centre (Figure 8).

Given n distinct sampling points, radial basis function surrogates can be represented as in Equation (4):

$$y(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \phi(\| \mathbf{x} - \mathbf{x}_i \|_2) + p(\mathbf{x}) \tag{4}$$

where $\lambda_i, ..., \lambda_n \in \mathbf{R}$ denote the weights to be determined; $\| . \|$ stands for the Euclidean norm and $\phi(.)$ represents the basis function.

Generally, radial basis functions are suitable to situations where Kriging surrogates can be used; however, they are not used as often as in the chemical engineering literature as Kriging. This may be the main reason why the parametrised basis function of Kriging (which may be considered to be a special form of a radial basis function) has higher accuracy, flexibility and ability to make predictions of model variance [53].

Neural networks follow the information processing scheme of biological neural networks, e.g., the brain (Figure 9).
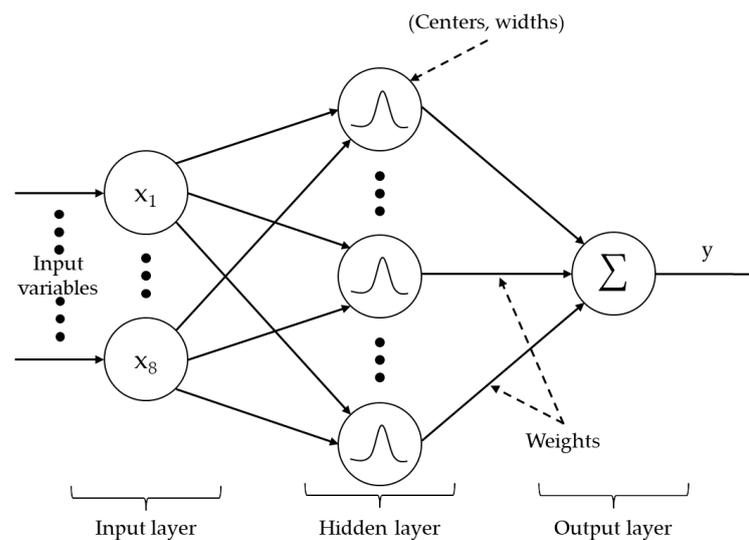
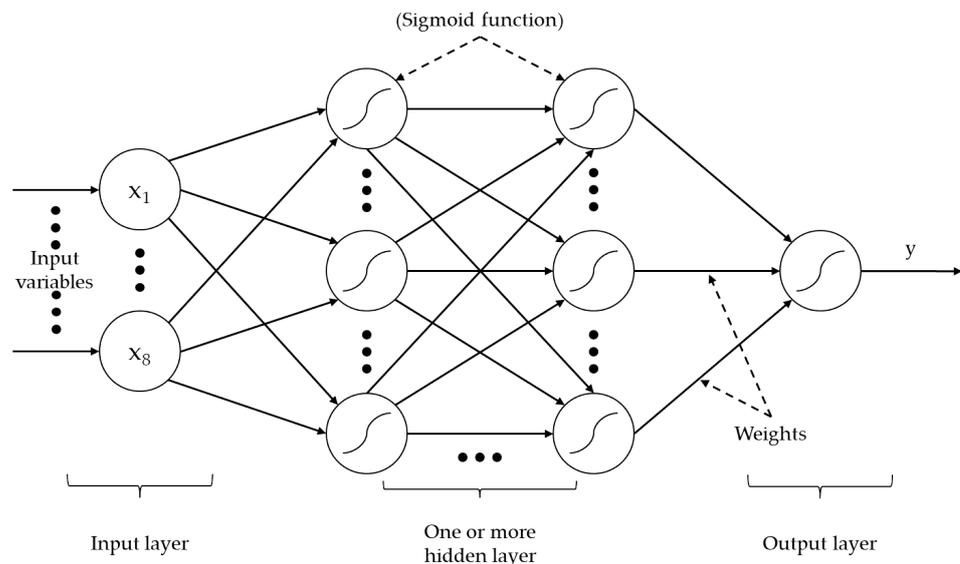**Figure 8.** Architecture of a radial basis function network.

Figure 9. Architecture of a multi-layer neural network.

These surrogates are suitable for fitting a wide variety of systems and have presented excellent results for many different tasks. The global characteristics of the design space for high-dimensional nonlinear systems can be described adequately by neural networks. The design of an appropriate network architecture, of which an infinite number of possibilities are possible, is the disadvantage of neural network modelling. Often a tremendous amount of data are required to fit the generally large number of parameters without overfitting. Therefore, artificial neural networks are recommended when a large amount of data are available or can be easily generated. Their use in computationally demanding simulations is not recommended, where the lots of function calls would become impractical, except if small networks of only a few neurons are applied. However, when these problems are insignificant or avoidable, neural networks are some of the most efficient surrogate models available [53].

When it is desirable to replace a complex, computational simulation with a surrogate model, then the following question arises: what type of surrogate model should be chosen? We found no consensus or clear-cut guidelines in the literature regarding the selection of surrogate models . A detailed review about the types of surrogate models and types of sampling algorithms can be found in [56]. In other papers, a rule-based method for

an automatic surrogate model selection called AutoSM [72] and a new selection criterion called the penalized predictive score [73] have been suggested. One of the key issues in the applicability of surrogate models is to determine their reliability, because if the surrogate model is not accurate enough, it has a negative effect on optimization, prediction or feasibility analysis. Some principles for selecting and building surrogate models have been identified, these were described in the presentation of the model types above. However, further research is still needed for exploring the appropriate approaches and methods.

3.3. Surrogate Model Validation and Maintenance

In general, model validation is the task of confirming that the outputs of a model have such fidelity to the outputs of the data-generating process that the objectives of the investigation can be achieved. Beyond assessing accuracy, validation techniques can be applied to select a surrogate model from the possible models and to fine-tune model parameters.

The validation procedure of data-driven models is independent of the model structure, so generally, the same method is applied in case of linear models or complex models generated by machine learning techniques. The data that were used to build the surrogate model should not be used to validate the model; therefore, during the building of the model, only a part of the available data should be used. The dataset used during the building of a model is called a training set and this set is used to validate the model referred to as a test set. The validation metrics can quantify the error of test set. Validation metrics that are commonly used to quantify this error using the re-sampling strategies are the explained variance score, the mean absolute error, the mean squared error, the median absolute error, the R^2 score, the relative absolute error and the relative maximum absolute error. These classical validation metrics are discussed in detail in [52]. It can be seen in Table 5 that the most often applied calculated type of error is the root mean square error, but often many types of metrics are calculated to determine the applicable surrogate model.

It must be emphasized that modelling should not be finished at this point. According to Table 3, one of the disadvantages of surrogate models is that their generalisation capability in systems that contain unforeseen problems can be limited. Besides, most processes do not operate around a true steady state due to changes in equipment, feedstock, sensors and operational strategy. These are the main reasons why it is extremely important to note that surrogate models require continuous maintenance during the whole life cycle of the model.

This requirement can be considered as generally valid for all models applied in industrial systems, namely models in advanced process control (APC), soft sensors, etc. For example, if an APC is left unsupervised, its performance will deteriorate over time. Deterioration is inevitable since a number of factors can change and affect the operation of the process. Maintenance of the APC system is essential to ensure continued performance [74]. In another study, guidelines are suggested for estimating the optimum maintenance cycles for APC projects [75]. Model maintenance is also crucial in the case of soft sensors. Several papers deal with maintenance approaches for soft sensors [76,77]. In both papers, different online soft sensor maintenance solutions are presented, e.g., semi-supervised [76] and Kalman filter-based [77] maintenance strategies. From the examples presented above, it can be seen that the maintenance of models is an inherently important task in digital twins too.

In [78], three main reasons are presented for a model to become impaired, namely non-stationary data distribution, degradation of hardware and system updates. This means that should the application environment or systems e.g., operating parameters, states, be changed, then it may become necessary to change the model parameters or even its structure and after that the model must be validated again. In order to assure or improve the quality of the model, a maintenance approach consisting of two sequential tasks (monitoring and updating) is proposed [78]. Another paper [17] draws attention to the need for the continuous maintenance of digital twins. They advise combining automated low-level adaptations for local updates with expert-driven revisions on higher levels.

Table 5. The applied calculated error types and surrogate model types for different application problems (Absolute Error (AE); Relative Mean Error (RME); Root Mean Square Error (RMSE)).

Problem Type	Surrogate Model Type	Calculated Error Type	Ref.
1D, 2D and 10D mathematical test functions; engineering problem: Prediction of the thrust on the whole rotor of a small unmanned aerial vehicle (UAV)	Polynomial Function; Radial Basis Function; Kriging	R^2	[79]
Multi-objective reliability-based design optimization of cementless hip prosthesis	Polynomial Function; Kriging	Maximum AE; RME; RMSE	[80]
Prediction of the load-bearing capacity of concrete-filled steel square hollow section members	Neural Network	RMSE; Mean AE; R^2	[81]
Mathematical and engineering test problems, e.g., Laval nozzle	Radial Basis Function; Kriging; Polynomial Function	RMSE	[82]
Evaluate gas-liquid flow in a horizontal pipe	Neural Network	RMSE; R^2	[83]
Mathematical test functions	Kriging	RMSE	[84]
Optimization of a horizontal axis tidal stream turbine blade	Radial Basis Function; Kriging; Polynomial Function	RMSE	[85]
Predict the microstructure of the final rod product	Polynomial Function; Kriging; Radial Basis Function	RMSE	[86]
Single-objective global optimization applied within an efficient global optimization framework	Kriging	RMSE	[87]

4. Potential Applications of Surrogate Models in Digital Twins

The application of digital twins is becoming more accepted in many industrial sectors from discrete manufacturing to process industries and even the utilities sector [88]. As an illustrative example, the digital twin of a water utility system can support the management with a complete, up-to-date view of the water system, can alert in the case of any anomalies and can provide accurate estimates about the operation of the system by integrating operational and business information flows. It can be considered as a model of the physical system that gets more and more accurate as fresh data or advanced machine learning solutions become available [89]. The scope of application ranges from management to control solutions, from operation through safety to optimality aspects, and can cover different parts of production facilities from the equipment to the enterprise levels. In the following subsections some illustrative examples and solutions representing recent application-related developments in process industries, control solutions and discrete manufacturing are discussed.

4.1. Potential Applications in Process Industries

In process industries (e.g., the oil refining and petrochemical industries), process systems engineering and related digital twin solutions play a key role in processing towards a smart operation [90]. Precursors of digital twinsdigital siblings have been used in large continuous plants for decades. They have been widely applied, for example, as operator training simulators. The core dynamic simulator of these solutions can provide an excellent foundation for digital twin applications [91]. A good example is the introduction of Honeywell in 2002, known as the Shadow Plant solution mimicking the process operation. In [92] the authors discuss that, when digital platforms are integrated within the concept

of Industry 4.0, transaction costs can be reduced, combining strengths of enterprises and realizing economies of scale as well as economies of scope.

The most relevant framework in which surrogate models are utilised in digital twins is depicted in Figure 10. Select process models describing the system's characteristics with sufficient accuracy are used in the forthcoming optimization and control steps. In this application, most of the calculations are conducted in commercial process flow-sheet simulators and process models were used as data generators to develop surrogate models. As the surrogate models map the input and output variables, the models can facilitate their integration with optimisation procedures and reduce the computational effort. The concept has been applied already in the petroleum industry, where artificial neural network is chosen as type of surrogate model with an adaptive sampling strategy, and the selected optimization procedure is genetic algorithm [93].
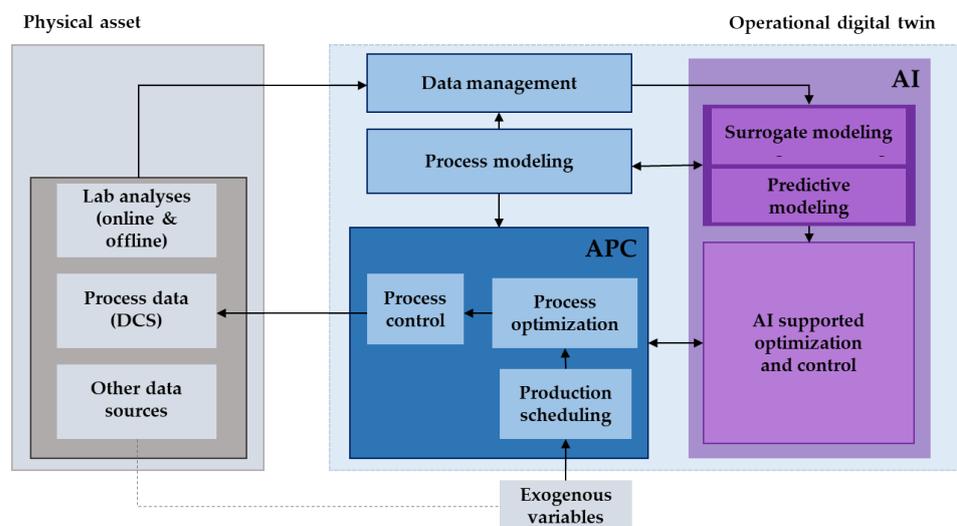

Figure 10. Operational digital twin framework supported by Artificial Intelligence (AI).

The main advances of the following areas are smart instrumentation, real-time optimisation, big data analytics (optimization, monitoring and management), advanced control and management platforms, predictive modelling techniques which allow a wide range of problems related to operational ability to be solved, abnormal situation management, planning and scheduling for oil refineries that data-driven modelling can take over when the application of complex first-principle models becomes prohibitive [90]. The possible main components of smart processing in the oil industry are summarised in Figure 11.

Energy efficiency is another key factor in process industries. Multi-stage compressors used in parallel arrangements, for instance, are often responsible for a considerable proportion of energy used in the chemical industries. A suitable real-time optimization framework combining short-term and long-term optimization for such systems offers an advantageous solution [94]. Since reliability and flexibility are ensured by using standby compressors, it is important to keep the compressor models up to date, while it is almost impossible that different compressors have identical characteristics and efficiencies. Hence, the use of data-driven modelling like surrogate models based on real-time process data is almost inevitable. The application of an integrated framework is demonstrated in an industrial case study and shows that it has a great potential to reduce the power consumption of compressor trains.
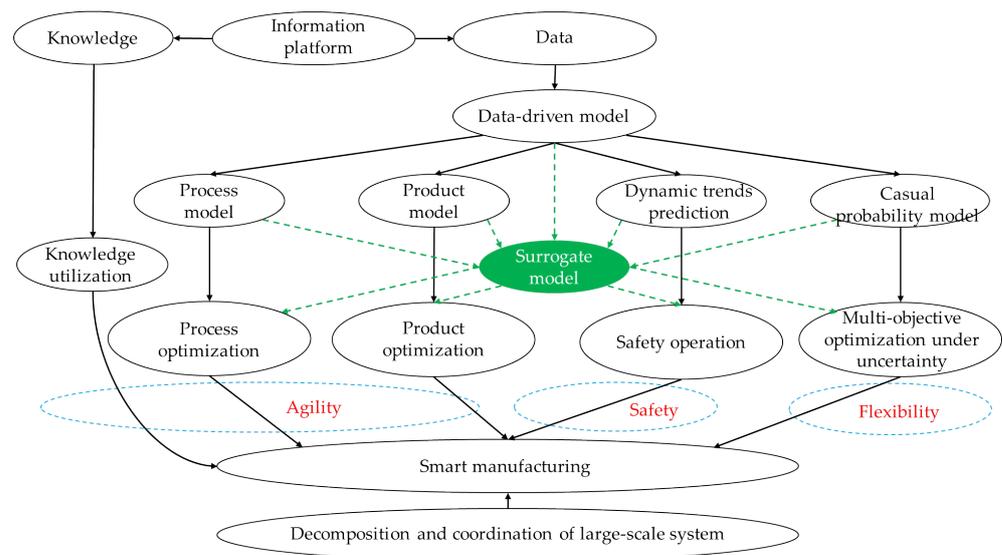
Figure 11. Components of smart processing in the oil industries [90].

4.2. Potential Applications in Control, Safety and Risk Management

Combining data-driven and physics-based models can help to describe the difference between physics-based mapping and experimental data. This approach allows for the performance of the model to be gradually improved as new data becomes available. The hybrid model approach for digital twins can be considered, e.g., it can be utilised in control applications and several types of surrogates [95]. Regarding control-related applications of digital twins, e.g., process analysis, optimisation, control design, commissioning and operator training, both static and dynamic models should be included. Since detailed physics-based models might be too complex and computationally intensive for optimization and control applications, one can refer to the surrogate models, which are simplified models that provide fast solutions and can be obtained based on more complex models.

Complex production processes involve several decision layers in a hierarchical structure, consequently, ensuring an optimal operation across the enterprise is a difficult task. Solving problems at different levels in an isolated manner frequently results in sub-optimal or inconsistent solutions. Therefore, compared to traditional solution strategies, an integrated approach that treats the different decision levels in an integrated framework offers solutions over a wider domain [96]. Although digital twins are not explicitly referred to, the methodology clearly relies on adequate and suitable up-to-date models of the production process; therefore, these models match the features of digital twins discussed earlier. Contrary to monolithic models that require centralised optimisation methods, the application of distributed optimisation is strongly based on surrogate modelling.

In terms of process industries, the complexity of the combined process and control systems as well as the severe and often extreme operating conditions can involve high risks to one's health, the environment, process safety and operational security. An important potential of digital twins is its ability to handle and lessen these risks by detecting faults and allowing operators to monitor the process and test deviations. A cooperative framework that integrates monitoring, diagnosis and optimised control can reduce process fluctuations as well as guarantee robust and safe process control [97]. This approach applies the subspace identification method to build discrete-time multivariable state-space models of the blocks of the relevant decomposed process. The framework consists of modelling, residual computing and evaluation, monitoring, diagnosis and fault detection, as well as tolerant control and optimisation. Sensor and actuator faults, process disturbances and cases involving multiple faults are all handled. The applicability and advantages of the proposed methodology are demonstrated in the Tennessee Eastman benchmark process [98], which is often utilised in operation-related research. Describing the process behaviour of digital twins adequately can help to assess the process risks and to prevent

possible losses. Well established commercial process simulators, such as Aspen HYSYS, serving as a steady-state simulation module, can be extended by a suitable process hazard analysis module and then can be applied for automated hazard analysis of chemical processes as it was demonstrated on a complete ammonia production plant [99]. Process safety analysis usually involves Hazard and Operability Studies (HAZOP). Using adequate dynamic simulation models, dynamic HAZOP analysis can be performed to explore hazard events like controller failures [100]. Validation of the functional model should rely on process knowledge and quantitative process simulation that is important in HAZOP studies [101,102]. A digital twin centred framework applying a reference model offers a more general solution for risk prediction and prevention. The proposed reference model covers all layers of the process plant: physical space of the process, communication system, digital twin and user space. The outlined digital twin system involves tools for simulation, control and execution, anomaly detection and prediction, as well as a cloud server platform supporting real-time data handling [103].

As assets of manufacturing facilities are becoming the focus of process management, asset management plays an important role throughout the process life cycle. Advanced simulation solutions can predict stochastic processes that take place in the physical assets. In this sense, a digital twin means a virtual representation of a physical system that employs field signals and is especially well-suited for asset-related decision-making, for example, risk management approaches [2]. In this way, asset-related decisions can be linked to any of the asset control levels (strategic, tactical or operational). Besides asset configuration, reconfiguration, planning and commissioning, digital twins facilitate effective asset diagnosis that helps asset condition monitoring as well as health assessment and consequently provide an asset-centric approach to safety problems. Features necessary for asset health diagnostics can be extracted from process data through the functionalities of digital twins.

Digital twins, as continually updated dynamic virtual instances of a physical system, adequately reflect the performance and health status of the physical system and can be applied everywhere in the system life cycle as model-based systems engineering tools [104]. Among other characteristics, they enable malfunctioning equipment to be troubleshooted by combining operation and maintenance data. The authors distinguish between four levels of digital twins. Pre-digital twins are built without physical twins and support, for example, risk assessment. Digital twins in general are such virtual models that incorporate performance, health and maintenance information of physical twins. Adaptive digital twins include an adaptive user/operator-sensitive interface that supports real-time decision-making. Intelligent digital twins employ machine learning to provide more granular information and have a high degree of autonomy.

4.3. Potential Applications in Discrete Manufacturing

Regarding discrete manufacturing processes, digital twins can embrace the complete product life-cycle. Digital twins of the product and the manufacturing process representing the designer's ideas and physical constraints can support an iterative process to reach optimal construction. In smart factory schemes, digital twins of the production line and the manufacturing process can be utilised to develop manufacturing plans and strategies, as well as real-time process monitoring and supervision. The digital twin of the product can be used for analysing the product state and the environmental effects, as well as predicting the lifespan of the product, diagnosing faults and consequently applying smart maintenance, repair and operations (MRO) strategies [105].

Particularly considering Industry 4.0 initiatives, smart manufacturing and progress in cyber–physical systems within discrete manufacturing can hardly be digitalised without the strong involvement of digital twins. The ISA-95 hierarchy model of manufacturing systems clearly expresses the need to integrate the different levels of manufacturing processes from enterprise resource planning through manufacturing execution management into the control of the production process. In discrete manufacturing, digital twin solutions most often rely on simulation modelling or discrete-event simulation and efforts aim to

automate model building. Recent advances in the field mainly focus on efficient model building strategies for digital twins. Comprehensive methodologies even allow special engineering and operational tasks to be solved, e.g., damage-tolerant planning based on damage diagnosis and prognosis supporting design, production and maintenance actions [106]. The scheme and the main components of the proposed solution are depicted in Figure 12.
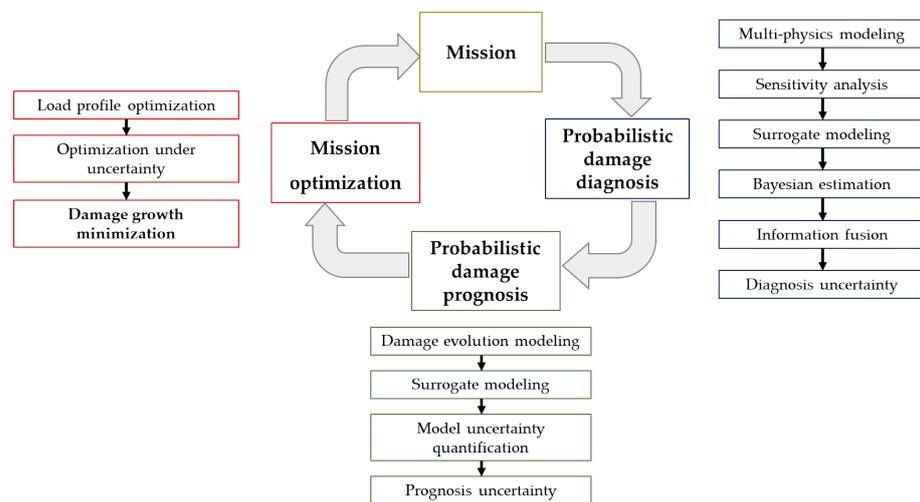

Figure 12. Components of a digital twin for damage-tolerant planning [106].

A novel methodology for the high-level model building of digital twins relies on statistical, discrete-event simulation and optimisation as well as automates many of the steps of model construction [107]. The method focuses on the bridging abstraction metamodel, which captures all discrete-event considerations in terms of manufacturing, supply chains, warehousing and distribution as well as transportation and logistics. The system model is then transformed into a metamodel by applying a universal modelling framework.

An automated discrete-event simulation model construction approach can also be developed by relying on information obtained from the production system [108]. Based on customer order data and the manufacturing process database, the method automatically constructs ad hoc models to solve different optimization tasks, e.g., modification of the factory layout to reduce the travel distance of products. The method is based on a commercial simulation tool and the ad hoc models were constructed by modifying standard XML files representing the model. A promising method can automatically create virtual factory models based on production configuration data given in generic XML format [109]. A virtual factory represents a high-fidelity, multilevel simulation involving a wide range of heterogeneous models. Another solution generates an complete flow shop model by automatically locating, linking and parametrising predefined objects [110]. The necessary information is extracted from the enterprise information system and manufacturing execution system. This later provides data on production sequences as well as current and planned progress of production regarding workstations and manufacturing orders.

5. Application Example

To demonstrate the applicability of the developed framework, we present the main results of the cycle time control of a production line of a wire-harness assembly line, where the control model was extracted from the digital twin of the process.

The studied assembly line consists of $w = 1, \ldots, N_w$ workstations where operators perform different sets of activities related to the production of different types of modular products (see Figure 13). We developed a digital twin that automatically updates the technology simulator based on the information extracted from the manufacturing execution system (MES) and a real-time location system (RTLS), where the RTLS was used to provide information about the material flows and the assembly times [111]. The scheduling, control

and the monitoring of the production are challenging due to multiple types of product being produced in the conveyor. The sequence of the N_p types of products is represented as $\pi(k) \in \{1, \ldots, N_p\}$ according to which type of product is being produced at the first workstation in the $k = 1, \ldots, N$-th cycle. The product types are defined based on their $m = 1 \ldots N_m$ modules according to the binary vectors p_p, the product definition of which can be considered as the bill of materials (BOM).

The proposed adaptive digital twin identifies the changes on the shop floor to make the process simulation adaptive and online (see Figure 14). The model is based on the estimated activity times of the operators that is identified based on the RTLS-measured duration of the product spent in each zone/cell [112].
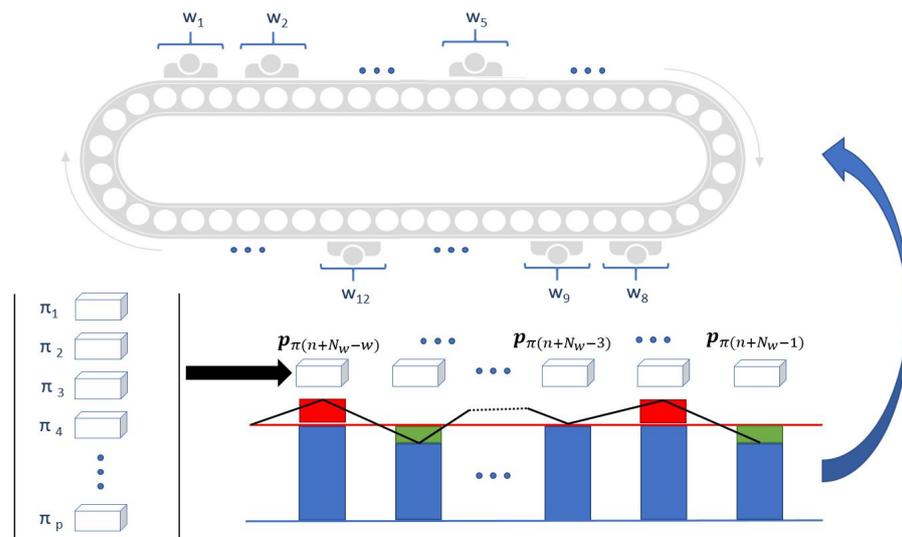

Figure 13. The scheme of the the open-station conveyor that is used to for wire-harness production.
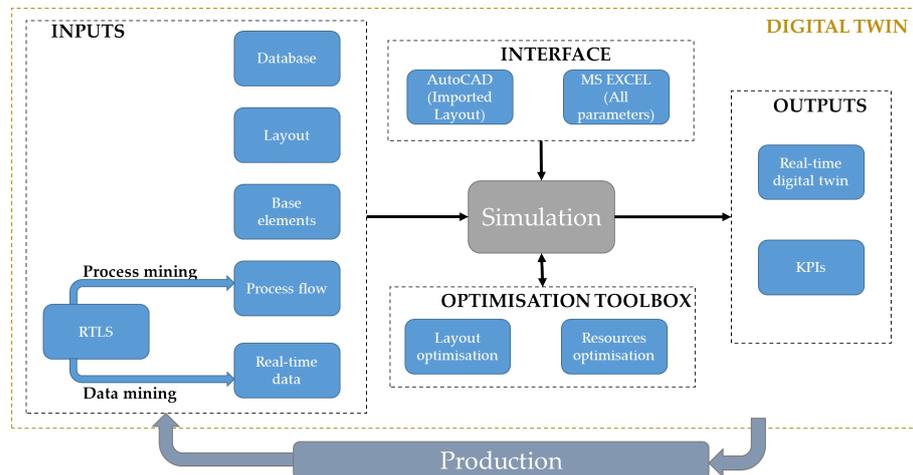

Figure 14. The developed framework of the real-time digital twin concept. The two main inputs of the simulation are the material flow (process model) and the real-time location system (RTLS)-based parameters of the model.

The key element of the digital twin is the simulator of the production line that has been developed in Siemens Plant Simulation Software. The discrete event simulator is generated by a program code taking into account the real-time information coming from the production line and the hierarchy of the process.

The optimisation and control of the production face challenges due to the complexity of production and the unpredictable nature of human activities. The cycle time of the

conveyor should be controlled as every operator can be delayed or work ahead and the conveyor should be stopped when the delay reaches a certain limit.

The main problem is that the complex simulation model cannot be applied in a real-time control scheme, so a control-relevant model should be extracted from the digital twin. As shown in Figure 15, we developed a model-based predictive controller (MPC) algorithm that utilises the distribution functions of the activity times and converts the problem into a simple linear model-based predictive controller [113]. The extracted model can be considered a surrogate model of the digital twin used in real-time optimisation.

The developed model predictive controller minimises the cycle time (represented as $u(k)$ at the k-th time instant) in a H_p prediction horizon, by determining a control sequence of length H_c $u^*(k) = [u(k), u(k+1), \ldots, u(k+H_c)]$ where H_c denotes the control horizon was formulated.

The cost function is formalized to minimise the cycle time which in turn minimizes any delay to the expected finishing times in Equation (5), which also optimizes the utilities of the operators and attempts to ensure a well-balanced workload.

$$\min_{u^*(k)} u^*(k)^T R u^*(k) \tag{5}$$

The optimisation problem of the model predictive controller is constrained to ensure that the control sequence seeks to avoid stoppages of the conveyor belt due to the accumulation of a delay.

As $u(k)$ denotes the cycle time set at the beginning of the kth cycle, the kth cycle starts at $t_c(k)$ and finishes at $t_c(k+1) = t_c(k) + u(k)$. According to this the end of cycle time in the prediction horizon can be calculated as $t_c(k) + \sum_{j=1}^{H_p} u(k+j-1)$.

The constrains represent the requirement that the $t_f(k+j|k)$ predicted finishing times should not exceed the cycle time over the c_{crit} time in the prediction horizon:

$$t_f(k+j|k) < t_c(k) + \sum_{i=1}^{j} u(k+i-1) + c_{crit}. \tag{6}$$

The resultant sets of equations can be represented in matrix form to define the quadratic optimisation problem along with the Equation (5).

$$A_R u^*(k) < b_R \tag{7}$$

where A_R is a lower triangular matrix with size H_pxH_c that sums up the u control signal over the prediction horizon and $b_R = [-t_f(k+1|k) \ldots, -t_f(k+H_p|k)]^T + t_c(k) + c_{crit}$ according to the rearranged form of the constraints

$$-\sum_{i=i}^{j} u(k+i-1) < t_c(k) + c_{crit} - t_f(k+j|k). \tag{8}$$

The concept has been validated in a reproducible simulation study in which the digital twin is used to check the performance of the controller. As the performance of the controller is illustrated in Figure 16. The performance of controller have been analysed and presented in detail in [113], in this paper we discuss the results from viewpoint of the model development.

The example illustrated the following benefits of surrogate modelling:

- Surrogate models should be applied when the discrete event simulator of digital twins cannot be utilised directly in control and optimisation.
- When the model is not linearisable, surrogate models can be extracted from the simulators by Monte Carlo simulation. In this case study, the distributions of the activity times were evaluated and approximated by fuzzy models.

- The resulted models can be further simplified to support problem-specific utilization. In this paper, a simple linear model was extracted from the fuzzy model that represents the activity times in a given confidence.

  The example also highlights the possible problems of the utilization of surrogate models:

- The modeller should validate not only the accuracy of the extracted model on a training and validation dataset, but also the performance of the application should be carefully analysed as most of the control and optimisation algorithms need extrapolation from the models.

- The surrogate models should not be oversimplified. Finding the optimal model complexity needs details analysis of the system and the application tasks. In this situation, the principle of Occam's Razors should be adapted; one should select the solution with the least complexity that makes predictions suitable for the given task.
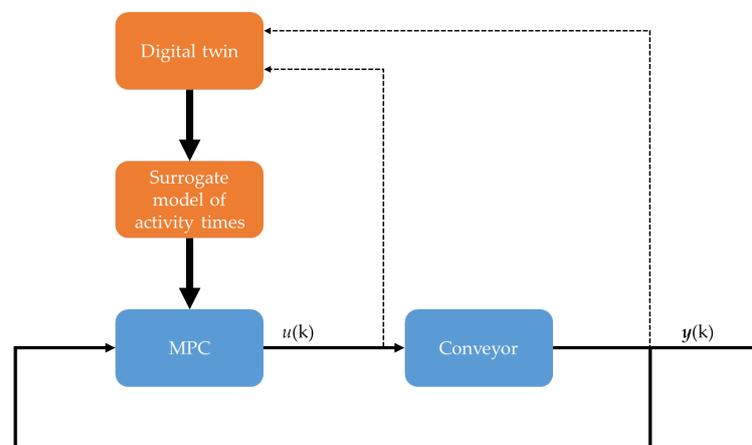

Figure 15. The scheme of the proposed surrogate model-based predictive controller (MPC). The model of the MPC is updated based on the information extracted from the digital twin.
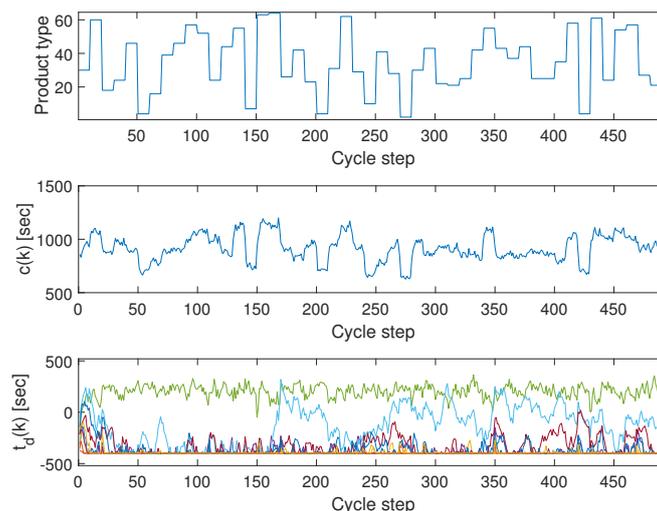

Figure 16. Production of $N = 500$ products in batches by applying model predictive control with regard to the cycle time of the wire-harness production line, $H_p = 5$, $H_c = 3$, $\alpha = 0.1$. Control of the cycle time maximizes productivity, so performance is enhanced by 20% in this complex problem compared when the cycle time was constant. The bottom plot shows the time delay $(t_d(k))$ at every workstation where the colours represent the operators.

6. Conclusions

In this paper, arguments supporting the suitability of surrogate models in digital twins are presented. In Figure 3 it was demonstrated how surrogate models link together physics-based modelling, data-driven modelling and hybrid solutions that lead to the combination of physics-based modelling and data-driven modelling with big data approaches. A comparison showing the advantages and disadvantages of each model type is presented in Table 3. Based on these characteristics, it can be stated that the use of surrogate models has many advantages, e.g., although of black-box-type, they still reflect some of the physics; once the models have been trained, they become stable for making predictionsinferences; errorsuncertainties can be bounded and estimated; they are less susceptible to bias. The main argument for applying surrogates in digital twins is that their ultimate computational demand could be significantly smaller than that of a detailed process simulation and even a Computational Fluid Dynamics (CFD)-based simulation.

Basically, two difficulties with the application of surrogate models can be identified, one of them is the huge amount of data needed for model building and the other is that they need continuous maintenance over the whole life cycle of the model. These are explained in detail in the following:

- Data requirement: As shown in Section 3, collecting data of adequate quantity and quality is the key component of suitable surrogate model building. These data may come directly from physical reality (measurements) or even from an adequate simulation of the system (process simulator, CFD-based simulator, etc.).
- Maintenance requirements: Since the generalisation capability of surrogate models can be limited, moreover, processes cannot really operate exactly in a true steady state, the continuous maintenance of surrogate models is considered a decisive task. Application-oriented validation should fit into the whole life cycle of the model as was discussed in Section 3.3.

In Section 3, the steps of surrogate model building and the commonly used surrogate models were presented in detail. The application limits and advantages of all types of models were described. Based on that section and the application areas introduced in the introduction and Section 4, it is obvious that surrogate models could be effectively applied in digital twins. The fundamental question is what type of surrogate models would be appropriate in particular cases. By examining the types of surrogate models used in each application, it can be concluded that the type of surrogate model used does not basically depend on the application task, rather on the kind of system modelled. In Section 3.2, two papers were cited in which different algorithms were presented to automate the selection of applicable surrogate model types for the current task and system.

Based on the discussed considerations, it is recommended to keep in mind the following steps when applying surrogate models in digital twins:

1. Definition of the physical process (manufacturing system) to be modelled.
2. Determination of the application task (optimization, control, scheduling, etc.).
3. Selection of the suitable modelling approach (physics-based, data-driven or hybrid) based on the knowledge of and information available on the underlying system.
4. Collection of appropriate quantity and quality data (depending on the modelling approach).
5. Model building (selection of the model type and determination of the model parameters) and model validation.
6. Regular supervision of the application environment and model quality.
7. Maintenance of the model (modification of the model parameters or structure if necessary).

We hope that this paper will serve as a guideline for the development of digital twins that utilise surrogate models.

In the paper we not only highlighted the benefits of surrogate models, but we also discussed that the utilisation of simplified models may have multiple disadvantages, e.g., it is not straightforward what accuracy should surrogate model have to be considered

suitable, in which regions of the operation will be the model utilised and what is the extrapolation and generalisation power of the extracted models.

To mitigate the highlighted drawbacks of the utilisation of surrogate models, we defined the following research topics for the future:

- Development of surrogate models for process safety: fusion of measured and simulated data for the modelling of process behaviour far from standard process conditions (e.g., runaway state, malfunction).
- Development of automated testing and validation tools that consider the application-specific preferences of the surrogate models.
- Development of automated tools that identify surrogate models based on process data and simulation models and determine the optimal model complexity.
- Study how semi-mechanistic models can be identified and utilized in the framework of digital twins.

Author Contributions: Conceptualization and supervision, J.A. and S.N.; writing—original draft preparation, T.C., Á.B. and J.A.; writing—review and editing, Á.B., T.C., S.N. and J.A.; funding acquisition, J.A. and S.N. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grieves, M. Digital twin: Manufacturing excellence through virtual factory replication. *White Pap.* **2014**, *1*, 1–7.
2. Almada-Lobo, F. The Industry 4.0 revolution and the future of Manufacturing execution systems (MES). *J. Innov. Manag.* **2015**, *3*, 16–21. [CrossRef]
3. Hicks, B. Industry 4.0 and Digital Twins: Key Lessons from NASA. 2020.Available online: https://www.thefuturefactory.com/blog/24 (accessed on 12 November 2020).
4. Jbair, M.; Ahmad, B.; Mus' ab H, A.; Harrison, R. Industrial cyber physical systems: A survey for control-engineering tools. In Proceedings of the 2018 IEEE Industrial Cyber-Physical Systems (ICPS), Saint Petersburg, Russia, 15–18 May 2018; pp. 270–276.
5. Wang, Q.; Jiao, W.; Wang, P.; Zhang, Y. Digital twin for human-robot interactive welding and welder behavior analysis. *IEEE/CAA J. Autom. Sin.* **2020**, *8*, 334–343. [CrossRef]
6. Rasheed, A.; San, O.; Kvamsdal, T. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access* **2020**, *8*, 21980–22012. [CrossRef]
7. Nentwich, C.; Varela, C.; Engell, S. Optimization of chemical processes applying surrogate models for phase equilibrium calculations. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.
8. Lv, Z.; Wang, L.; Han, Z.; Zhao, J.; Wang, W. Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 838–849. [CrossRef]
9. Burnak, B.; Diangelakis, N.A.; Katz, J.; Pistikopoulos, E.N. Integrated process design, scheduling, and control using multiparametric programming. *Comput. Chem. Eng.* **2019**, *125*, 164–184. [CrossRef]
10. Shirmohammadi, R.; Aslani, A.; Ghasempour, R.; Romeo, L.M. CO2 Utilization via Integration of an Industrial Post-Combustion Capture Process with a Urea Plant: Process Modelling and Sensitivity Analysis. *Processes* **2020**, *8*, 1144. [CrossRef]
11. Tsay, C.; Baldea, M. Integrating production scheduling and process control using latent variable dynamic models. *Control Eng. Pract.* **2020**, *94*, 104201. [CrossRef]
12. Pereira, L.M.; Vega, L.F. A systematic approach for the thermodynamic modelling of CO2-amine absorption process using molecular-based models. *Appl. Energy* **2018**, *232*, 273–291. [CrossRef]
13. Egorov, V.; Novakovskii, A.; Zdrachek, E. Modeling of the effect of diffusion processes on the response of ion-selective electrodes by the finite difference technique: Comparison of theory with experiment and critical evaluation. *J. Anal. Chem.* **2017**, *72*, 793–802. [CrossRef]
14. Peitz, S.; Dellnitz, M. A survey of recent trends in multiobjective optimal control—Surrogate models, feedback control and objective reduction. *Math. Comput. Appl.* **2018**, *23*, 30. [CrossRef]
15. Tabatabaei, M.; Hakanen, J.; Hartikainen, M.; Miettinen, K.; Sindhya, K. A survey on handling computationally expensive multiobjective optimization problems using surrogates: Non-nature inspired methods. *Struct. Multidiscip. Optim.* **2015**, *52*, 1–25. [CrossRef]

16. Melab, N.; Gmys, J.; Korosec, P.; Chakroun, I. Synergy between parallel computing, optimization and simulation. *J. Comput. Sci.* **2020**, *44*, 101168. [CrossRef]

17. Borth, M.; Verriet, J.; Muller, G. Digital Twin Strategies for SoS 4 Challenges and 4 Architecture Setups for Digital Twins of SoS. In Proceedings of the 2019 14th Annual Conference System of Systems Engineering (SoSE), Anchorage, AK, USA, 19–22 May 2019; pp. 164–169.

18. Moher, D.; Shamseer, L.; Clarke, M.; Ghersi, D.; Liberati, A.; Petticrew, M.; Shekelle, P.; Stewart, L.A.; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **2015**, *4*, 1. [CrossRef] [PubMed]

19. Zhu, W.; Chebeir, J.; Romagnoli, J.A. Operation optimization of a cryogenic NGL recovery unit using deep learning based surrogate modeling. *Comput. Chem. Eng.* **2020**, 106815. . [CrossRef]

20. Duchêne, P.; Mencarelli, L.; Pagot, A. Optimization approaches to the integrated system of catalytic reforming and isomerization processes in petroleum refinery. *Comput. Chem. Eng.* **2020**, *141*, 107009. [CrossRef]

21. Rall, D.; Schweidtmann, A.M.; Kruse, M.; Evdochenko, E.; Mitsos, A.; Wessling, M. Multi-scale membrane process optimization with high-fidelity ion transport models through machine learning. *J. Membr. Sci.* **2020**, 118208. [CrossRef]

22. Kim, S.H.; Boukouvala, F. Surrogate-Based Optimization for Mixed-Integer Nonlinear Problems. *Comput. Chem. Eng.* **2020**, *140*, 106847. [CrossRef]

23. Ali, W.; Khan, M.S.; Qyyum, M.A.; Lee, M. Surrogate-assisted modeling and optimization of a natural-gas liquefaction plant. *Comput. Chem. Eng.* **2018**, *118*, 132–142. [CrossRef]

24. Zhong, W.; Qiao, C.; Peng, X.; Li, Z.; Fan, C.; Qian, F. Operation optimization of hydrocracking process based on Kriging surrogate model. *Control Eng. Pract.* **2019**, *85*, 34–40. [CrossRef]

25. Christelis, V.; Kopsiaftis, G.; Mantoglou, A. Performance comparison of multiple and single surrogate models for pumping optimization of coastal aquifers. *Hydrol. Sci. J.* **2019**, *64*, 336–349. [CrossRef]

26. Candelieri, A.; Archetti, F. Sequential model based optimization with black-box constraints: Feasibility determination via machine learning. In Proceedings of the AIP Conference Proceedings, Leiden, The Netherlands, 18–21 September 2018; AIP Publishing LLC: Leiden, The Netherlands, 2019; Volume 2070, p. 020010.

27. Keßler, T.; Kunde, C.; McBride, K.; Mertens, N.; Michaels, D.; Sundmacher, K.; Kienle, A. Global optimization of distillation columns using explicit and implicit surrogate models. *Chem. Eng. Sci.* **2019**, *197*, 235–245. [CrossRef]

28. Deng, Y.; Yu, B.; Sun, D. Multi-objective optimization of guide vanes for axial flow cyclone using CFD, SVM, and NSGA II algorithm. *Powder Technol.* **2020**, *373*, 637–646. [CrossRef]

29. Beykal, B.; Boukouvala, F.; Floudas, C.A.; Pistikopoulos, E.N. Optimal design of energy systems using constrained grey-box multi-objective optimization. *Comput. Chem. Eng.* **2018**, *116*, 488–502. [CrossRef]

30. Olofsson, S.; Mehrian, M.; Calandra, R.; Geris, L.; Deisenroth, M.P.; Misener, R. Bayesian multiobjective optimisation with mixed analytical and black-box functions: Application to tissue engineering. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 727–739. [CrossRef]

31. Nguyen, T.H.; Nong, D.; Paustian, K. Surrogate-based multi-objective optimization of management options for agricultural landscapes using artificial neural networks. *Ecol. Model.* **2019**, *400*, 1–13. [CrossRef]

32. Quirante, N.; Caballero, J.A. Optimization of a Sour Water Stripping Plant Using Surrogate Models. In *Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2016; Volume 38, pp. 31–36.

33. Brevault, L.; Balesdent, M.; Hebbal, A.; Patureau De Mirand, A. Surrogate model-based multi-objective MDO approach for partially Reusable Launch Vehicle design. In Proceedings of the AIAA Scitech 2019 Forum, San Diego, CA, USA, 7–11 January 2019; p. 0704.

34. Nentwich, C.; Engell, S. Application of surrogate models for the optimization and design of chemical processes. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 1291–1296.

35. Quirante, N.; Javaloyes-Antón, J.; Caballero, J.A. Hybrid simulation-equation based synthesis of chemical processes. *Chem. Eng. Res. Des.* **2018**, *132*, 766–784. [CrossRef]

36. Yang, H.; Hong, S.H.; ZhG, R.; Wang, Y. Surrogate-based optimization with adaptive sampling for microfluidic concentration gradient generator design. *RSC Adv.* **2020**, *10*, 13799–13814. [CrossRef]

37. Shi, H.; You, F. A novel adaptive surrogate modeling-based algorithm for simultaneous optimization of sequential batch process scheduling and dynamic operations. *AIChE J.* **2015**, *61*, 4191–4209. [CrossRef]

38. Dias, L.S.; Ierapetritou, M.G. Integration of planning, scheduling and control problems using data-driven feasibility analysis and surrogate models. *Comput. Chem. Eng.* **2020**, *134*, 106714. [CrossRef]

39. Omairey, S.L.; Dunning, P.D.; Sriramula, S. Multiscale surrogate-based framework for reliability analysis of unidirectional FRP composites. *Compos. Part B Eng.* **2019**, *173*, 106925. [CrossRef]

40. Rafiei, M.; Ricardez-Sandoval, L.A. Integration of design and control for industrial-scale applications under uncertainty: A trust region approach. *Comput. Chem. Eng.* **2020**, *141*, 107006. [CrossRef]

41. Wang, K.; Chen, J.; Xie, L.; Su, H. Decision making scheme of integration design and control under uncertainty for enhancing the economic performance of chemical processes with multiplicity behaviors. *Chem. Eng. Res. Des.* **2019**, *150*, 327–340. [CrossRef]

42. Rodič, B. Industry 4.0 and the new simulation modelling paradigm. *Organizacija* **2017**, *50*, 193–207. [CrossRef]

43. Kljajic, M.; Bernik, I.; Skraba, A. Simulation approach to decision assessment in enterprises. *Simulation* **2000**, *75*, 199–210. [CrossRef]

44. Rasheed, A.; San, O.; Kvamsdal, T. Digital Twin: Values, Challenges and Enablers. *arXiv* **2019**, arXiv:1910.01719.

45. Löcklin, A.; Müller, M.; Jung, T.; Jazdi, N.; White, D.; Weyrich, M. Digital Twin for Verification and Validation of Industrial Automation Systems–a Survey. In Proceedings of the 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria, 8–11 September 2020; Volume 1, pp. 851–858.

46. Aivaliotis, P.; Georgoulias, K.; Arkouli, Z.; Makris, S. Methodology for enabling digital twin using advanced physics-based modelling in predictive maintenance. *Procedia CIRP* **2019**, *81*, 417–422. [CrossRef]

47. Wang, X.; Wang, Y.; Tao, F.; Liu, A. New Paradigm of Data-Driven Smart Customisation through Digital Twin. *J. Manuf. Syst.* **2020**, *58*, 170–280.

48. Mencarelli, L.; Pagot, A.; Duchêne, P. Surrogate-based modeling techniques with application to catalytic reforming and isomerization processes. *Comput. Chem. Eng.* **2020**, *135*, 106772. [CrossRef]

49. Dimitrov, N.; Kelly, M.; Vignaroli, A.; Berg, J. From wind to loads: Wind turbine site-specific load estimation using databases with high-fidelity load simulations. *Wind Energ. Sci. Discuss* **2018**, *375*, 767–790. [CrossRef]

50. Slot, R.M.; Sørensen, J.D.; Sudret, B.; Svenningsen, L.; Thøgersen, M.L. Surrogate model uncertainty in wind turbine reliability assessment. *Renew. Energy* **2020**, *151*, 1150–1162. [CrossRef]

51. Li, M.; Wang, Z. Surrogate model uncertainty quantification for reliability-based design optimization. *Reliab. Eng. Syst. Saf.* **2019**, *192*, 106432. [CrossRef]

52. Bhosekar, A.; Ierapetritou, M. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput. Chem. Eng.* **2018**, *108*, 250–267. [CrossRef]

53. McBride, K.; Sundmacher, K. Overview of surrogate modeling in chemical process engineering. *Chem. Ing. Tech.* **2019**, *91*, 228–239. [CrossRef]

54. Kim, S.H.; Boukouvala, F. Machine learning-based surrogate modeling for data-driven optimization: A comparison of subset selection for regression techniques. *Optim. Lett.* **2019**, *14*, 1–22. [CrossRef]

55. Asher, M.J.; Croke, B.F.; Jakeman, A.J.; Peeters, L.J. A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* **2015**, *51*, 5957–5973. [CrossRef]

56. Razavi, S.; Tolson, B.A.; Burn, D.H. Review of surrogate modeling in water resources. *Water Resour. Res.* **2012**, *48*, 1–32. [CrossRef]

57. Chatterjee, T.; Chakraborty, S.; Chowdhury, R. A critical review of surrogate assisted robust design optimization. *Arch. Comput. Methods Eng.* **2019**, *26*, 245–274. [CrossRef]

58. Díaz-Manríquez, A.; Toscano, G.; Barron-Zambrano, J.H.; Tello-Leal, E. A review of surrogate assisted multiobjective evolutionary algorithms. *Comput. Intell. Neurosci.* **2016**, *2016*, 1–14. [CrossRef]

59. Bartz-Beielstein, T.; Zaefferer, M. Model-based methods for continuous and discrete global optimization. *Appl. Soft Comput.* **2017**, *55*, 154–167. [CrossRef]

60. Aversano, G.; Bellemans, A.; Li, Z.; Coussement, A.; Gicquel, O.; Parente, A. Application of reduced-order models based on PCA & Kriging for the development of digital twins of reacting flow applications. *Comput. Chem. Eng.* **2019**, *121*, 422–441.

61. Tabar, R.S.; Wärmefjord, K.; Söderberg, R. A new surrogate model–based method for individualized spot welding sequence optimization with respect to geometrical quality. *Int. J. Adv. Manuf. Technol.* **2020**, *106*, 2333–2346. [CrossRef]

62. Meng, F.; Li, Y.; Yuan, S.; Wang, W.; Zheng, Y.; Osman, M.K. Multiobjective Combination Optimization of an Impeller and Diffuser in a Reversible Axial-Flow Pump Based on a Two-Layer Artificial Neural Network. *Processes* **2020**, *8*, 309. [CrossRef]

63. Deng, T.; Ran, Y.; Yin, Y.; Chen, X.; Liu, P. Multi-objective optimization design of double-layered reverting cooling plate for lithium-ion batteries. *Int. J. Heat Mass Transf.* **2019**, *143*, 118580. [CrossRef]

64. Wang, X.; Xu, H.; Wang, J.; Song, W.; Wang, M. Multi-objective optimization of discrete film hole arrangement on a high pressure turbine end-wall with conjugate heat transfer simulations. *Int. J. Heat Fluid Flow* **2019**, *78*, 108428. [CrossRef]

65. Shi, J.; Chu, L.; Braun, R. A Kriging Surrogate Model for Uncertainty Analysis of Graphene Based on a Finite Element Method. *Int. J. Mol. Sci.* **2019**, *20*, 2355. [CrossRef]

66. Trucchia, A.; Egorova, V.; Pagnini, G.; Rochoux, M.C. On the merits of sparse surrogates for global sensitivity analysis of multi-scale nonlinear problems: Application to turbulence and fire-spotting model in wildland fire simulators. *Commun. Nonlinear Sci. Numer. Simul.* **2019**, *73*, 120–145. [CrossRef]

67. Vessaz, C.; Andolfatto, L.; Avellan, F.; Tournier, C. Toward design optimization of a Pelton turbine runner. *Struct. Multidiscip. Optim.* **2017**, *55*, 37–51. [CrossRef]

68. Ali, W.; Duong, P.L.T.; Khan, M.S.; Getu, M.; Lee, M. Measuring the reliability of a natural gas refrigeration plant: Uncertainty propagation and quantification with polynomial chaos expansion based sensitivity analysis. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 103–117. [CrossRef]

69. Yang, M.; Xiao, Z. POD-based surrogate modeling of transitional flows using an adaptive sampling in Gaussian process. *Int. J. Heat Fluid Flow* **2020**, *84*, 108596. [CrossRef]

70. Nentwich, C.; Engell, S. Surrogate modeling of phase equilibrium calculations using adaptive sampling. *Comput. Chem. Eng.* **2019**, *126*, 204–217. [CrossRef]

71. Dias, L.; Bhosekar, A.; Ierapetritou, M. Adaptive Sampling Approaches for Surrogate-Based Optimization. In *Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 47, pp. 377–384.

72. Jia, L.; Alizadeh, R.; Hao, J.; Wang, G.; Allen, J.K.; Mistree, F. A rule-based method for automated surrogate model selection. *Adv. Eng. Inform.* **2020**, *45*, 101123. [CrossRef]

73. Salem, M.B.; Tomaso, L. Automatic selection for general surrogate models. *Struct. Multidiscip. Optim.* **2018**, *58*, 719–734. [CrossRef]

74. Bauer, M.; Craig, I.K. Economic assessment of advanced process control—A survey and framework. *J. Process. Control* **2008**, *18*, 2–18. [CrossRef]

75. Mayo, S.M.; Rhinehart, R.R.; Madihally, S.V. *Advanced Process Control Capital Decisions Must Include Operations Planning for APC Maintenance*; School of Chemical Engineering, Oklahoma State University: Stillwater, OK, USA, 2020.

76. Lu, B.; Chiang, L. Semi-supervised online soft sensor maintenance experiences in the chemical industry. *J. Process. Control* **2018**, *67*, 23–34. [CrossRef]

77. Chen, K.; Castillo, I.; Chiang, L.H.; Yu, J. Soft sensor model maintenance: A case study in industrial processes. *IFAC Pap.* **2015**, *48*, 427–432.

78. Studer, S.; Bui, T.B.; Drescher, C.; Hanuschkin, A.; Winkler, L.; Peters, S.; Mueller, K.R. Towards CRISP-ML (Q): A Machine Learning Process Model with Quality Assurance Methodology. *arXiv* **2020**, arXiv:2003.05155.

79. Lv, L.; Song, X.; Sun, W. Modify Leave-One-Out Cross Validation by Moving Validation Samples around Random Normal Distributions: Move-One-Away Cross Validation. *Appl. Sci.* **2020**, *10*, 2448. [CrossRef]

80. Dammak, K.; El Hami, A. Multi-objective reliability based design optimization using Kriging surrogate model for cementless hip prosthesis. *Comput. Methods Biomech. Biomed. Eng.* **2020**, *23*, 854–867. [CrossRef]

81. Le, T.T. Surrogate Neural Network Model for Prediction of Load-Bearing Capacity of CFSS Members Considering Loading Eccentricity. *Appl. Sci.* **2020**, *10*, 3452. [CrossRef]

82. Ye, Y.; Wang, Z.; Zhang, X. An optimal pointwise weighted ensemble of surrogates based on minimization of local mean square error. *Struct. Multidiscip. Optim.* **2020**, *62*, 529–542. [CrossRef]

83. Seong, Y.; Park, C.; Choi, J.; Jang, I. Surrogate Model with a Deep Neural Network to Evaluate Gas–Liquid Flow in a Horizontal Pipe. *Energies* **2020**, *13*, 968. [CrossRef]

84. Kang, K.; Qin, C.; Lee, B.; Lee, I. Modified screening-based Kriging method with cross validation and application to engineering design. *Appl. Math. Model.* **2019**, *70*, 626–642. [CrossRef]

85. Kumar, P.M.; Seo, J.; Seok, W.; Rhee, S.H.; Samad, A. Multi-fidelity optimization of blade thickness parameters for a horizontal axis tidal stream turbine. *Renew. Energy* **2019**, *135*, 277–287. [CrossRef]

86. Alizadeh, R.; Jia, L.; Nellippallil, A.B.; Wang, G.; Hao, J.; Allen, J.K.; Mistree, F. Ensemble of surrogates and cross-validation for rapid and accurate predictions using small data sets. *Artif. Intell. Eng. Des. Anal. Manuf.* **2019**, *33*, 484–501. [CrossRef]

87. Palar, P.S.; Shimoyama, K. On efficient global optimization via universal Kriging surrogate models. *Struct. Multidiscip. Optim.* **2018**, *57*, 2377–2397. [CrossRef]

88. Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital Twin: Enabling Technologies, Challenges and Open Research. *IEEE Access* **2020**, *8*, 108952–108971. [CrossRef]

89. Gigi, K.E. Digital twins for greater insights. *WaterWorld* **2020**, *36*, 11.

90. Yuan, Z.; Qin, W.; Zhao, J. Smart Manufacturing for the Oil Refining and Petrochemical Industry. *Engineering* **2017**, *3*, 179–182. [CrossRef]

91. Lee, J.; Cameron, I.; Hassall, M. Improving process safety: What roles for Digitalization and Industry 4.0? *Process. Saf. Environ. Prot.* **2019**, *132*, 325–339. [CrossRef]

92. Müller, J.M. Antecedents to digital platform usage in Industry 4.0 by established manufacturers. *Sustainability* **2019**, *11*, 1121. [CrossRef]

93. Örs, E.; Schmidt, R.; Mighani, M.; Shalaby, M. A Conceptual Framework for AI-based Operational Digital Twin in Chemical Process Engineering. In Proceedings of the 2020 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), Cardiff, UK, 15–17 June 2020; pp. 1–8. [CrossRef]

94. Xenos, D.P.; Cicciotti, M.; Kopanos, G.M.; Bouaswaig, A.E.; Kahrs, O.; Martinez-Botas, R.; Thornhill, N.F. Optimization of a network of compressors in parallel: Real Time Optimization (RTO) of compressors in chemical plants—An industrial case study. *Appl. Energy* **2015**, *144*, 51–63. [CrossRef]

95. Lie, B. Surrogate and Hybrid Models for Control. In Proceedings of the 60th SIMS Conference, Västerås, Sweden, 12–16 August 2019; Volume 170, pp. 1–8.

96. Chu, Y.; You, F. Model-based integration of control and operations: Overview challenges advances and opportunities. *Comput. Chem. Eng.* **2015**, *83*, 2–20. [CrossRef]

97. He, R.; Chen, G.; Dong, C.; Sun, S.; Shen, X. Data-driven digital twin technology for optimized control in process systems. *ISA Trans.* **2019**, *95*, 221–234. [CrossRef] [PubMed]

98. Sales-Cruz, M.; Cameron, I.; Gani, R. Tennessee Eastman Plant-wide Industrial Process Challenge Problem. In *Product and Process Modelling*; Elsevier: Amsterdam, The Netherlands, 2011; pp. 273–303.

99. Janošovský, J.; Danko, M.; Labovský, J.; Jelemenský, L. Software approach to simulation-based hazard identification of complex industrial processes. *Comput. Chem. Eng.* **2019**, *122*, 66–79. [CrossRef]

100. Kummer, A.; Varga, T. Process simulator assisted framework to support process safety analysis. *J. Loss Prev. Process. Ind.* **2019**, *58*, 22–29. [CrossRef]

101. Wu, J.; Lind, M.; Jørgensen, S.B.; Jensen, N.; Sin, G. Functional Modeling for Process Safety. In Proceedings of the 2nd International Workshop on Functional Modelling for Design and Operation of Engineering Systems and Infrastructures, Okayama, Japan, 30–31 March 2015.

102. Wu, J.; Lind, M.; Zhang, X.; Jørgensen, S.; Sin, G. Validation of a functional model for integration of safety into process system design. In *Computer Aided Chemical Engineering*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 37, pp. 293–298.

103. Bevilacqua, M.; Bottani, E.; Ciarapica, F.E.; Costantino, F.; Di Donato, L.; Ferraro, A.; Mazzuto, G.; Monteriù, A.; Nardini, G.; Ortenzi, M.; et al. Digital twin reference model development to prevent operators' risk in process plants. *Sustainability* **2020**, *12*, 1088. [CrossRef]

104. Madni, A.M.; Madni, C.C.; Lucero, S.D. Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Systems* **2019**, *7*, 7. [CrossRef]

105. Qi, Q.; Tao, F. Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access* **2018**, *6*, 3585–3593. [CrossRef]

106. Karve, P.M.; Guo, Y.; Kapusuzoglu, B.; Mahadevan, S.; Haile, M.A. Digital twin approach for damage-tolerant mission planning under uncertainty. *Eng. Fract. Mech.* **2020**, *225*, 106766. [CrossRef]

107. Thiers, G.; Sprock, T.; McGinnis, L.; Graunke, A.; Christian, M. Automated production system simulations using commercial off-the-shelf simulation tools. In Proceedings of the 2016 Winter Simulation Conference (WSC), Washington, DC, USA, 11–14 December 2016, pp. 1036–1047.

108. Rodič, B.; Kanduč, T. Leveraging Digital Twin Technology in Model-Based Systems Engineering. *Int. J. Math. Model. Methods Appl. Sci.* **2015**, *9*, 1–13.

109. Jain, S.; Lechevalier, D. Standards based generation of a virtual factory model. In Proceedings of the 2016 Winter Simulation Conference (WSC), Washington, DC, USA, 11–14 December 2016; pp. 2762–2773.

110. Kirchhof, P. Automatically generating flow shop simulation models from SAP data. In Proceedings of the 2016 Winter Simulation Conference (WSC), Washington, DC, USA, 11–14 December 2016; pp. 3588–3589.

111. Ruppert, T.; Abonyi, J. Integration of real-time locating systems into digital twins. *J. Ind. Inf. Integr.* **2020**, *20*, 1–12. [CrossRef]

112. Ruppert, T.; Abonyi, J. Software Sensor for Activity-Time Monitoring and Fault Detection in Production Lines. *Sensors* **2018**, *18*, 2346. [CrossRef] [PubMed]

113. Tamas Ruppert, G.D.; Abonyi, J. Fuzzy activity time-based model predictive control of open-station assembly lines. *J. Manuf. Syst.* **2020**, *54*, 12–23. [CrossRef]