Proceedings Paper:
Munko, M.J., Vidmar, M., Ó Brádaigh, C.M. et al. (3 more authors) (2024) Do digital twins require physical simulations? A study of developing digital twins with varying reliance on physics-based models. In: Proceedings of 2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC). 2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC), 24-28 Jun 2024, Funchal, Portugal. Institute of Electrical and Electronics Engineers (IEEE) , pp. 1-8. ISBN 9798350362442

https://doi.org/10.1109/ice/itmc61926.2024.10794252

Do Digital Twins Require Physical Simulations? A Study of Developing Digital Twins with Varying Reliance on Physics-Based Models

Marek Jan Munko
School of Engineering
The University of Edinburgh
Edinburgh, United Kingdom
M.Munko@ed.ac.uk*

Matjaz Vidmar
School of Engineering
The University of Edinburgh
Edinburgh, United Kingdom
Matjaz.Vidmar@ed.ac.uk

Conchúr M. Ó Brádaigh
Faculty of Engineering
The University of Sheffield
Sheffield, United Kingdom
C.OBradaigh@sheffield.ac.uk

Fergus Cuthill
School of Engineering
The University of Edinburgh
Edinburgh, United Kingdom
Fergus.Cuthill@ed.ac.uk

Miguel Angel Valdivia Camacho
School of Engineering
The University of Edinburgh
Edinburgh, United Kingdom
MA.Valdivia@ed.ac.uk

Sergio Lopez Dubon
School of Engineering
The University of Edinburgh
Edinburgh, United Kingdom
Sergio.LDubon@ed.ac.uk

***Abstract*— Digital Twins (DTs) rely on the alignment of a physical asset and its digital representation to provide a useful insight into its operation. By bridging the gap between a physical entity and process data collected in real time, DTs have been at the forefront of the most recent wave of industrial digitisation. Due to the great variety of assets which can be modelled with a DT, there is no standardised way to develop one. Successful DT implementations found in literature vary from being strongly dependent on physical simulations, to being solely data-driven. Therefore, the consideration of the DT design process addressed in this work is necessary for incorporating physics-based models. To this end, we group DTs into three categories, namely: purely data-driven DTs, physics-based DTs and physics-informed DTs. We choose representative cases from literature to explain their distinguishing features, describe the intrinsic differences, as well as draw conclusions on their advantages and limitations. We then present the case study of developing a DT for FastBlade, a facility for regenerative testing of tidal turbine blades. We discuss the challenges and opportunities associated with the facility to assess the suitability of each of the three distinct DT development strategies. The complexity of the energy-recovery system, unknown asset internals, as well as the broad scope of the sensing and logging network are identified to be the key decisive factors. Finally, we suggest developing a physics-informed DT for FastBlade as the optimal route.**

Keywords—* *Digital Twins, Data-Driven Engineering, Machine Learning, Physical Models

I. Introduction

A. *Digital Twins – Definition and Implementation*

A Digital Twin (DT) is a coupling of a physical asset and its corresponding model in the digital domain. While the concept of using a virtual model to improve the operation of an asset or an engineering process is not novel, the key distinguishing feature of DTs is the bi-directional flow of data between the physical and virtual twins [1]. Unlike a digital shadow (DS), the DT not only updates together with the state of the physical asset, but it can also have a direct impact on the operation of its real-word counterpart. The influence of the DT on the physical asset can either require human intervention, or be fully-automated [2]. A schematic DT model is presented in Fig. 1. Another important feature of many DT applications is its ability to operate close to real-time conditions [3]. By seamlessly transferring data between the physical and digital assets, competitive advantage can be achieved across a variety of fields, including production [4], [5], [6], hydraulic systems [7], [8] and modelling physical phenomena [9]. This is sometimes referred to as moving towards "cyber-physical" systems operation and is the cornerstone of a system-of-systems approach trying to integrate DTs into a seamless workflow process at a facility or factory [10].

The major benefits of incorporating DTs into production and operation pipelines include improved asset monitoring and management, enhanced simulation capabilities, increased efficiency, as well as better planning and analysis [11]. Recent years have brought about a unified view on what characterises a DT. However, due to their great dependence on a particular asset modelled, there does not exist a unified approach to developing DTs.

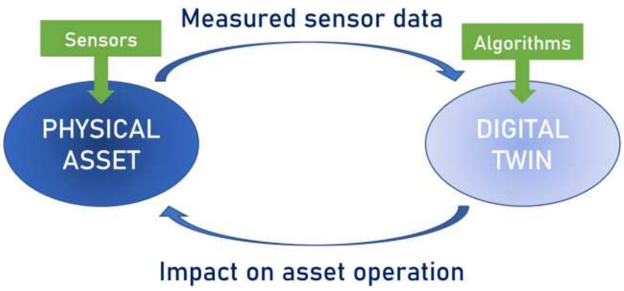

Fig. 1. Schematic interconnection between a physical asset and its DT. Source: authors, based on [1].

B. *Development of a Digital Twin*

Since any industrial asset can have its DT, each use case will result in a unique set of challenges. The variation in asset complexity, its operation environment and availability of historic operation data makes it difficult to adopt a common DT development framework. Therefore, it is possible to find successful DT implementations which will be seemingly completely different from one another.

Addressing the fundamental matter of *what it takes to create a digital twin*, soon leads to one of the most important design dilemmas, namely – the need for a physical simulation. The recent developments of simulation packages have made modelling physical processes unprecedently accurate. Complex physical models constitute an indispensable tool in areas of science and engineering which are too many to mention. However, creating a physics-based model comes with an upfront development cost and running it is often associated with a high computational demand. Therefore, the study of different industrial applications shows the use cases of both physics-based and purely data-driven DTs. The aim of this work is to introduce the inherent trade-offs between DTs with varying reliance on physical models in the light of relevant literature and the case study of FastBlade, a tidal turbine blade testing facility. We then aim to discuss the considerations for deciding on a given DT development strategy for FastBlade.

C. Paper Outline and Methodology

In this paper we first aim to review the existing systematic review studies which have been carried out to characterise the nature of DTs. We then propose a practical division of DTs into three categories and consider the gap which has been identified in the process of DT development and which relates to the use of physical knowledge of the modelled system. We review specific use cases from each of the suggested categories to understand the reasons for the design choices made. We then draw conclusions on the extent of the use of physics-based modelling based on available resources and desired DT functionality. The findings are applied to a particular industrial use case and the insights into the choice of the suitable DT development strategy are shared. Finally, we use the knowledge gained to propose a generalised strategy for choosing the optimal DT development plan. The division of DTs into the suggested categories is regarded as a helpful tool in the process of systemising knowledge about building DTs, rather than an unfailing method of classifying mature DTs. Therefore, this work does not constitute a contribution in the area of DT ontology, but rather focuses on deriving a particular heuristic, applying it to a specific industrial application and generalising the learning outcomes.

II. Literature Review

In recent years, considerable research has been carried out to characterise DTs. Table 1 presents five extensive systematic reviews of DTs, together with their main contributions.

TABLE I. THE SUMMARY OF SELECTED DT REVIEWS

Ref.	Title	Insight
[12]	Digital Twin applications toward Industry 4.0: A Review	Presents DTs in the light of Industry 4.0, discusses their features and presents a list of the most common applications.
[13]	Digital twin paradigm: A systematic literature review	Provides an up-to-date picture of the main DT components, features and associated problems, considering the ongoing research in various application domains.
[14]	Characterising the Digital Twin: A systematic literature review	Provides DT characterisation, identifies knowledge gaps and indicates the areas of future research.
[15]	A systematic review of digital twin about physical entities, virtual models, twin data, and applications	Provides DT definition, presents an overview of the DT components, characteristics and applications in the light of the current research and suggests future recommendations.
[16]	Systematic review of digital twin technology and applications	Explains the DT concept and definition, provides a historical context, suggests a conceptual DT model and discusses current challenges and future development.

The review papers listed in Table 1 focus on establishing the most suitable definition of DTs, list their applications and discuss the future directions of research. Therefore, they should be referred to for the more in-depth study of the DT specifications. However, as it was stated in the introduction, DTs lack a common development strategy. This might be particularly important for deciding on the extent to which physics-based models need to be integrated into them, since physical modelling requires substantial resources. In the following subsections, the DTs found in literature are divided into three classes to help systemise the conclusions drawn.

A. Division of Digital Twins and Nomenclature

Special care must be exercised when differentiating between different kinds of DTs based on their dependence on physical models. In essence, all DTs are data-driven, as in line with their definition, DTs rely on a continuous, bi-directional flow of information (i.e., data). What is more, although not all DTs seem to strictly rely on *physical models*, collection of physical data over time at pre-defined system locations may also constitute a *physical model* in the broader sense. Although every DT may show certain features of hybridisation, in this work we refer to three different classes of DTs, based on the extent to which they rely on physics-based models, namely:

- Purely *data-driven digital twins* (*DDDTs*), which rely solely on statistical correlations;
- *Physics-based digital twins* (*PBDTs*), which use a comprehensive physics-based simulation modelling entire asset operation;
- *Physics-informed digital twins* (*PIDTs*), which rely on a limited physical insight into the operation of an asset.

The division is made to distinguish between these classes of DTs to allow for generalisation and drawing relevant conclusions. The key differences between these DT types will be further discussed in the following sections.

B. Nature of the Problem

Although the question of necessity and scope of physics-based simulations in DTs is universal, only the publication by Erikstad [17] is found to discuss differences between PBDTs and DDDTs directly. The author observes the difference between structural physics-based models, which rely on finite element analysis (FEA), and machine learning (ML) models, which are based on statistics and data science to monitor, diagnose and predict system output based on the input data streams. Comparing the two approaches, it is concluded that though fundamentally different, they each offer unique benefits and are often complementary to each other. Erisktad also provides examples of using statistical methods to improve performance of a DT by incorporating computational offloading and surrogate models. The aim of this literature review is to deepen the understanding of the differences between DDDTs, PBDTs and PIDTs, and to find criteria facilitating the choice of a suitable DT development strategy.

C. Purely Data-Driven Digital Twins

DTs incorporating no physics-based models are almost universally found in systems whose operation is easily divisible into states and events, resulting in discrete process outcomes, such as *success* or *failure*. DDDTs are commonly found in assembly line processes, where a number of distinct stages needs to be completed to output a desired product.

Resman et al. [18] develop a framework for DDDT development for discrete manufacturing systems. They validate their work using a process in which a robotic arm inserts blocks into pallets to create desired patterns. The DT is capable of measuring the start and end time of the process and classifies whether the assembled output matches the desired template. Friedrich et al. [19] describe a case study of an assembly line producing parts for a quadcopter drone. The work showcases the development of a high-fidelity DDDT, integrated into the pilot assembly line to continuously monitor the process. The DT tracks the assembly line failures and by using statistical methods, outputs a reliability function for each asset. Both publications demonstrate the success of implementing DDDTs into a discrete production system, providing valuable insight into the probability of a successful process completion. However, the nature of the solution limits the insight into a problem once it has occurred. The solutions developed do not provide information about the severity of a failure or suggestions how an asset can be fixed.

The publication by Vishnu et al. [20] describes the development of a DDDT for predicting key performance indicators (KPIs) in a CNC machining process, namely energy and surface roughness. The critical cutting parameters for optimisation are identified using a correlation matrix and are used to train predictive models. A support vector machine (SVM), a fully connected deep neural network (FCDNN) and a gaussian process regression (GPR) are model types used for KPI prediction. The predicted values are communicated to the operator, who can manually adjust feed rate and spindle speed and see the predicted power product quality and associated energy consumption. The DDDT developed bypasses lack of physical knowledge about the system by using an extensive training dataset, at the same time providing just a *black-box* image of the entire process.

D. Physics-Based Digital Twins

Considering the reliance of DTs on physics-based simulations, PBDTs appear to be on the other side of the spectrum relative to the examples described in the previous section, as they stem from a comprehensive physical representation of an asset. Rituraj and Scheidl [7] describe the development of a DT for a counterbalance valve (CBV), based on its physical model. Experimental data are obtained under various operational conditions and is subsequently used for parameter identification to match the model's output with empirical results. Thanks to the physical relationships described, the model is capable of predicting outputs of an asset in operating conditions beyond those used for parameter identification. The physics-based model also offers valuable insight into the valve's operation. However, expert knowledge of the CBV topology is a prerequisite for accurate physical model development. Anda et al. [8] describe the development of a control valve. Their aim is to predict the local valve flow coefficient at various operation positions of the valve. The methodology relies on a computational fluid dynamics (CFD) model, which is calibrated by running a real valve in the laboratory. Subsequently, an artificial neural network (ANN) is used to link chosen operation parameters with the valve flow coefficient outputs, providing a satisfactory training accuracy. The approach combines the use of a physics-based model with ML. The calibrated CFD model serves multiple purposes, including behaviour prediction and ease of synthetic data generation. Similar to Rituraj and Scheidl [7], the detailed knowledge of asset internals is demonstrated by the authors in the development of a physics-based model.

The work done by Kapteyn et al., presented in [21], [22] shows the development of a DT for a 12-ft wingspan unmanned aerial vehicle (UAV). The component-based approach is acquired by producing reduced order models of the asset to generate structural models suitable for real-time deployment. A plurality of physical models is developed for the UAV, representing a variety of degradation levels of the asset. Based on the sensor readings, the most suitable model is chosen in real time to best optimise the mission trajectory of the UAV. The authors succeed in creating models which are both accurate and computationally-feasible. However, as it is the case with any high-fidelity model, detailed knowledge of the behaviour of the asset is required. In another work, Kapteyn et al. [23] offer a broader insight into the concept of interaction between the physical and digital works. Basing their idea on a probabilistic graphical model, they create an abstraction of a coupled asset-twin dynamical system. This allows them to use concepts from areas such as Bayesian statistics and control theory to generalise the corelation between observed quantities (i.e., deterministic values) and estimated quantities (usually characterised by a probability distribution), making the problem of DT development scalable. The probabilistic model can be used for tailoring a computational model, to reflect the individual characteristics of a physical asset with great fidelity. Although acquiring the probabilistic approach improves the performance of a DT, the suggested method still relies on the existence of pre-developed physical models, which in the case of the UAV are finite-element structural models.

E. Physics-Informed Digital Twins

Somewhere in-between the DDDTs and PDDTs are the Physics-Informed Digital Twins (PIDTs). These rely heavily on data-derived models, which are, however, constrained using partial physical modelling of key parameters, even though a comprehensive physical model is not known or possible to develop.

For example, similar to the work presented by Kapteyn et al. in [23], Møller et al. [24] consider a probabilistic approach to DT development. The reasons for their decision result from the fact that the phenomena which they are modelling, namely ultrafiltration and microfiltration membrane separation, are characterised by associated uncertainties and complex interactions and do have a rigorous theoretical description. The authors propose Stochastic Greybox Modelling and Control (SGMC) based on stochastic differential equations describing certain physical phenomena known to occur in the filtration process and quantify various inherent uncertainties. The method serves as a means of data-driven future forecasting in a problem which is not represented by a conventional physical model, such as a CFD-based simulation. Although the solution provided by the authors is complex, they manage to prove the efficacy of the method and show that DTs can be developed for systems which are hard to describe accurately with known physical tools.

The paper by Zhang and Zhao [25] describes the process of developing a DT for a wind farm, which also bypasses the development of a comprehensive physical model. The key operational parameters of the windfarm identified in the publication are the energy output and the structural fatigue. Both parameters are determined by the spatiotemporal flow field around the turbines, which however, does not have an accurate physical model due to the limitations of current measurement, modelling and prediction tools. As a solution to the problem, available theoretical models based on Navier-Stokes equations and actuator disk method for wind turbine modelling are coupled with Lidar measurements and operational data of the turbines via a physics-informed ANN. High-fidelity models are compared against predictions made by the ANN to prove that even complex air dynamics phenomena are accurately reconstructed by the DT created. The publication presents how both physical and digital domains are linked using a physics-informed ANN, where information obtained from sparse Lidar data are used alongside a simplified Navier-Stokes module and a turbine module, informing the ANN about the wind aerodynamics and the wind turbine properties respectively. This way, the lack of an overarching physical model is circumvented in the system and satisfactory performance is achieved. A similar approach is described by Yucesan and Viana [26], who develop a PIDT to monitor fatigue of the main bearing of a wind turbine and estimate the product life. The physics-informed model quantifies output uncertainties and uses a smaller dataset relative to a DDDT. As a result, it is concluded that the DT reaches satisfactory performance and the impact of varying lubricant quality in the bearings is mitigated. The authors claim that the model is more reliable than a purely data-driven, black-box approach. However, the physical models need to follow linear algebra requirements dictated by the ANN structure and frequent bearing inspections are initially required to build up a representative dataset.

F. *Literature Review Findings*

The literature review describes numerous case studies of DT development, grouped into three categories based on the extent to which they rely on physics-based models. The aim of this subsection is to draw conclusions from the study and formalise them to aid the decision-making process for choosing the most optimal DT development strategy.

It is evident that DDDTs are suitable for discrete processes, typically in production/assembly lines, where the division into individual process states is evident. Statistical methods can be used to locate failures within the process and estimate the reliability of a system. The fact that DDDTs do not offer a detailed insight into the operation of an asset has more serious implications for processes with non-discrete success measures, such as CNC machining [20]. Although a purely data-driven approach may offer a satisfactory solution, a solely statistical correlation of process parameters may hinder the operator's understanding of asset operation and would require collecting a new dataset in the case of e.g., adding a new process parameter.

The literature review has shown that PBDTs offer considerable insight into the operation of an asset, and for the discussed cases, result in satisfactory DT performance. In fact, it seems natural to choose a comprehensive physical model to represent the behaviour of an asset as faithfully as a *twin* would. There also exist ways in which computational demand of such models can be decreased, such as by using reduced-order models [22], [27], to make them more suitable for near real-time deployment. The use of detailed physical simulations does not prohibit incorporation of advanced statistical methods either, which can be used, e.g., to aid the decision-making process.

However, creating a comprehensive physical model of an asset is not always possible. The referenced literature shows cases when a faithful theoretical model cannot be constructed due to unpredictable behaviour of the filtration process [24] or when creating such a model is not feasible due to strong nonlinearity and many degrees of freedom present in the system [25]. Therefore, PIDTs are seen as a solution which allows some physical dependences to be introduced into the model of an asset, while relying heavily on experimental data. Such solutions offer more control over the behaviour of the model due to the deterministic part present, at the same time taking advantage of statistical methods.

In the next section, the distinct advantages and disadvantages of these models will be explored in context of a specific application of a DT at an advanced research facility.

III. FastBlade Case Study

A. *FastBlade – Facility at the Forefront of Innovation*

FastBlade, pictured in Fig. 2, is the world's first regenerative fatigue test facility [28]. It is run by the University of Edinburgh and is located in Rosyth, Scotland. The site allows for testing mechanical performance of metal and composite structures of up to 14 metres, such as beams, columns and blades. The predominant purpose of the facility is to investigate the properties of tidal turbine blades by subjecting them to both static and cyclic loads. This way, fatigue applied to a specimen simulates the loads a blade would experience over its lifetime of subsea deployment at an accelerated pace. The tidal energy potential in the UK is estimated at 50 TWh per year, constituting about half of the potential European resource [29]. By validating the design and manufacturing processes at FastBlade, the tidal energy generation technology can be de-risked, thus helping exploit this great potential in the UK and worldwide.

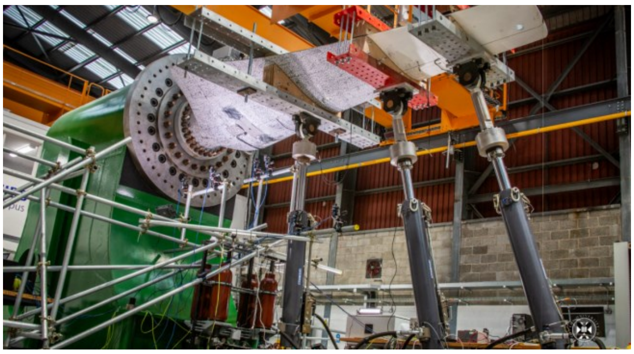

Fig. 2. The main experimental setup at FastBlade [30]. The specimen under test is mounted on the reaction frame and deflected by three actuators. Source: The University of Edinburgh.

FastBlade is equipped with a proprietary energy-recovery system which allows efficient energy transition between electric motors, pumps and the deforming specimen. Thanks to this, FastBlade can carry out tests at significantly reduced energy consumption, making the process much more economical. This state-of-the-art solution circumvents the challenge of the high natural frequency of short and stiff tidal blades, which prohibits the energy-efficient actuation of a specimen at its resonant frequency. A great number and variety of sensors can be found across the site, including accelerometers, strain gauges and thermocouples on the test specimen, as well as power transducers, rotational encoders, flowmeters and other sensors installed on the pumps, electric motors, inverters, actuators and oil pipes. The sensors are connected to synchronised data loggers, which store data for analysis. Such variety of data recorded provides not only a detailed insight into the behaviour of the specimen, but also the operation of the facility.

B. *Reasons for Digital Twin Development*

FastBlade is a pioneering site in the area of regenerative structural tests, enabling industry-academia collaboration on an integrated research and development process within the Open Engineering framework [31]. However, the success of the facility depends on how reliably its complex system can operate, delivering high-quality tests with minimum downtime. DTs, as introduced in the introduction to this paper, answer the need for a comprehensive monitoring tool which is capable of not only storing data and providing a real-time insight into its operation, but also indicating faults and suggesting system improvements. The most important reasons for investing in the development a system of DTs for FastBlade are:

- **Anomaly detection**: detecting anomalies in the operation of an asset protects the equipment and allows for safe operation of the system, reducing operational downtime by enhanced maintenance scheduling. It is also an enabling step in allowing unmanned operation of the facility.
- **Close monitoring of test parameters**: applying common standards to evaluate test performance, such as process capability index (PCI) [32], ensures that the test results are credible and is necessary for certification purposes.
- **Close monitoring of the proprietary energy recovery system**: monitoring the system is essential not only to validate its operational concept, but also quantify and minimise the energy consumption, which is the most significant running cost of a test.
- **Recording historic data**: a systemised way of storing historic data is crucial in assessing the degradation of the assets and developing further system improvements.

C. *Considerations for a Digital Twin at FastBlade*

Although there is an evident need for a DT to be developed for FastBlade, the unique characteristics of the facility pose a number of challenges. The most significant factors identified to affect the DT development strategy include the use of very specific, niche system components, such as the Digital Displacement® pumps [33], pictured in Fig. 3. The pumps do not only meet the operational pressure requirements to meet the target loads, but also allow for bi-directional fluid flow, which is indispensable for recovering energy in the system. However, the products are off-the-shelf and their physical models are not supplied. Moreover, complex asset operation and some of its internal design being protected by trade secrets, seem to be prohibitive obstacles to the development of faithful and reliable physics-based models. The problem scales up to the entirety of the system at FastBlade, as due to the one-off setup, there does not exist a comprehensive physical simulation which would just require calibration to turn it into a DT, but rather ground-up DT development. The efforts which have already been invested in building a physical model of the hydraulic system, including piping topology, pumps, valves and PID controllers, have turned out to be unreliable and computationally-slow, which would require further upgrades and running-time improvement. Considering the development of a DDDT for FastBlade, another challenge discovered is limited historic data available, since the site only became operational in 2022. While there exist numerous records of tests run at different load and frequency specifications, information about the degradation of some of the critical assets, such as electric motors, reaction frame or pumps, is scarce.

One of the greatest opportunities for creating a DT for FastBlade is the synchronised sensing infrastructure present, holistically monitoring most of the assets present on site. Moreover, the availability of significant computational resources at FastBlade, including GPUs with a total memory of 192 GB, opens the door to taking full advantage of recent developments in data-driven methods. The literature study shows that some of the aforementioned problems, such as the lack of a comprehensive physics-based models, can be mitigated with the use of ML or statistical modelling. The advent of efficient feature extraction algorithms, such as Short-time Fourier transform (STFT) [34], or wavelet scattering [35], and ANN architectures, such as autoencoders [36], also allows feasible and accurate detection of anomalies using recorded data alone [37], [38], [39], [40].

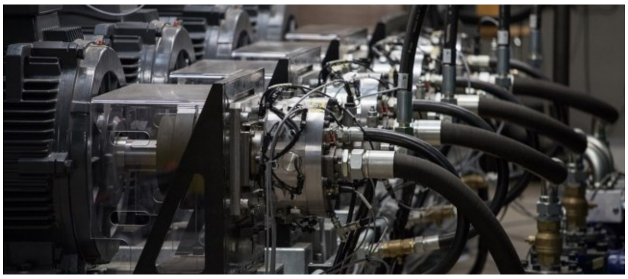

Fig. 3. Four Digital Displacement® pumps in the FastBlade's machinery room. Each pump is mechanically coupled to the shaft of an electric motor. Source: The University of Edinburgh.

D. *Choosing the Optimal Digital Twin Development Strategy for FastBlade*

The aim of this subsection is to consider how the characteristics of FastBlade impacts the choice of the most suitable DT development route considering both challenges and benefits of each of the strategies discussed in this work. Considering DDDTs, the question to answer is how useful the insight into the operation of the plant can be without the use of any physical information about the system. On the one hand, the static and fatigue tests run at FastBlade are not processes whose outcome can be easily classified as either *success* or *failure*, but they have a more descriptive character. However, there is certainly value in certain statistical methods, such as PCI, in assessing how consistent the minimum and maximum load exerted on a specimen is. Similarly, no physical models are required for detecting anomalies in e.g., vibrational traces recorded by accelerometers installed on the electric motors, which can highlight bearing issues, or analysing audio acquired on site for non-specific anomaly detection. Therefore, developing a DDDT for FastBlade would certainly provide much insight into its operation, showing potential for a real-time impact on the operation of the physical assets, e.g., when anomalous behaviour is discovered. On the other hand, considering the arguments drawn from the literature review, there would certainly be limitations to how in-depth the insight into operation of the facility would be. It would not be possible to pinpoint the cause of an anomaly once it has been found using audio data, it would be hard to find out how to improve a poor PCI score or predict the results of a test without having an extensive dataset collected for a given specimen.

Certainly, many such issues would be addressed by developing a PBDT. Having a faithful physical model of the setup at FastBlade, the specimen under test would potentially represent the only unknown in the system. Referring to one of established cantilever models, it would be possible to calibrate the simulation in a feasible number of cycles and try to predict the outcome of the test long before it is completed. This way, test parameters such as nominal motor speed or test frequency could be optimised to enhance the efficiency of the process. Similarly, updating the state of each asset in real time would enable precise localisation of identified faults. However, due to the limitations mentioned, it might not only be unfeasible to invest resources into the development of such a faithful model, but doing so might even be impossible considering our limited knowledge of some of the asset internals. Yet, in line with the literature review findings, it should be possible to fill in the gaps in our understanding of the system thanks to the richness of the data collected on site. Introducing a physical constraint to the models developed, even if based on simplified assumption (such as a spring-damper model for the specimen, or a simple rotating rigid body representation for the electric motors) should not only help the algorithms train faster, but should also provide additional credibility of the predicted results. Therefore, having identified and discussed the key routes for DT development with varying dependence on physics-based models, it is decided that developing a PIDT for FastBlade should constitute a practicable solution, at the same time providing desired operational benefits.

E. Dgital Twin Development and Integration at FastBlade

Based on the classification of DTs considered in this work, and the evaluation of the system found at FastBlade, the optimal DT development strategy has been chosen. The aim of this subsection is to outline the key considerations of creating a PIDT for FastBlade, highlighting how approaching individual problems will help combine the reliance on data and statistical models with the physical insight into the DT.

Primarily, the purely data-driven solutions may be applied in some of the anomaly detection and test quality assurance methods. Microphone data can be used for non-specific, audio-based anomaly detection requiring little historic data. By recording audio of the normal system operation, it is then possible to incorporate a signal processing pipeline to extract relevant features and train an ANN to classify abnormal behaviour [41]. Such a tool may be used as the first line of prevention against running the system with a fault.

Moreover, no physical insight is needed for real-time evaluation of the test quality. The aforementioned techniques, such as PCI, are widely used in industry and provide a statistical measure of how close the distribution of recorded values is to the expected outcomes. They can be applied to a variety of system parameters in FastBlade, including the maximum and minimum load recorded cycle-to-cycle in a fatigue test, where pre-defined values must be matched to meet certification requirements. Purely data-driven approaches can also contribute to the process of developing the DT by optimising sensor locations across the system. As an example, high-resolution strain maps of the specimen under test, obtained through Digital Image Correlation (DIC), can be used to optimise the location of point strain gauges across the sample [42]. This way, the redundancy of information collected is reduced and the measurements from sparsely located sensors can be used to reconstruct high-resolution strain data.

As discussed in the previous sections, data alone will be used to model the performance of the pumps in the system. Due to the complexity of their operation, metamodels can be incorporated to corelate recorded operational parameters, such as rotational speed, temperature or valve control signals, to the output pressure. Similarly, it should be possible to track the degradation of the reaction frame, which is one of the most expensive utilities in the system, without running computationally-expensive FEA models, as long as the change in the surface strain relative to the system loads is monitored.

The assets whose simplified physical models can provide useful insight and improve the quality of ML models are the electric motors and the oil delivery system. The motors can be modelled using rotating rigid body theory, calibrated by verifying their theoretical moment of inertia. The existing physical model of the oil system, containing detailed information on hose lengths, bend radii and valves used, can be simplified with losses averaged over larger sections. Lastly, the specimen under test, whose dynamics is not verified until the test commences, can be modelled using a spring-damper model. The test sample stores energy as it is deflected (like a spring) and both actuation and relaxation have a damping factor, which should be correlated with the actuation velocity. This way, a simplified specimen model could be calibrated when the first cycles are run, and used to predict the test outcomes.

IV. Discussion and Conclusions

In this paper we address the question of the need for physical models in developing digital twins, which has not been considered explicitly in literature. We chose to assign cases found in literature to one of the three DT categories, namely: purely data-driven (DDDTs); relying on comprehensive physical simulations (PBDTs); or using limited physical knowledge of the system (PIDTs). The framework for deciding on the most optimal DT development strategy resulting from the insights gained in this work is presented using a decision tree in Fig. 4.

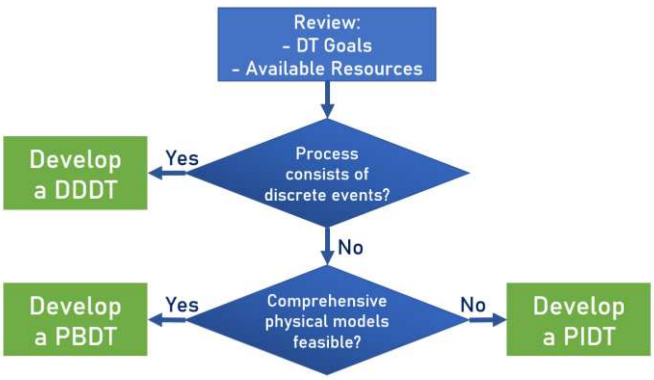

Fig. 4. Decision tree intended to help optimise the DT development route. Source: authors.

The requirement for DDDTs is usually an extensive dataset. They are suitable for discrete processes, and while precious feedback can be provided for other cases, there will almost certainly be limitations to the insight provided into the operation of an asset. PBDTs generally offer the most detailed description of the asset operation, but they are associated with high upfront cost of developing a comprehensive physical model. Although they were historically expensive to run, there exist ways to improve their feasibility. They can also be integrated with statistical solutions, e.g., in the decision-making process. PIDTs can be utilised in cases when it is either impossible or impractical to create a high-fidelity physical model. Relying to a large extent on historical data and statistical methods, they incorporate certain physical information about the system which can mitigate some of the downsides of not having a complete physical simulation. The in-depth analysis of FastBlade's system needs and the available infrastructure shows that developing PIDTs can maximise the advantage gained, not restricting the parallel incorporation of purely data-driven solutions.

Looking critically at the process of optimising the DT development route for FastBlade presented in the paper, it could be concluded that the chosen solution, namely creating a PIDT, constitutes some form of middle ground between DDDTs and PBDTs which leverages the benefits of the both worlds. Seen in such light, opting for the *hybrid* solution would not require such an in-depth analysis. However, it has to be noted that although a PIDT can be seen as a compromise in some ways, there exist systems where DDDTs or PBDTs would be the preferred DT solution. DDDTs would typically offer faster development and better system transferability, while the presence of physical models does not limit the extent to which data-driven solutions are incorporated in PBDTs. Therefore, the multi-level analysis of system requirements and limitations is recommended each time a new DT development strategy is considered.

Acknowledgments

M. J. M. thanks Babcock International and The Data Lab for funding his PhD project.

S. L. D. thanks the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 801215 and the University of Edinburgh Data-Driven Innovation programme, part of the Edinburgh and South East Scotland City Region Deal.

References

[1] G. Pronost, F. Mayer, B. Marche, M. Camargo, and L. Dupont, "Towards a Framework for the Classification of Digital Twins and their Applications," in *2021 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, IEEE, Jun. 2021, pp. 1–7. doi: 10.1109/ICE/ITMC52061.2021.9570114.

[2] M. Singh, E. Fuenmayor, E. Hinchy, Y. Qiao, N. Murray, and D. Devine, "Digital Twin: Origin to Future," *Applied System Innovation*, vol. 4, no. 2, p. 36, May 2021, doi: 10.3390/asi4020036.

[3] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital Twin in Industry: State-of-the-Art," *IEEE Trans Industr Inform*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019, doi: 10.1109/TII.2018.2873186.

[4] K.-J. Wang, Y.-H. Lee, and S. Angelica, "Digital twin design for real-time monitoring – a case study of die cutting machine," *Int J Prod Res*, vol. 59, no. 21, pp. 6471–6485, Nov. 2021, doi: 10.1080/00207543.2020.1817999.

[5] T. H.-J. Uhlemann, C. Schock, C. Lehmann, S. Freiberger, and R. Steinhilper, "The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems," *Procedia Manuf*, vol. 9, pp. 113–120, 2017, doi: 10.1016/j.promfg.2017.04.043.

[6] R. Söderberg, K. Wärmefjord, J. S. Carlson, and L. Lindkvist, "Toward a Digital Twin for real-time geometry assurance in individualized production," *CIRP Annals*, vol. 66, no. 1, pp. 137–140, 2017, doi: 10.1016/j.cirp.2017.04.038.

[7] R. Rituraj and R. Scheidl, "Towards digital twin development of counterbalance valves: Modelling and experimental investigation," *Mech Syst Signal Process*, vol. 188, p. 110049, Apr. 2023, doi: 10.1016/j.ymssp.2022.110049.

[8] I. Anda, R. Mishra, S. Iram, A. M. Aliyu, F. F. Jackson, and L. Fleming, "A Machine Learning Model for the Development of a Digital Twin for a Control Valve for Oil and Gas Pipelines," in *2022 International Conference on Maintenance and Intelligent Asset Management (ICMIAM)*, IEEE, Dec. 2022, pp. 1–6. doi: 10.1109/ICMIAM56779.2022.10146806.

[9] J.-H. Lee, Y.-S. Nam, Y. Kim, Y. Liu, J. Lee, and H. Yang, "Real-time digital twin for ship operation in waves," *Ocean Engineering*, vol. 266, p. 112867, Dec. 2022, doi: 10.1016/j.oceaneng.2022.112867.

[10] J. Michael, J. Pfeiffer, B. Rumpe, and A. Wortmann, "Integration challenges for digital twin systems-of-systems," in *Proceedings of the 10th IEEE/ACM International Workshop on Software Engineering for Systems-of-Systems and Software Ecosystems*, New York, NY, USA: ACM, May 2022, pp. 9–12. doi: 10.1145/3528229.3529384.

[11] L. Guzina, E. Ferko, and A. Bucaioni, "Investigating Digital Twin: A Systematic Mapping Study," 2022. doi: 10.3233/ATDE220164.

[12] M. Javaid, A. Haleem, and R. Suman, "Digital Twin applications toward Industry 4.0: A Review," *Cognitive Robotics*, vol. 3, pp. 71–92, 2023, doi: 10.1016/j.cogr.2023.04.003.

[13] C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, "Digital twin paradigm: A systematic literature review," *Comput Ind*, vol. 130, p. 103469, Sep. 2021, doi: 10.1016/j.compind.2021.103469.

[14] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the Digital Twin: A systematic literature review," *CIRP J Manuf Sci Technol*, vol. 29, pp. 36–52, May 2020, doi: 10.1016/j.cirpj.2020.02.002.

[15] X. Liu *et al.*, "A systematic review of digital twin about physical entities, virtual models, twin data, and applications," *Advanced Engineering Informatics*, vol. 55, p. 101876, Jan. 2023, doi: 10.1016/j.aei.2023.101876.

[16] J.-F. Yao, Y. Yang, X.-C. Wang, and X.-P. Zhang, "Systematic review of digital twin technology and applications," *Vis Comput Ind Biomed Art*, vol. 6, no. 1, p. 10, May 2023, doi: 10.1186/s42492-023-00137-4.

[17] S. Erikstad, "Merging Physics, Big Data Analytics and Simulation for the Next-Generation Digital Twins," Mar. 2017. Accessed: Mar. 06, 2024. [Online]. Available: https://www.researchgate.net/publication/320196420_Merging_Physics_Big_Data_Analytics_and_Simulation_for_the_Next-Generation_Digital_Twins?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoicHJvZmlsZSIsInByZXZpb3VzUGFnZSI6InB1YmxpY2F0aW9uIn19#full-text

[18] M. Resman, J. Protner, M. Simic, and N. Herakovic, "A Five-Step Approach to Planning Data-Driven Digital Twins for Discrete Manufacturing Systems," *Applied Sciences*, vol. 11, no. 8, p. 3639, Apr. 2021, doi: 10.3390/app11083639.

[19] J. Friederich, D. P. Francis, S. Lazarova-Molnar, and N. Mohamed, "A framework for data-driven digital twins of smart manufacturing systems," *Comput Ind*, vol. 136, p. 103586, Apr. 2022, doi: 10.1016/j.compind.2021.103586.

[20] V. S. Vishnu, K. G. Varghese, and B. Gurumoorthy, "A data-driven digital twin framework for key performance indicators in CNC machining processes," *Int J Comput Integr Manuf*, vol. 36, no. 12, pp. 1823–1841, Dec. 2023, doi: 10.1080/0951192X.2023.2177741.

[21] M. G. Kapteyn, D. J. Knezevic, and K. Willcox, "Toward predictive digital twins via component-based reduced-order models and interpretable machine learning," in *AIAA Scitech 2020 Forum*, Reston, Virginia: American Institute of Aeronautics and Astronautics, Jan. 2020. doi: 10.2514/6.2020-0418.

[22] M. G. Kapteyn, D. J. Knezevic, D. B. P. Huynh, M. Tran, and K. E. Willcox, "Data-driven physics-based digital twins via a library

of component-based reduced-order models," *Int J Numer Methods Eng*, vol. 123, no. 13, pp. 2986–3003, Jul. 2022, doi: 10.1002/nme.6423.

[23] M. G. Kapteyn, J. V. R. Pretorius, and K. E. Willcox, "A probabilistic graphical model foundation for enabling predictive digital twins at scale," *Nat Comput Sci*, vol. 1, no. 5, pp. 337–347, May 2021, doi: 10.1038/s43588-021-00069-0.

[24] J. K. Møller, G. Goranović, P. Brath, and H. Madsen, "A data-driven digital twin for water ultrafiltration," *Communications Engineering*, vol. 1, no. 1, p. 23, Sep. 2022, doi: 10.1038/s44172-022-00023-6.

[25] J. Zhang and X. Zhao, "Digital twin of wind farms via physics-informed deep learning," *Energy Convers Manag*, vol. 293, p. 117507, Oct. 2023, doi: 10.1016/j.enconman.2023.117507.

[26] Y. A. Yucesan and F. A. C. Viana, "Physics-informed digital twin for wind turbine main bearing fatigue: Quantifying uncertainty in grease degradation," *Appl Soft Comput*, vol. 149, p. 110921, Dec. 2023, doi: 10.1016/j.asoc.2023.110921.

[27] C. Huang, K. Duraisamy, and C. Merkle, "Component-Based Reduced Order Modeling of Large-Scale Complex Systems," *Front Phys*, vol. 10, Aug. 2022, doi: 10.3389/fphy.2022.900064.

[28] S. Lopez Dubon, F. Cuthill, C. Vogel, C. Ó Brádaigh, and E. D. McCarthy, "A full-scale composite tidal blade fatigue test using single and multiple actuators," *Compos Part A Appl Sci Manuf*, vol. 181, p. 108140, Jun. 2024, doi: 10.1016/j.compositesa.2024.108140.

[29] N. Yates, I. Walkington, R. Burrows, and J. Wolf, "Appraising the extractable tidal energy resource of the UK's western coastal waters," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1985, p. 20120181, Feb. 2013, doi: 10.1098/rsta.2012.0181.

[30] The University of Edinburgh, "FastBlade, Services Offered." Accessed: Mar. 13, 2024. [Online]. Available: https://www.fastblade.eng.ed.ac.uk/services-offered

[31] M. Vidmar, "Aligning Open Engineering Practices," in *2022 IEEE 28th International Conference on Engineering, Technology and Innovation (ICE/ITMC) & 31st International Association For Management of Technology (IAMOT) Joint Conference*, IEEE, Jun. 2022, pp. 1–9. doi: 10.1109/ICE/ITMC-IAMOT55089.2022.10033245.

[32] C.-W. Wu, W. L. Pearn, and S. Kotz, "An overview of theory and practice on process capability indices for quality assurance," *Int J Prod Econ*, vol. 117, no. 2, pp. 338–359, Feb. 2009, doi: 10.1016/j.ijpe.2008.11.008.

[33] Danfoss, "Digital Displacement® pumps." Accessed: Mar. 15, 2024. [Online]. Available: https://www.danfoss.com/en/products/dps/pumps/digital-displacement-pumps/digital-displacement-single-and-multiple-outlet-pumps/#tab-news

[34] N. Kehtarnavaz, "CHAPTER 7 - Frequency Domain Processing," in *Digital Signal Processing System Design (Second Edition)*, Second Edition., N. Kehtarnavaz, Ed., Burlington: Academic Press, 2008, pp. 175–196. doi: https://doi.org/10.1016/B978-0-12-374490-6.00007-6.

[35] J. Bruna and S. Mallat, "Invariant Scattering Convolution Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013, doi: 10.1109/TPAMI.2012.230.

[36] W. H. Lopez Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, "Chapter 11 - Autoencoders," in *Machine Learning*, A. Mechelli and S. Vieira, Eds., Academic Press, 2020, pp. 193–208. doi: https://doi.org/10.1016/B978-0-12-815739-8.00011-0.

[37] S. Givnan, C. Chalmers, P. Fergus, S. Ortega-Martorell, and T. Whalley, "Anomaly Detection Using Autoencoder Reconstruction upon Industrial Motors," *Sensors*, vol. 22, no. 9, p. 3166, Apr. 2022, doi: 10.3390/s22093166.

[38] R. N. Toma *et al.*, "Classification Framework of the Bearing Faults of an Induction Motor Using Wavelet Scattering Transform-Based Features," *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228958.

[39] Z. Liu, G. Yao, Q. Zhang, J. Zhang, and X. Zeng, "Wavelet Scattering Transform for ECG Beat Classification," *Comput Math Methods Med*, vol. 2020, 2020, doi: 10.1155/2020/3215681.

[40] M. Z. Ahmad, A. M. Kamboh, S. Saleem, and A. A. Khan, "Mallat's Scattering Transform Based Anomaly Sensing for Detection of Seizures in Scalp EEG," *IEEE Access*, vol. 5, pp. 16919–16929, Aug. 2017, doi: 10.1109/ACCESS.2017.2736014.

[41] M. J. Munko, M. A. Valdivia Camacho, F. Cuthill, C. M. Ó Brádaigh, and S. Lopez Dubon, "Using audio-data for anomaly detection in the fatigue test of a composite tidal turbine blade," *Proceedings of the International Conference on Condition Monitoring and Asset Management*, vol. 2023, no. 1, pp. 1–2, Jan. 2023, doi: 10.1784/cm2023.2e8.

[42] J. McLoughlin, M. J. Munko, M. A. Valdivia Camacho, F. Cuthill, and S. L. Dubon, "Use of digital image correlation and machine learning for the optimal strain placement in a full-scale composite tidal turbine blade," *Proceedings of the International Conference on Condition Monitoring and Asset Management*, vol. 2023, no. 1, pp. 1–12, Jan. 2023, doi: 10.1784/cm2023.2d4.