The digital twin in Industry 4.0: a wide-angle perspective

Ron S. Kenett[1] and Jacob Bortman[2]*

[1]KPA, Raanana and Samuel Neaman Institute, Technion, Israel
[2]Department of Mechanical Engineering, Ben Gurion University of the Negev, Israel
*corresponding author: ron@kpa-group.com

Abstract

The move towards advanced manufacturing and industry 4.0 is fed by increased demand for speeding up innovation, increasing flexibility, improving maintenance, becoming more customized while saving on total cost of operations. The move is accompanied by increased dependence on virtual product and process development, including data driven process and product knowledge. Other characteristics affecting modern design, manufacturing as well as maintenance, include big data intelligence in product, process and maintenance, globally competitive operations, added manufacturing, flexible manufacturing and robotics, sensor technology, smart value/supply chains and industrial information technology backbones.

This paper is about surrogate models, also called *digital twins*, that are providing an important complementary capacity to physical assets. Digital twins capture past, present and predicted behavior of physical assets. Digital twin models are updated periodically to represent the current state of physical assets. The type of curated information on the state of physical asset's history depends on how digital twin are used. For example, if a digital twin is used for fault classification, the history captured can be operational data from specific equipment in healthy and faulty state.

We provide here a review of digital twins, with an emphasis on Industry 4.0 applications. The paper maps the background, the current state of the art and the future directions of digital twins in a wide range of process engineering, product design and analytic disciplines.

Keywords: Digital Twins, Simulations, Model Calibration, Condition Based Maintenance, Process Control, Information Quality, Industry 4.0.

1. Introduction

The term "digital twin" has been defined in different forms including as a high-fidelity simulation, a virtual organization, a virtual reality representation and an emulation facility. Its uses are focused, among other applications in optimization, monitoring, diagnostic, prognostic and prescriptive capabilities (Kenett et al, 2016). It originated with the concept of a digital factory (Jain and Shao 2014). The definition we use here is that a digital twin is the digital representation of a physical asset or system across its life-cycle using operational real-time data and other sources, adopted to drive business outcomes.

The digital twin concept has been implemented by leading manufacturing companies. Ford Motor Company enhanced assembly line performance by evaluating and optimizing the designs using digital twins (IMT 2013). Volvo Group Global (2017) showed how to validate changes using a digital twin. General Electric developed digital twins of aircraft engines[ab]. Major commercial software vendors support development of virtual factories via integrated

[a] https://www.youtube.com/watch?v=gE0Z-lQdMtI
[b] https://www.ge.com/research/offering/digital-twin-creation

solutions for product, process and system design, simulation, and visualization (Tolio et al. 2013). A standardization of process control technologies is provided by ANSI (2010). On the other hand, Jain et al. (2001) attempted to implement a multi-resolution digital twin but found it highly challenging due technology limitations and information availability.

Virtual data management, automatic model generation, static and dynamic simulation, and integration and communication are paramount to realizing a digital twin (Choi et al 2014). However, most software tools are, in general, not supplied with these capabilities making it a challenge to develop a digital twin. There are efforts addressing different aspects of the challenge. To enhance conventional simulations for a digital twin, Bal and Hashemipour (2009) use Product-Resource-Order-Staff Architecture for modelling controls while the Quest simulation tool models the physical elements. To integrate models and enhance communication, Hints et al. (2011) developed a software tool named Design Synthesis Module. Debevec et al (2014) use a simulative model to test and improve the schedules before implementation in the real factories of small and medium size enterprises. For production planning, Terkaj at all (2015) present an ontology for a virtual factory or digital twin to aid planning decisions.

The recent concept of 'Industry 4.0', or the fourth industrial revolution, includes Cyber-Physical Systems (CPS) as a key component. The function of CPS has been identified as monitoring physical processes and creating a virtual copy of the physical world to support decentralized decision-making (Mario et al 2015). The 'virtual copy' and 'virtual plant models' discussed in that paper is in the context of Industry 4.0 and matches the concept of twining discussed in this paper.

In Industry 4.0 applications, one sees a growing role of twinning a physical plant with simulation-based surrogates. A digital twin is a simulation model which replicates a physical object or process. By means of sensors, real-time data about physical items are collected and used to duplicate the physical state of the item and the impact of ongoing changes (Kenett et al, 2020). Digital twins include five main components: physical part, virtual part, connection, data, and service. The virtual and physical parts exchange information collected through the connection part. The interaction between the human and the digital twin is provided by the service part. Digital twins are traditionally used to improve the performance of engineering devices, like wind turbines or jet engines. In this context, they also serve to model systems of devices, to collect and analyze information about processes and people, and to help solve complex problems. Such digital twins provide powerful planning and troubleshooting capabilities and statistical methods play a significant role in both the design and analysis of simulations and computer experiments on digital twin platforms.

Simulation models of systems and processes are increasingly used in order to shorten time to market, reduce design, operations and maintenance costs while improving quality. As an example, consider the PENSIM simulation software modeling penicillin production in a fed-batch fermentor[c]. The model includes variables such as pH, temperature, aeration rate, agitation power, and feed flow rate of the substrate. It is used in monitoring and process trouble shooting activities. Such simulators are used, for example, in fault diagnosis of semiconductors, biotechnological and chemical production processes[1]. They are also used for research and educational purposes (Reis and Kenett, 2017).

---

[c] https://nl.mathworks.com/matlabcentral/fileexchange/49041-industrial-scale-penicillin-simulationv2

A digital twin is a power full infrastructure for the condition-based maintenance (CBM) approach which is based on the concept that maintenance operations should be done only when necessary (Gruber et al, 2020). The purpose of CBM is to prevent a reduction in the effectiveness of a system which can evolve to a total failure of the system. It is aimed to reduce the maintenance costs by enabling planning of maintenance operations in advance and replacement of the damaged component. For effective CBM, prediction of the remaining useful life (RUL) of each component is required. Evaluation of the components' RUL requires not only diagnosis of the fault existence, but also the estimation of the fault location and severity.

Simulating the behaviour of a system, with computer software, is a cost-efficient approach to knowledge building. Computer experiments differ from physical experiments since the output of a computer run is deterministic, unaffected by measurement error and experimental noise. On the other hand, like with physical experiments, models and knowledge derived from computer experiments require validation.

In this paper we provide a wide-angle perspective of digital twins, simulations and computer experiments, highlighting areas of interest for practitioners, researchers and educators in the context of Industry 4.0. we cover several complementary areas: i) integration of physical systems and computer simulations, ii) a case study of a dynamic model for gear fault diagnosis used to develop diagnostic algorithms, iii) examples of digital twins, iv) calibration of digital twins and v) the role of digital twins in training and education. A final section, with discussion and extensive references, concludes the paper.

Digital twins complement or substitute physical experiments with software-based simulation experiments, Santner et al (2003), Kenett and Zacks (2014). Digital twins permit building knowledge of a physical system and supporting decision-making in the design and monitoring of such systems, via simulations. A physical experiment consists of executing real life experiments where modifications to input variables are taking place, see Figure 1. Similarly, a digital twin can be used to run experiments by a number of runs of a simulation code where factors are parametrized as a subset of the code's inputs.

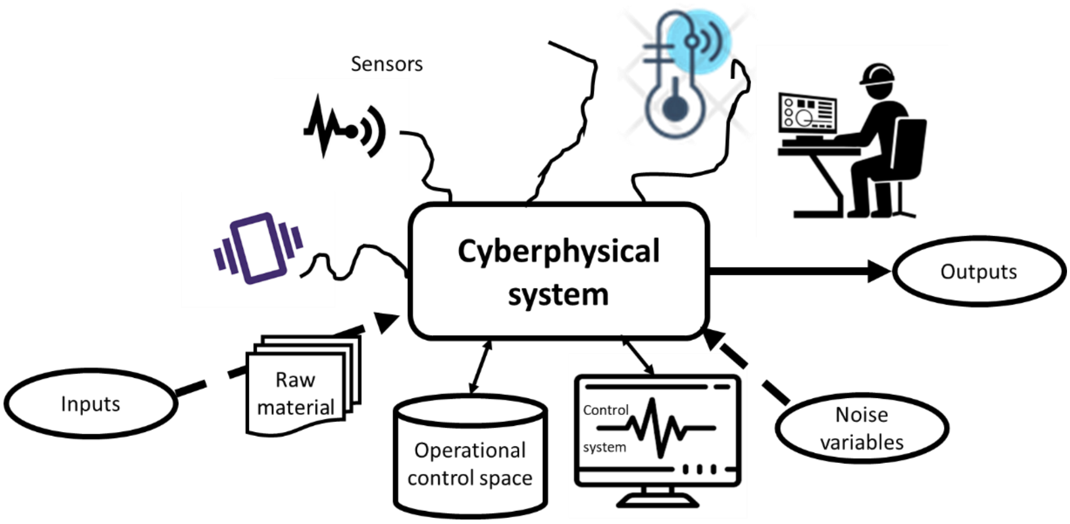

Figure 1 – Representation of the behavior of a cyberphysical system in terms of inputs (controlled variables and noise variables) and outputs (or responses).

Some examples of limitations in physical experiments that are overcome in digital twins are:

- *Inability to measure the response*. For instance, time evolution of internal stress during a material test or nanoscale phenomena.
- *Topological complexity*. For example, microsystems in semiconductor industry with thousands of interacting components.
- *Non-reproducibility of experiments*. For example, in customer interactions and supply chain systems.

Typically, the complexity of physical systems does not lead to a simple analytic formula that describes in detail the phenomena under study, and a simulation is the only available tool for experimentation. In order to overcome difficulties in carrying out physical experiments, one can resort to a digital twin that provides the option of extensive experimentation and to collecting information that would otherwise not be possible to collect. Complex simulation models are, for practical purposes, black boxes that work on a what-if basis. The proactive experimentation is a way to extract usable knowledge by modelling digital twin outputs.

A standard goal of a digital twin is optimisation of the system performance. In this case, computer experiments are used to create a fast-to-compute "copy" of the simulator used for the optimisation, see Figure 2. Specifically, given a simulator f, digital twin methods search for an approximate model g which mimics the behavior of the simulator but runs hundreds or thousands of times faster.

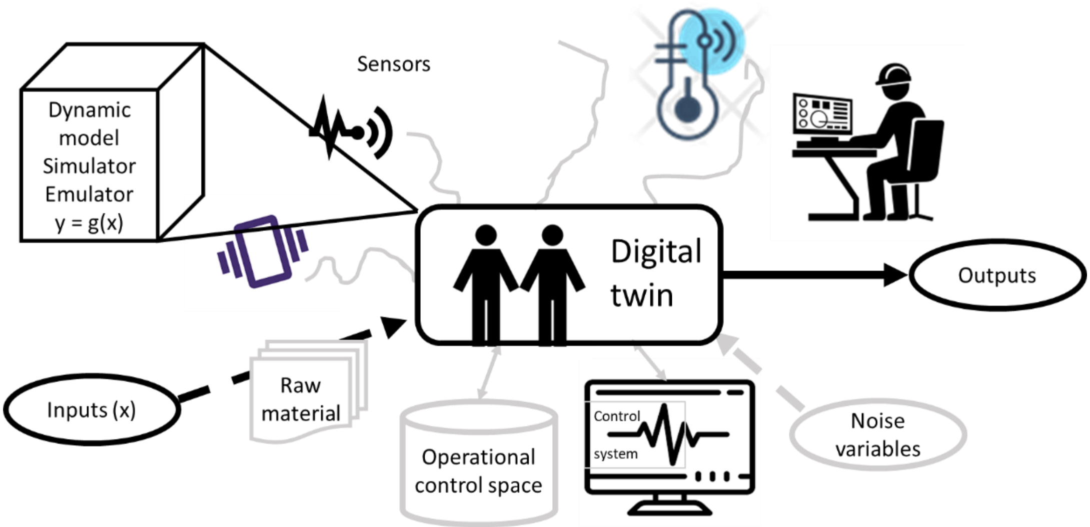

Figure 2 – Representation of the behavior of a digital twin and the use of emulators. The emulator $y = g(\boldsymbol{x})$ approximates the model $y = f(\boldsymbol{x})$ using smooth functional relationships in order to best represent the model and be computationally fast.

Digital twin simulation experiments can be combined with physical experiments. Combining digital twin simulations with physical experimentation validates both models and the digital twin code. It allows experts to create and calibrate models used in optimization applications (Romano et al, 2004). In the next sections we provide a case study of gear fault diagnosis

3. A dynamic model for gear fault diagnosis

This section is a case study of a subsystem which is prevalent in manufacturing systems and systems in general. a rotating machine with a gear box. It is focused on monitoring, diagnostic and prognostic capabilities typical to have in digital twins.

3.1 Digital twin of a rotating machine

Common health monitoring methods for rotating machines are based on signal monitoring, usually acceleration signals and recently strain sensing (Alian et al (2019). These methods are designed to alert users on damage in the monitored component. However, the dynamic behavior of the component, with and without a fault, needs to be studied to certify the diagnostic capabilities. A relation between the dynamic behavior and its expression in the vibration signal should be created for determination of the defect severity. Using a combination of models and simulations contributes to the understanding of how the system functions when there is a fault and, in addition, to explain why and how the fault is expressed in the vibration signature. These models are used during the life cycle of a machine, as part of the machine's digital twin, to diagnose the machine condition, to optimize the operation such that its key performance goals are achieved

The methodology presented here is based on understanding the physics of the machine components and their expression in the vibration signal in the presence of a fault. Each rotating machine component is analyzed separately. It is assumed that isolation of the signal excited by each component is part of the preliminary signal processing task.

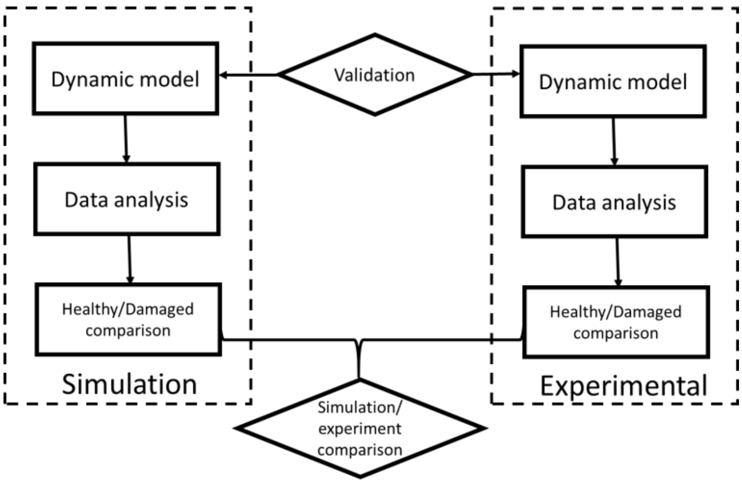

Figure 3: The rotating machine model development and validation approach (Madar et al, 2019).

The methodology is based on several stages as described schematically in Figure 3. First, a general dynamic model of the system under investigation is built. The dynamic equations are solved using numerical tools which describe the behavior of the component in healthy conditions as well as with faults. Each model undergoes validation and verification processes [(Madar et all (2019)) before the results are interpreted. The simulated signals from the models establish the physical understanding of the process occurring in different working modes. Additionally, prominent events in the simulated signal are investigated. Various experiments are performed, and the measured signals are analyzed. The insights from the simulation allow a better interpretation of the experimental signals and development of diagnostic and prognostic algorithms, especially, the condition of the system using Condition Indicators (CIs). Similarly, the experimental signals are used to validate the model and to learn about the robustness of the model limitations (Madar et all (2019), Kogan et all (2018) and Gazizulin et all (2019)).

In the current example, a nonlinear dynamic model was developed in order to describe the vibration response of a gear system in the presence of faults and surface roughness (Dadon et all (2018). The gear model considers the two gears, as well as their supporting system. The reason for modeling all the components is that the phenomena which occurs in the different system components affects the gear dynamics, and, as a result, also the vibration signal. In addition to the gears, the model includes simulations of bearings, shafts, a motor, and a brake. The components in the modeled system were assumed to be rigid bodies. The two shafts were assumed to be of finite torsional stiffness and infinite bending stiffness. The interaction between teeth meshing was modeled as a network of linear springs, which simulate the contact across the tooth face. The stiffness of each spring is variable and depends on the angular position of the gears. Other interactions between the rigid bodies were modeled as linear springs with constant stiffness. In the dynamic model, the clearance between meshing teeth is not considered. This is acceptable due to the distance to the system resonances.

The model predicts the gear system vibrations signature of healthy and damaged gears. The mesh properties, such as those related to gear faults and surface roughness, are considered in the gear mesh model. The surface roughness is given as a displacement-excitation function along the pressure line.

3.2 The tooth surface

The tooth surface quality of the gear has an important effect on the ability to detect faults via the vibration signal. For the case of a local fault, the expression of the fault in the measured vibration signal differs from its expression in the simulated signal. Modeling the surface roughness by adding profile errors to the tooth face contributed to the understanding of the difference between the measured and simulated signal (Figure 4). The effect of the gear surface quality on the generated signal with and without the fault sheds light on the possibility to detect a local fault. In the measured signal, the impact generated by the local fault is masked by the gear profile errors and, therefore, is undetectable. However, when a high-quality gear was tested experimentally, the local faults were clearly visible in the vibration signal. In the simulations of the high-quality gears, the fault was obviously observable starting from the first fault severity, where for the low-quality gears the local fault was not detected. This insight was verified experimentally on gears of different quality.

For a specific precision grade, the manufacture produces the surface with predefined tolerances, the gearwheel surface roughness varies in the tolerances, and gears from the same production line can still appear differently in the vibration signal. The model allows the running of various simulations of the gear with random profile errors in the tolerances defined by the manufacturer. This capability generates scatter in the results without changing the system or the operating conditions. Scattering of the results makes the model more realistic, contributing to testing different condition indicators on the scattered results that characterize the fault. Additionally, the efficiency of signal processing methods which better emphasize the local faults can be tested. The main insight is that the tooth surface quality has a profound effect on the vibration signal and, consequently, on the capability to detect faults.

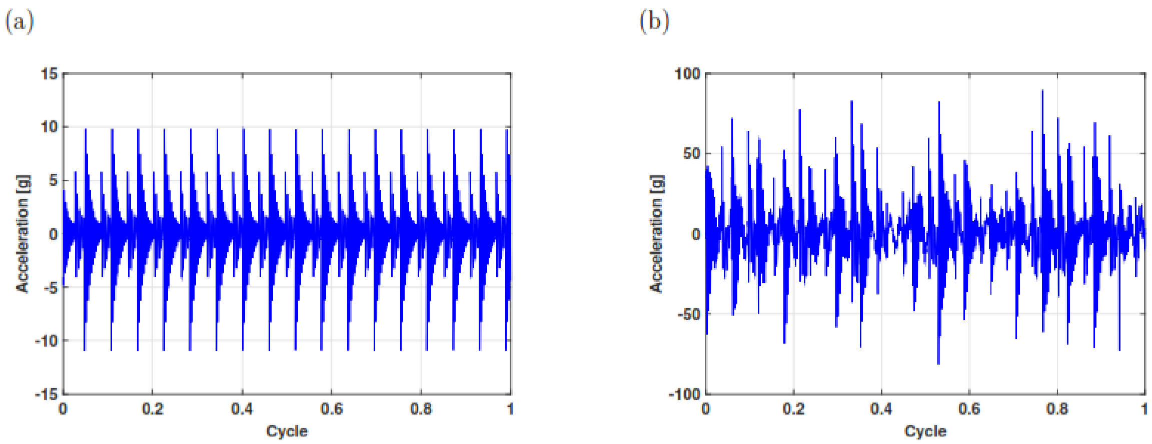

Figure 4: Simulated vibration cycle of the driving wheel. (a) Without profile errors; (b) With profile errors.

The model allows to check the possibility of identifying faults in the gears. The common local faults that we considered are, spall-like faults and chipped, broken, and missing teeth. Periodic Hertzian contact stresses on a tooth's surface produce detachment of small chunks of material. A broken tooth is caused by excessive impact load or unstable load which causes a part of the tooth edge to separate from it. Cases of broken teeth can be sectioned into three different grades of severity. The three states influence the involute profile differently. The first is a chipped tooth – non-uniform removal of material from the tooth edge. During meshing of the defected tooth, the involute profile is not altered throughout the tooth. The second state is a broken tooth, i.e. the tooth height is lowered, influencing the involute profile. The last state is a missing tooth, when a substantial amount of material is removed from the tooth. In that stage, the height of the remaining tooth does not matter. Each of these cases was tested and simulated. The model was validated by comparing the experiment results to the simulations. A database of simulations for gears with different local faults was built. For each local fault, several variables were examined, testing different severity levels, under different working conditions, with different surface qualities. The database serves to study the diagnostic capabilities systematically.

Currently, there are a variety of CIs to diagnose gear faults. The capability of a CI to detect a fault does not necessarily indicate that this CI will work on the same fault when the system operating condition is changed. Moreover, the same CI may be appropriate for different type of faults in different conditions. Hence, it is necessary to understand the limits of each CI, i.e. the type and the severity of faults that the CI can reflect. Extensive experiments (including different rotating speeds, loads, surface qualities, sizes of faults, and repetitions) are required to set boundaries for each CI. The usage of a validated model to simulate different cases and to test the CI on the simulations provides a more effective solution. Using simulations from a validated model, it is possible to find the characteristics of each local fault in the vibration signal. Examining each of the CIs on the simulation results indicates on the best CI or combination of CIs for each local fault. Then the results can be generalized and employed in a twin model to diagnose the health of gears during the life of the machine.

Many insights about the ability to detect different faults have emerged from this study. As expected, the best results of the CIs were in the difference signal due to removal of the signals irrelevant to the faults. A gear with a spall-like fault which does not go throughout the whole tooth is very challenging to detect because its effect is a small change in the frequency response for a very short period of time during the mesh of the damaged tooth. However, when the spall-like fault is all along the tooth, its presence is visible in the synchronous average signal because of the contact loss in the involute profile of the gear i.e. the two meshing teeth are in contact, but the location has changed. For this case, the fault can be detected using the classical CIs, but the severity of the defect cannot be determined.

3.3 Installation errors

Modeling local faults in gear transmission and studying their effect on the vibration signature of a gear transmission contributed to the development of a new algorithm for separation between local and distributed faults. Traces of a local fault may be obscured by distributed transmission defects, such as eccentricity or misalignment, whereas these phenomena may appear in parallel with the local fault.

Using the dynamic model, we simulated the distributed and local faults. Simulations of a gear transmission with different eccentricity levels have been conducted (Koen et al (2016)). Here, we examined three cases of simulated signals to understand the influence of the eccentricity on the vibration signal. The local and distributed fault combination in the vibration signal caused concealment of the local fault. Insight from the investigation of the simulated signals enabled us to develop and test a method to reduce the distributed fault effect masking the traces of the local fault. A new algorithm, which calculates a new adaptive difference signal, was demonstrated to emphasize the local tooth fault in the presence of a distributed fault that masked the effect of the local fault Koen et al (2016)).

In this section, we test the ability to detect different faults based on vibration signals using insights from the gear dynamic model. The main insight is that the tooth surface quality has a significant effect on the ability to detect faults via the vibration signal. By considering the tooth surface quality in the model, a random factor is added to the simulations which causes a scattering in the results. This addition contributes toward understanding the limits and reliability of fault detection. Detecting a partial spall or a chipped tooth is difficult, even if the local fault is the only defect in the system because the effect of those faults on the vibration signal is a small change of frequency for a short period of time. In general, if there is a continuous contact along the involute profile, i.e. the involute profile is not altered, the vibration signal does not change significantly compared to the healthy case. However, when the involute profile is changed as in the cases of spall throughout or a broken tooth, the displacement of the gear will be expressed in the vibration signal, enabling fault detection.

A database of signals from simulations obtained from the experimentally validated model was established. The database was used to test classic and new CIs on different types of faults in order to define their ability to detect different faults. Using the dynamic model, it is possible to explain why certain indicators are ineffective and some are more relevant for each case. In ongoing research, the database from the simulations is being used for sensitivity tests by studying the effects of different parameters on the vibrations.

This model and the data base can be used as infrastructure for constructing a gear twin model. This is an example how a validated model can be used to construct an individual realistic model twin for a complicated mechanical component that might deteriorate or be damaged.

This model may be very useful when moving to Condition Based Maintenance (CBM) in a critical system like airplane engine.

4. Other examples of digital twins

This section presents a variety of applications of digital twins in industry. We start with the behavior of a robot arm, whose shoulder is fixed at the original in a plane, the reference plane, see Pistone and Vicario, 2010. The robot arm is made up of a number m of segments; each segment has a length L_i and is at an angle θ_i, $i = 1, 2, \ldots, m$, with respect to the horizontal coordinate axis of the plane. The controlled factors are the $2m$ variables lengths $L_i \in [0, 1]$ and angles $\theta_i \in [0, 2\pi]$; the response y is the Euclidean distance between the end of the arm and the origin of the plane.

Another example is a simulation model of an analog integrated circuit Lo et al, (2000). The input variables are the different circuit parameters (transistor characteristics) and the response is the measurement of the circuit performance, such as the output voltage. The interest of the design engineers is to study the variation of the output current due to external and internal noises. and find the best combination of circuit parameters so that the output response variability is minimized. In yet another example, a finite element model (FEM) is used to design an engine block and a head joint sealing assembly in automotive internal combustion engines (Bray, 1981). For an introduction to FEM see Zienkiewicz, 1971. This type of design is quite complex since multiple components (block and head structures, gasket and fasteners) are involved, and the geometry is convoluted if sealing of combustion products, high pressure oil, oil drain, and coolant have to be maintained properly. The selection of the parameter setting in the assembly may not be analyzed separately, owing to the presence of strong assembly interaction effects. In addition, design decisions for this system must be made during the product development stage, prior to the availability of a physical prototype. FEM is used to capture the effects of the three-dimensional part geometry, the compliance in the components, the nonlinear gasket material properties and contact interface among block, gasket, head, and fasteners.

Digital twins have been used to simulate of the Rockwell C hardness test to assessing resistance of materials: (Romano et al, 2000). In such tests an indenter is pressed into the material and the depth of penetration is measured. After performing extensive numerical experiments, a few experimental runs were conducted in the lab to validate the digital twin model. Measurement uncertainty obtained by physical experiments conducted in the lab closely matched that obtained by the simulation experiment.

A thermal spraying process where the in-process parameters cannot be measured and soft sensors derived from a digital twin is presented in Rudak et al (2014). Soft sensors are sensor type data derived from a model. These are used when data cannot be collected, like in the thermal spraying process, or expensive to collect.

A combination of physical and simulated experiments to fine-tune predictions for the manufacturing process injection moulding process outputs is used in Villarreal-Marroquın et al (2016), see also Reese et al (2004). The authors design both physical and simulation experiments. The values of the performance measures are estimated at a grid of process control variables, and a set of predicted Pareto solutions is identified using non-dominance criteria. Refinements of the obtained solutions are also proposed. The physical process implementing a subset of the solution is used for the validation procedure. The next section is about the calibration of digital twins.

5. Calibration of digital twins

An example of a continuous production line that incorporates several sequential units where a biological raw material (oil) is transformed into biodiesel is provided in Brasio et al (2018). In the reactor, at an appropriate temperature and under the action of a catalyst, the oil is subject to transestericationby methanol. In order to improve the reaction, methanol is added in stoichiometric excess. The mixture leaving the reactor therefore contains not only the products of reaction (ester and glycerol) but also significant non-reacted methanol and traces of non or only partially reacted oil (triglycerides, diglycerides, and monoglycerides). To attain commercially acceptable biodiesel, there is need to isolate the produced ester to a purity of at least 96.5%, which requires to submit the obtained mixture to several separation procedures. At first, the separation of the immiscible compounds glycerol and ester is tackled by gravitational settling. The mixture is then passed through a heat exchanger to cool it down before it enters the decanter unit. In the decanter, the mixture separates into two phases. Glycerol (denser) tends to migrate to the down heavy phase while ester (less dense) is prone to go to the upper light phase. The light phase, collected after overflowing a baffle, is sent to the washer where water is used to wash away the existing methanol, catalyst, and any remaining glycerol. Only after having dried up the remnant water, biodiesel is finally obtained. All this is included in a digital twin. Its main elements include a reactor, a heat exchanger, a decanter, valves, filters, washer, dryer and controllers. All these elements are represented using mathematical equations. To use this complex simulator, one needs to validate and calibrate it. This applies to the individual parts and the integrated system which includes interactions between the various elements. Specifically, one is interested to investigate the impact of faults or degradation in one part of the system on its overall performance.

The idea of model inadequacy as the difference between a computer code's output (using the 'best' input settings) and reality is introduced in Kennedy and Ohagan (2001). They list problems that concern modelers and model users, the sources of uncertainty that arise when tackling such problems, and many aspects of modelling with Gaussian processes.

Computer models generally require two distinct groups of inputs. One group is the (unknown) parameter set about which we wish to make inferences; these are the calibration inputs. It is assumed that there is a true but unknown parameter set with the property that if the model were run with these values, it would reproduce the observations with a given model inadequacy term and noise. The other group comprises all the other model inputs whose value might change when the calibration set is fixed.

The calibration data consists of n physical observations $\mathbf{z} = (z_1,\ldots, z_n)$ representing observations of $f(x_i)$, $I = 1, \ldots, n$, the outputs of a model $f(\bullet)$ at values $\mathbf{x} = (x_1,\ldots, x_n)$. The simulation model is subject to error. Calibration can be conducted using frequentists and Bayesian methods. The Bayesian paradigm allows the incorporation of a priori knowledge (beliefs) about the parameters of any model being used. The code output is computed at a new set of parameters. Under the Bayesian view the true value of the code output is a random variable, drawn from a distribution that is conditioned by prior knowledge and previous computer runs. The computer code is thus viewed as a random function. Assuming smoothness, if the code has been run before at a slightly different point in the parameter space, one can assert with some confidence that the new run output shouldn't be too different from the last run. An emulator of the computer simulator, Simpson et al, 2001, Crary, 2002, Bates et al 2006, provides an estimate that is close to the value that code itself would produce. However, even a good model may be significantly affected by uncertainty in the input.

Quantification of the uncertainty distribution is needed to reduce uncertainty in predictions. Moreover, in order to derive good predictions, calibration is necessary by using observational data. In calibration you alter the model's parameters to make it fit the data.

A framework combining validation and calibration is proposed in Bayari et al (2007). The authors fit an emulator using only simulated data. The Bayesian methodology they propose combines validation and calibrations by incorporating the posterior distribution of the tuning parameters in the overall assessment of uncertainties. This is important because simply replacing a tuning parameter by an optimally "tuned" parameter using least squares hides the interaction between bias and tuning and can lead to overly optimistic assessments of validity.

An approach for modularized calibration is developed in Gramcy et al (2015). Such a decoupling has computational advantages and therefore applies to large simulations. They apply a flexible derivative-free optimization method to pair local approximate Gaussian process. Their proposal yields excellent predictions even when the emulator is far from the true value. This is attributed to the flexibility of coupled nonparametric regression models. The next section discusses the use of digital twins in education processes.

6. Digital twins in training and educational processes

Introductory engineering courses in academia combines the study of basic concepts with hands-on practice using tasks that resemble problems faced by engineers and technicians in their professional lives. Such practice exposes students to the nature and challenges of their future profession and fosters their motivation to learn following graduation. Industrial robotics laboratories, generally implement three types of learning scenarios setting up a robot system, programming different industrial robots and performing advanced robot-handling tasks. These laboratories offer learning practice in hands-on, virtual, and remote environments (Verner et al, 2019). In hands-on environments, students are present in the laboratory and directly control the installed robot systems. In virtual environments, they work with computer simulations of robot systems. In remote environments, the students are distant from the laboratory and control the robot systems through teleoperation. To perform robot system setup, programming, and operation assignments, the student needs immediate and detailed visual information from the robot workspace. In the hands-on environment, the student is near the robot system, and so all needed information is acquired directly through observation. In the remote-control system, visual feedback is transmitted from video cameras via a computer screen, and so is incomplete and delayed. In the virtual environment, the student works with a graphic simulation of the robot system on the computer screen, under limitations of the given software.

Simulators of batch and continuous processes are used both as digital twins and as educational infrastructures (Reis and Kenett, 2017). One important example is the Tennessee Eastman (TE) process simulator[d]e. The TE process plant involves five major units: a reactor, a condenser, a vapor-liquid separator, a product stripper and a recycle compressor. The plant produces two liquid products from four gaseous reactants through a reaction system composed of four irreversible and exothermic reactions. It also produces an inert product and a by-product purged as vapors from the system through the vapor-liquid separator. Reactants flow into a reactor where the reaction takes place. The output from the reactor is fed to a

[d] https://github.com/camaramm/tennessee-eastman-profBraatz
[e] http://depts.washington.edu/control/LARRY/TE/download.html

condenser. Some non-condensable vapors join the liquid products, but the following vapor-liquid separator again splits the substances into separate flows. Vapor is partially recycled and partially purged together with the inert product and the by-product. The stripper separates the remaining reactants from the liquid and another reactant is added to the product. The final products then exits the process and the remaining reactants are recycled. The TE process has 12 manipulated variables and 41 measured variables. The combination of mass ratios and production rates of the final products define six different operating modes of the process. The user can also choose to activate 20 pre-set process disturbances. TE is a complex simulator that can be used in advanced courses.

Educational activities, in life-long learning and academic environments, link concepts, mathematical tools and applications[fg]. This is especially true in reskilling the workforce to an adequate analytic level required by Industry 4.0 implementations. In this context, student engagement is an important activity aiming at effective acquisition of skills and competences. Flipped class learning, for example, is a pedagogical approach reinforcing the link between theory and practice by calling on trainees to play an active role in the learning process. The learning is often done by off class individual activities of students supported by on-line material. The classroom is a discussion arena where alternative interpretations are discussed. This has been used in both academic and adult education programs. Training with robots requires to engage students in the learning process, from problem elicitation and definition, to the analysis of collected data, follow-up experimentation and drawing of conclusions (Verner et al, 2019).

Simulations play a key role in distance learning programs, such as Massive Open Online Courses (MOOCs) and on-line education. A hierarchical classification scheme (HCS) of educational process simulators reflecting their inherent complexity, and describe training situations with integrated, comprehensive and coherent pedagogical solutions is proposed in Reis and Kenett, 2017. HCS builds on three dimensions capturing different aspects of complexity, namely:

1) Presence of non-linear modelling elements in the simulated models [NL]
- o NL=1: Linear
- o NL=2: Non-Linear

2) Presence of time-dependent behavior (e.g., dynamics, non-stationarity) [TD]
- o TD=1: Static
- o TD=2: Time-dependent

3) Size of the simulated system [SI].
- o SI=1: Small-scale
- o SI=2: Large-scale

These three dimensions describe eight classes of simulators. Class 1 are basic simulators with NL=1, TF=1 and SI=1. They represent applications based on the generation of random numbers from common probability distributions. The pedagogical goals of Class 1 simulators are: (i) demonstrate the existence of variability, (ii) show how randomness represents ignorance about the process outcomes and (iii) illustrate consequences of variability and non-linearity as contributors to systems complexity. Class 5 simulators are non-linear simulators (NL=2, TD=1, SI=1). These correspond to controlled continuous processes, with process kept

[f] https://intelitek.com/
[g] https://phet.colorado.edu

in narrow windows of operation. In these cases, first order Taylor approximation of the response surface provides a good description of the system behaviour. However, when the system is tested in more extreme conditions, non-linear behaviour may show up. This class of simulators is used for teaching quality engineering courses. Class 8 simulators are simulators of advanced process optimization, control, monitoring and diagnostics. These are characterised by medium/large-scale non-linear dynamics (NL=2, TD=2, SI=2). An example of Class 8 simulators is the Tennessee Eastman process simulator mentioned earlier (Capaci et al 2019). A list of such simulators is provided in Kenett and Reis, 2017.

7. Discussion and future developments

This final section provides a discussion and sketch for further developments related to digital twins. We emphasize here three aspects: the general structure of industry 4 manufacturing, the role of soft sensors and the potential of prescriptive analytics.

In the fourth industrial revolution, advanced communication, computation, and control infrastructures are integrated into cyber-physical systems incorporating a network of multiple manufacturing systems. This involves the use of embedded systems, such as sensors and programmable logic controllers to achieve an unprecedented level of automation in manufacturing. With the extensive use of embedded systems in the manufacturing industry, the fourth industrial revolution has significantly improved the throughput, efficiency, and product quality in the entire manufacturing industry, while considerably reducing human labour (Kenett et al., 2020).

In the Industry 4.0 and Big Data era, there is an increasing interest in the exploitation of the huge amounts of industrial data that are being routinely collected and stored. Product quality variables tend to be registered less frequently than process data. This is due complex measurement protocols and to time and cost to extract process samples. Moreover, the laboratories testing for product quality often require expensive equipment and highly trained staff. This leads to slower acquisition rates, creating sparse data structures in the plant databases, also known as multirate data structures. These data sources are integrated and further explored through advanced process analytics methodologies with predictive, monitoring and diagnosis solutions (Kenett and Zacks, 2014). To mitigate the effects of long delays and low sampling rates for the product quality variables there has been an increase of interest to apply process analytics methods for developing inferential models that are able to provide real-time estimates of their values. These are known as soft sensors and are part of a digital twin platform. As in the example in section 3, to improve the diagnostic tools in rotating machines, virtual sensors that measure the Instantaneous Angular Speed (IAS) are used. A common approach in diagnostics of rotating machinery is based on vibrations. Vibration analysis is an effective method for detecting various faults and malfunctions. The various methods of vibration signal processing require knowing the rotational speed of the machine in question, since, with rotating parts, events occur at specific angular positions rather than at specific times. For this reason, an accurate estimate of the Instantaneous Angular Speed (IAS) is important for reliable diagnostics. Inaccurate angular speed, due to dynamic phenomena such as unbalance, misalignment or eccentricity, can mask the effects of incipient localized faults. In practice direct measurement of the rotational speed is often impossible, expensive, or inaccurate. One example is Leclère (2016) who introduced a new approach for robust IAS estimation called the “Multi-Order Probabilistic Approach” (MOPA).

Machine learning is naturally integrated in digital twin platforms. Most of the machine learning today is supervised learning, where the model learns from labeled examples. Other forms of learning enable the identification of patterns in the data. An example is reinforcement learning, where the model learns in an unsupervised way when taking actions in a given (simulated) environment. In most cases of reinforcement learning, such conditions wouldn't be possible in the real world. Training models could however be highly time consuming. The OpenAI Five[h] neural network took 180 years of effective play time to train and still lost to professional players of the game. In the digital twin environment you can repeat a scenario or do a test without breaking the system and reinforcement learning agents can find novel ways to optimize the system.

Prescriptive analytics provides information about possible situations or scenarios, available resources, past performance, and current performance, and suggests a course of action or strategy. It can be used to make decisions on any time horizon, from immediate, to long term. In the case of a system and subsystem, prescriptive analytics provides focused recommendations on what line replaceable unit (LRU) to change and when, in order to ensure ongoing operations.

In providing monitoring, diagnostic, prognostic and prescriptive analytic capabilities, digital twins play a central role. The digital twin integrates information from empirical models with first principle simulation applications. For examples of such hybrid models in biotechnology see von Stosch et al, 2017.

As the title implies, our goal in this paper is to provide a wide-angle perspective of digital twins, and their role in the fourth industrial revolution. We discuss a range of applications with a more in-depth case study of rotating machines. We emphasize that the topics presented in this article should be taught in the advanced study programs in the Master of Science in Systems Engineering program under the topic of Engineering 4.0.

References

Alian, H., Konforty, S., Ben-Simon, U., Klein, R., Tur, M. and Bortman, J. (2019). Bearing fault detection and fault size estimation using fiber-optic sensors. Mechanical Systems and Signal Processing, 120, 392-407.

ANSI (2010). ANSI/ISA-95.00.01 Enterprise-control System Integration – Part 1: Models and Terminology. Washington, DC: American National Standards Institute.

Bal M, Hashemipour M. (2009) Virtual Factory Approach for Implementation of Holonic Control Industrial Applications: A Case Study in Die-casting Industry. Robotics and Computer-Integrated Manufacturing.; 25:570–581.

Bates, R.A., Kenett, R.S., Steinberg D. and Wynn, H. (2006) Robust design using computer experiments, *Quality Technology and Quantitative Management*, 3(2): 161-177.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, C.L.and Tu, J. (2007) A Framework for Validation of Computer Models, *Technometrics*, 49(2):138-154.

Brasio, A., Romanenko, A. and Fernandes, N. (2018) Modeling, Dynamics, Optimization and Bioeconomics III, Springer Proceedings in Mathematics and Statistics; Springer International Publishing, 224: 127-146.

[h] https://openai.com/blog/openai-five/

Bray, A. (1981) The role of stress analysis in the design of Force Standard Transducers, SESA 1980 W. M. Murray Lecture, Experimental Mechanics, 21, 1-20.

Capaci, F., Vanhatalo, E., Kulahci, M. and Bergquist, B. (2019) The Revised Tennessee Eastman Process Simulator as Testbed for SPC and DoE Methods. *Quality Engineering*, 31(2): 212–229.

Choi S, Kim BH, Noh SD. (2014) A Diagnosis and Evaluation Method for Strategic Planning and Systematic Design of a Virtual Factory in Smart Manufacturing Systems. International Journal of Precision Engineering and Manufacturing; 16:1107–1115.

Crary, S.B. (2002) Design of computer experiments for metamodel generation. *Analog Integrated Circuits and Signal Processing*, 32: 7-16.

Dadon, I., N. Koren, R. Klein and J. Bortman (2018) A realistic dynamic model for gear fault diagnosis, Engineering Failure Analysis, Volume 84 pp. 77-100.

Debevec M, Simic M, Herakovic N. (2014) Virtual Factory as an Advanced Approach for Production Process Optimization. International Journal of Simulation Modeling. 13(1):66–78.

Gazizulin, D., Cohen, E., Klein, R and Bortman, J. (2019) Critical Rotating Machinery Protection by Integration of a "Fuse" Bearing. International Journal of Critical Infrastructure Protection.

Gazizulin, D., Cohen, E., Klein, R and Bortman, J. (2019) Critical Rotating Machinery Protection by Integration of a "Fuse" Bearing. International Journal of Critical Infrastructure Protection, 27, 100305. 10.1016/j.ijcip.2019.100305

Gazizulin, D., Rosado, L., Schneck, R., Klein, R. and Bortman, J. (2019). A New Efficient Rolling Element–Spall Edge Interaction Model. International Journal of Fatigue, 105330.

Gramacy, R. B., Bingham, D., Holloway, J. P., Grosskopf, Michael J., Kuranz, C., Rutter, E., Trantham, M. and Drake, R. (2015) Calibrating a large computer experiment simulating radiative shock hydrodynamics. *Annals of Applied Statistics* 9(3): 1141—1168.

Gruber, A. Yanovski, S. and Ben-Gal, I. (2020) Condition-Based Maintenance via a Targeted Bayesian Network Meta-Model in Kenett, R.S., Swarz, R. and Zonnenshain A. (editors) *Systems Engineering in the Fourth Industrial Revolution: Big data, Novel Technologies, and Modern Systems Engineering*, John Wiley and Sons.

Hints, R., Vanca, M., Terkaj, W., Marra, ED., Temperini, S., Banabic, D. (2011). A Virtual Factory Tool to Enhance the Integrated Design of Production Systems, Proceedings of the DET2011 7th International Conference on Digital Enterprise Technology, Athens, Greece; p. 28-30.

IMT (2013). Are Virtual Factories the Future of Manufacturing?. Thomasnet Industry News. Mar 13. http://news.thomasnet.com/IMT/2013/03/13/are-virtual-factories-the-future-of-manufacturing/

Jain, S., Shao, G. (2014) Proceedings of the Winter Simulation Conference. Piscataway, NJ: IEEE; 2014. Virtual Factory Revisited for Manufacturing Data Analytics; p. 887-898.

Kenett, R. S., and Zacks, S. (2014) *Modern Industrial Statistics: With Applications in R, Minitab and JMP* (2nd ed.), John Wiley and Sons.

Kenett, R.S., Swarz, R. and Zonnenshain A. (2020) *Systems Engineering in the Fourth Industrial Revolution: Big data, Novel Technologies, and Modern Systems Engineering*, John Wiley and Sons.

Kenett, R.S., Zonnenshain, A. and Fortuna, G. (2016) A road map for applied data sciences supporting sustainability in advanced manufacturing: the information Quality dimensions, Procedia Manufacturing, 21, pp 141-148.

Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussions). Journal of the Royal Statistical Society B, 63: 425–464.

Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models (with discussions). *Journal of the Royal Statistical Society* B, 63: 425–464.

Kogan, G. R. Klein, J. Bortman (2018), A physics-based algorithm for the estimation of bearing spall width using vibrations, Mechanical Systems and Signal Processing, Volume 104, Pages 398-414.

Leclère, H. (2016) A multi-order probabilistic approach for Instantaneous Angular Speed tracking," Mechanical Systems and Signal Processing, pp. 375-386.

Lo Y. K., Zhang W. J. and Han, M. (2000) Application of the uniform design to quality engineering, Journal of the. Chinese Statistical Association, 38, 411–428.

Madar, E., Klein, R. and Bortman, J. (2019). Contribution of dynamic modeling to prognostics of rotating machinery. Mechanical Systems and Signal Processing, 123, 496-512.

Mario, H., Tobias, P., Boris, O. (2015) . Design Principles for Industrie 4.0 Scenarios: A Literature Review. Working Paper No. 01 / 2015. Technische Universität Dortmund; Dortmund, Germany

Pistone, G. and Vicario, G. (2010) Comparing and generating Latin Hypercube designs in Kriging models, AStA Adv Stat Anal 94: 353.

Reese C., Wilson A., Hamada M., Martz H. and Ryan K. (2004) Integrated Analysis of Computer and Physical Experiments, *Technometrics*, 46(2): 153-164.

Reis, M. and Kenett, R.S. (2017) A structured overview on the use of computational simulators for teaching statistical methods, *Quality Engineering*, 29(4): 730-744.

Romano D., Varetto M. and Vicario G. (2004) Multiresponse Robust Design: a General Framework Based on Combined Array, Journal of Quality Technology, 36(1): 27-37.

Romano D., Vicario G., and Galetto M. (2000) Robust FEM experiments, in: Proce. 3rd International conference on Quality, Reliability and Maintenance, Mc-Nulty editor, 171–174, Oxford, UK.

Rudak, N., Kuhnt, S. and Riccomagno E. (2014) Modelling of a Thermal Spraying Process by Gaussian Chain Graphs, *Quality Technology and Quantitative Management*, 11(1): 85-98.

Santner, T. J., Williams, B. J. and Notz, W. I. (2003) *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York.

Simpson, T. W., Peplinski, J. D., Koch, P. N. and Allen, J. K. (2001) Metamodels for Computer-based Engineering Design. Survey and Recommendations. *Engineering with Computers*, 17: 129-150.

Terkaj W, Tolio T, Urgo M. (2015) A Virtual Factory Approach for in Situ Simulation to Support Production and Maintenance Planning. CIRP Annals – Manufacturing Technology. ; 64:451–454.

Tolio, T., Sacco, M., Terkaj, W., Urgo, M. (2013) Virtual Factory: an Integrated Framework for Manufacturing Systems Design and Analysis, Proceedings of 46th CIRP Conference on Manufacturing Systems. Amsterdam: Elsevier p. 25-30.\

Verner, I., Cuperman, Gamer, S. and Polishuk, A. (2019) Training Robot Manipulation Skills through Practice with Digital Twin of Baxter, International Journal of Online and Biomedical Engineering, 15(9): 58-70.

Villarreal-Marroquın M.G., Po-Hsu Chen, Mulyana, R., J. Santner T.J., Dean A. M. and Castro J.M. (2016) Multiobjective Optimization of Injection Molding Using a Calibrated Predictor Based on Physical and Simulated Data. *Polymer Engineering and Science*. DOI 10.1002/pen.24405.

Volvo Group Global (2017) Virtual twin plant shorten lead times https://www.volvogroup.com/en-en/news/2017/mar/virtual-twin-plant-shorten-lead-times.html

von Stosch, M., Oliveira, R., Peres, J., and de Azevedo, S. (2014). Hybrid semi-parametric modelling in process systems engineering: Past, present and future. Computers and Chemical Engineering, 60(0), 86–101.

Zienkiewicz O. C. (1971) *The Finite Element Method in Engineering Science*, McGraw-Hill, New York.