Challenges of modeling and analysis in cybermanufacturing: a review from a machine learning and computation perspective

SungKu Kang[1] · Ran Jin[1] · Xinwei Deng[1] · Ron S. Kenett[2]

Abstract

In Industry 4.0, smart manufacturing is facing its next stage, cybermanufacturing, founded upon advanced communication, computation, and control infrastructure. Cybermanufacturing will unleash the potential of multi-modal manufacturing data, and provide a new perspective called computation service, as a part of service-oriented architecture (SOA), where on-demand computation requests throughout manufacturing operations are seamlessly satisfied by data analytics and machine learning. However, the complexity of information technology infrastructure leads to fundamental challenges in modeling and analysis under cybermanufacturing, ranging from information-poor datasets to a lack of reproducibility of analytical studies. Nevertheless, existing reviews have focused on the overall architecture of cybermanufacturing/SOA or its technical components (e.g., communication protocol), rather than the potential bottleneck of computation service with respect to modeling and analysis. In this paper, we review the fundamental challenges with respect to modeling and analysis in cybermanufacturing. Then, we introduce the existing efforts in computation pipeline recommendation, which aims at identifying an optimal sequence of method options for data analytics/machine learning without time-consuming trial-and-error. We envision computation pipeline recommendation as a promising research field to address the fundamental challenges in cybermanufacturing. We also expect that computation pipeline recommendation can be a driving force to flexible and resilient manufacturing operations in the post-COVID-19 industry.

Keywords Computation pipelines · Cybermanufacturing · Industry 4.0 · Machine learning · Manufacturing modeling and analysis

Introduction

In the era of Industry 4.0, advanced communication, computation, and control infrastructures are integrated into cyber-physical systems (CPS) incorporating a network of multiple manufacturing systems. Such a CPS opens a new paradigm of smart manufacturing, called cybermanufacturing. Conventional smart manufacturing has relied on traditional data-driven decision-making methods (e.g., statistical process control) to improve the performance of individual manufacturing systems. On the other hand, the next smart manufacturing, cybermanufacturing, is founded upon a CPS that constitutes a physical entity (e.g., physical equipment with sensor systems) and its corresponding cyber entity (e.g., a computer simulation model of the physical system, also referred to as digital twin (Kendrik et al. 2020)). Thus, the concept of digital twins can be closely integrated at all scales and levels (i.e., from individual equipment level to systems and networks of equipment) in the context of manufacturing operations, while the application of the general concept of digital twin is not restricted to cybermanufacturing. In cybermanufacturing, the behavior of physical entities can be timely sensed, predicted, and controlled via cyber entities.

Cybermanufacturing is the next stage of smart manufacturing, and it is a significant leap from conventional smart

✉ Ron S. Kenett
  ron@kpa-group.com

  SungKu Kang
  kangsungku@vt.edu

  Ran Jin
  jran5@vt.edu

  Xinwei Deng
  xdeng@vt.edu

[1] Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

[2] KPA Group, the Samuel Neaman Institute, Technion, Israel and University of Turin, Turin, Italy

manufacturing. Specifically, cybermanufacturing allows recording and registering all manufacturing events to improve the availability of manufacturing data, as well as provides ubiquitous computation resources and autonomous robotics to improve manufacturing efficiency and flexibility (Chen et al. 2018). With state-of-the-art machine learning methods, cybermanufacturing is expected to unleash the power of multi-modal manufacturing data, thus enhancing manufacturing efficiency, cost reduction, product quality, flexibility, and domain knowledge discovery. For example, case studies report on the applications of cybermanufacturing to innovate important sectors of manufacturing, including manufacturing logistics (Strandhagen et al. 2017), anomaly detection (Stojanovic et al. 2016; Pagani et al. 2020; Yan et al. 2020), and quality improvement (Oliff and Liu 2017). In addition, not only large manufacturing corporations like Stanley Black & Decker (Cisco 2019) and Siemens (Siemans 2019), but also small/medium-sized companies can benefit from cybermanufacturing. For example, a recent study has demonstrated the feasibility of adopting cybermanufacturing in small/medium-sized spring manufacturing companies by utilizing affordable sensor systems (Kuo et al. 2017). In other words, the transition into cybermanufacturing is not only about maintaining competitiveness, but also becoming essential to survival in Industry 4.0.
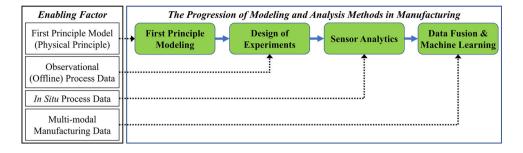
In the meantime, disruptive events, such as the COVID-19 pandemic, are calling for a more resilient manufacturing industry from unexpected downs of facilities, supply chains, and human resources (Loayza and Pennings 2020; Ivanov and Dolgui 2020). This accelerates the adoption of data-driven decision-making methods in manufacturing, so that the majority of manufacturing decisions in production, supply chain, quality engineering, and reliability engineering will be automatically made. This will allow the workforce to focus more on the insight-needed and situation-dependent tasks, such as business planning and new process/product development, rather than the basic and repetitive tasks (Chen and Jin 2017). In this direction, the concept of computation services in manufacturing has been proposed (Chen and Jin 2021), to describe a situation where manufacturing data is automatically collected and processed in ubiquitous computation units, such as the Cloud and the Fog nodes (Chen et al. 2018; Zhang et al. 2019). The goal is to provide real-time or online computation results and meet on-demand computation requests from manufacturing processes, systems, and users for decision-making. Thus, minimal labor-intensive training and tuning works are needed to effectively use data analytics and machine learning methods in such a computation service. In this paper, we focus on the "modeling and analysis" type of problems in manufacturing (Van Luttervelt et al. 1998; Shi 2006), which has served as the foundation for process monitoring, root-cause diagnosis, and control (Monostori and Prohaszka 1993; Duffie and Malmberg 1987; Prabhu

and Duffie 1995). Specifically, we review the modeling and analysis methods in smart manufacturing, and discuss the fundamental challenges to use those methods in cybermanufacturing. Other important research topics, such as sensor systems, computation infrastructure, production planning, and control/robotics, will not be discussed in this paper. For more detailed information about the advancement of sensor, robotics technologies, and human-machine interaction, the readers are referred to the recent reviews (Schütze et al. 2018; Bahrin et al. 2016; Kendrik et al. 2020; Cadavid et al. 2020; Derigent et al. 2020; Ansari et al. 2018; Paelke et al. 2020).

To streamline computation service in cybermanufacturing, it is necessary to identify a proper configuration of data analytics and/or machine learning methods providing the best computation performances for a contextualized computation task, such as variation analysis and anomaly detection. Here, one effective configuration is called a "computation pipeline", which is a sequence of method options involving data collection, data preprocessing, data analytics methods, etc. (Sparks et al. 2017; Chen and Jin 2018; Chen and Jin 2021; Chen et al. 2021). Adopting the concept of computation pipeline can comprehensively show how the data are sourced and processed to generate computation outcomes (Chen and Jin 2018). In cybermanufacturing, efficient identification of a proper computation pipeline under a given contextualized computation task ensures the resiliency of cybermanufacturing by improving the flexibility of the configuration of machine learning methods. In this paper, we will explore the recent advancement in selecting the computation pipelines, especially focusing on computation pipeline recommendation.

The remainder of this work is organized as follows. Section 2 provides an overview of smart manufacturing with respect to modeling and analysis. Section 3 focuses on a systematic review of the challenges of modeling and analysis in cybermanufacturing. Then, we review the computation services and computation pipelines as a promising research area in cybermanufacturing in Sect. 4. Lastly, we summarize the contribution of this paper with some discussions in Sect. 5. We envision that active investigation of computation pipeline selection/recommendation will catalyze the adoption of cybermanufacturing by providing rapid and effective decision-making in the post-COVID-19 industry.

Revisiting smart manufacturing: a modeling and analysis perspective

The third industrial revolution involved the use of simple embedded systems, such as sensors and programmable logic controllers to achieve an unprecedented level of automation in manufacturing. With the extensive use of embedded

Fig. 1 The progression of modeling and analysis methods in the manufacturing industry

systems in the manufacturing industry, the third industrial revolution has significantly improved the throughput, efficiency, and product quality in the entire manufacturing industry, while considerably reducing human labor. This innovation opened the era of "smart manufacturing", which is capable of utilizing the sensor data to enable data-driven decision-making (e.g., variation analysis and anomaly detection) and then precisely controlling the equipment with numerical controllers (Kenett et al. 2019). However, smart manufacturing has experienced a stagnant period due to its fundamental limitations, which calls for the transition to its next stage: cybermanufacturing.

In the rest of this section, we revisit the progression of modeling and analysis under smart manufacturing in Sect. 2.1. Then, we review the limitation of conventional smart manufacturing, and the advancement of enabling technologies that stimulates the transition into cybermanufacturing in Sect. 2.2.

Progression of modeling and analysis in conventional smart manufacturing

In the manufacturing industry, modeling and analysis involve quantifying the relationship between process variables and manufacturing performance variables (e.g., product quality) as well as the inter-relationships among process variables for accurate prediction of the results of manufacturing operations (Van Luttervelt et al. 1998). Modeling and analysis play the most important role in manufacturing, by providing the foundation for process monitoring, root-cause diagnosis, and control (Monostori and Prohaszka 1993; Duffie and Malmberg 1987; Prabhu and Duffie 1995). The modeling and analysis methods in manufacturing have evolved in accordance with the advancement of domain knowledge and technologies. Figure 1 shows the progression of modeling and analysis methods in the manufacturing industry.

Motivated by manufacturing modeling needs, first principle modeling have been firstly developed for modeling and analysis, based on physical principles in Thermodynamics, Fluid Mechanics, etc. This approach involves deriving exact solutions for ordinary/partial differential equations (ODE/PDEs) (e.g., Navier–Stokes equations) or solving an approximation of ODE/PDEs via numerical methods (e.g., finite element analysis, FEA) for efficiency (Meguid et al. 1999; Roberts et al. 2009; Wen et al. 2018). However, first principle modeling is limited when there is a significant gap between the assumption of physical principles (e.g., ideal boundary conditions) and the actual manufacturing condition. Therefore, the statistical design of experiments (DOE) became an active research area for modeling and analysis in manufacturing.

Originated from the design and analysis of agriculture experiments (Fisher 1919), DOE has been widely applied to manufacturing applications and computer simulations (Box et al. 2005; Wu and Hamada 2011; Santner et al. 2003; Kenett et al. 2013). The design and analysis of physical experiments provide valuable information about the manufacturing process under controllable settings. While the design and analysis of computer experiments can seek information on the process when the physical experiments are too costly to even impossible to conduct. Founded upon DOE, computer model calibration has been advanced (Kennedy and O'Hagan 2001; Oakley and O'Hagan 2002; Higdon et al. 2008), such that first principle models can be calibrated with the observational process data to provide more accurate results. In addition, DOE has also contributed to modeling and analysis via (1) variable screening, (2) treatment comparison, (3) uncertainty quantification, (4) system optimization, and (5) system robustness (Wu and Hamada 2011).

In the last few decades, the paradigm of smart manufacturing has significantly advanced sensing technologies, which allows the incorporation of *in situ* process variables in modeling and analysis. This advancement has enabled compensation of modeling and analysis methods based on *in situ* process data to improve product quality, efficiency, as well as the automation of manufacturing process. An example includes the compensation of machine tool positioning error based on the *in situ* sensor data from multiple thermal sensors (Yang et al. 1999; Yang and Ni 2005). In the meantime, the advancement of sensing technology also provides a data-rich environment, where a large amount of data can be collected from heterogeneous manufacturing systems in multiple modalities. Specifically, the manufacturing industry is known to generate more data than any other industries,

thus efficiently utilizing the data has become one of the most important factors to be successful in the manufacturing industry (Hartmann et al. 2015).

There are two promising directions for utilizing manufacturing data: data fusion and machine learning methods. Data fusion refers to the methods of integrating different models and data sources (e.g., first principle model and data-driven model) (Zhao et al. 2011) or different types of data sets (e.g. observational data and DOE data) (Jin and Deng 2015; Dalla Valle and Kenett 2018). In the manufacturing industry, data fusion has demonstrated its significant advantages in modeling and analysis by utilizing data with different sources, formats, or distributions (Zhao et al. 2011; Jin and Deng 2015; Deng and Jin 2015; Sun et al. 2017). On the other hand, machine learning refers to the methods of building a mathematical model based on data, such that the model can make predictions or decisions without being explicitly programmed to do so. Given abundant data, machine learning methods have shown superior performances in modeling complex manufacturing processes in the past few years. Recently, with the advancement of computation infrastructure and increased availability of manufacturing data, machine learning and artificial intelligence (AI) algorithms, such as deep learning, are actively adopted in the manufacturing industry (Wang et al. 2018a). The most notable feature of deep learning is that its multiple hidden layers allow learning important features purely from the complex manufacturing data (e.g., image or temporal data), without expert domain knowledge. For example, the convolutional neural network (CNN) and its variants, originally designed for image analysis, are well suited for the analysis of image data from manufacturing process. The CNN has been widely adopted for product quality inspection and fault assessment in manufacturing process (Park et al. 2016; Weimer et al. 2016). Another example is recurrent neural network (RNN) and its variants, originally designed for the analysis of temporal behavior, are suitable for the predictive analysis based on historical records. Therefore, RNN has been widely adopted for prognosis in manufacturing process (Malhi et al. 2011; Zhao et al. 2018). In the meantime, a known limitation of such AL algorithms is that they are mostly black-box models in nature with highly nonlinear model structure, thus not interpretable to human (Carvalho et al. 2019). Recent trends in machine learning community investigate model interpretability to improve the black-box models, such as the works on explainable machine learning (XAI) (Arrieta et al. 2020) and knowledge-based artificial neural network (KBANN) (Coatanéa et al. 2018). Such advancements are promoting a trend to develop machine learning methods for modeling and analysis in manufacturing (Dagli 2012). In this context, data fusion and machine learning methods are expected to play crucial roles in cybermanufacturing.
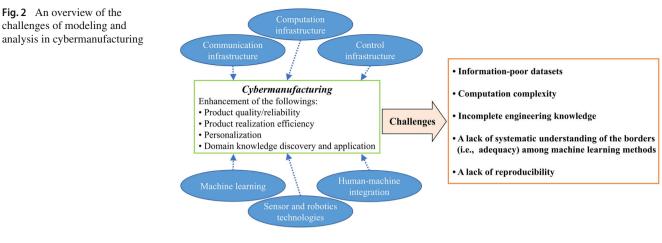
Transition into cybermanufacturing

Regardless of the success of smart manufacturing for decades, the fundamental limitations of conventional smart manufacturing have been proposed as follows. First, the scalability of the data fusion and machine learning methods is limited, as the manufacturing systems and cyber resources (e.g., computation resource, data storage) are not connected to each other. In addition, the manufacturing systems are not operated in a synergistic manner nor jointly using the data collected from different manufacturing systems. Second, conventional smart manufacturing cannot fully take advantage of state-of-the-art machine learning models nor use the potential of multi-modal manufacturing data, as the computation capability of simple embedded systems is limited. While utilizing external computation resources (i.e., Cloud computing service) can be an option to alleviate this issue, it cannot be a reliable solution since its reliability is highly affected by network conditions. This limits the efficiency and reliability of modeling and analysis under conventional smart manufacturing, especially where the given contextualized computation task is computationally demanding and/or requires responsive and reliable decision making. Third, the flexibility (i.e., the ability for re-configuration) of the entire system is limited, as conventional smart manufacturing systems typically assume limited flexibility in terms of its operation (e.g., production plan, facility layout, etc.). This requires extensive cost (i.e., additional human labor, downtime, and considerable lead time to stabilize the new configuration) when an adjustment or variant management is required due to disruptive events. In this context, there has been a pressing need to innovate smart manufacturing via enhanced communication, computation, and control infrastructure. In recent years, CPS has been actively adopted in the manufacturing industry to innovate important sectors of manufacturing (e.g., manufacturing logistics (Strandhagen et al. 2017), anomaly detection (Stojanovic et al. 2016; Pagani et al. 2020; Yan et al. 2020), and quality improvement (Oliff and Liu 2017)) aiming at addressing the aforementioned limitations (Tao et al. 2019). By incorporating CPS, cybermanufacturing is expected to enhance intelligent production planning, predictive maintenance, flexible/autonomous manufacturing process, and human-machine integration (Jeschke et al. 2017). There are three enabling technologies towards cybermanufacturing to form the infrastructure to support the future generation of machine learning and computation in manufacturing (Gorecky et al. 2014; Xiong et al. 2015; Tao et al. 2019; Jaloudi 2019).

Communication Infrastructure There have been two important advancements in communication infrastructure. First, the advancement of wireless communication technologies, such as 5G and Wi-Fi 6, enables a seamless connection among manufacturing systems, which allows efficient data

Fig. 2 An overview of the challenges of modeling and analysis in cybermanufacturing

Challenges of modeling and analysis in cybermanufacturing

Cybermanufacturing with advanced communication, computation, and control infrastructures is expected to enhance product quality/reliability, product realization efficiency, personalization, and domain knowledge discovery. However, cybermanufacturing faces several challenges of modeling and analysis with respect to machine learning and computation. Figure 2 shows an overview of the challenges.

In this section, we review five important challenges of modeling and analysis in cybermanufacturing: information-poor datasets, computation complexity, incomplete engineering knowledge, a lack of systematic understanding of the borders among different machine learning methods, and a lack of reproducibility.

Information-poor datasets

In this section, we make the distinction between data quality and information quality. Data quality is an intrinsic feature of the collected data. Information quality reflects the analysis and management of its outcomes, conditioned on the goals of the analysis. Data quality has been recognized as a serious challenge by statisticians and data analysts, since almost all data requires some pre-processing (e.g., filtering or cleaning) before it can be useful for a certain type of modeling and analysis via the transition to information (Kenett and Shmueli 2016). In that transition, many additional dimensions need to be considered in order to ensure information quality. To evaluate data quality, several assessment frameworks have been proposed, e.g. Wang et al. (1995). To ensure information quality, a framework based on four components and eight dimensions has been proposed in Kenett and Shmueli (2016), which is called InfoQ. The application of InfoQ in the context of Industry 4.0 was reviewed in (Kenett et al. 2018). In the InfoQ framework, information quality

collection, exchange, and analysis to improve efficiency and productivity in a synergistic manner (Cheng et al. 2018). Second, the advancement of standardized, open, and manufacturer-independent communication protocols, such as Message Queuing Telemetry Transport (MQTT), Open Platform Communications Unified Architecture (OPC UA), and Data Distribution Service (DDS), improves scalability and interoperability under cybermanufacturing (Profanter et al. 2019). Coupled with Time Sensitive Networking (TSN) standard, the protocols will allow deterministic real-time machine to machine communications with low latency, which is one of the most important features of cybermanufacturing (Arestova et al. 2021).

Computation Infrastructure There have been two important advancements in computation infrastructure. First, the capability of computation resources (e.g., graphics processing unit (GPU) and tensor processing unit (TPU)) have been significantly improved to afford computation-intensive tasks (e.g., analyzing video streams for inspection). Second, combined with communication infrastructure, such computation resources become ubiquitous, thus accessible as cloud node (high performance, but high latency due to remote connection) as well as Edge or Fog nodes (low performance, but low latency due to proximity) via the network.

Control Infrastructure There have been two important advancements in control infrastructure. First, the advancement of autonomous robotic systems (e.g., autonomous mobile robots (AMR) (Fragapane et al. 2020) and automated guided vehicles (AGV) (Mehami et al. 2018)) enhances the flexibility of production to respond to dynamic markets. Second, the advancement of human-machine integration technologies (e.g., exoskeletons (Huysamen et al. 2018) and augmented reality (Paelke 2014)) allows efficient collaboration between humans and machines to achieve high productivity as well as responsive decision-making.

is defined as the utility, U, in applying a method of analysis, f, to a data set, X, conditioned on the analysis goal, i.e., $InfoQ(U, f, X, g) = U(f(X|g))$. To assess InfoQ, the following eight dimensions are considered: 1) Data Resolution, 2) Data Structure, 3) Data Integration, 4) Temporal Relevance, 5) Chronology of Data and Goal, 6) Generalizability, 7) Operationalization and 8) Communication. Specifically, each of the eight dimensions is quantified with respect to each of the components U, f, X, and g by answering a set of relevant questions, and the then components are weighted to derive InfoQ value. For more details on the calculation of InfoQ in a practical setting, the readers are referred to the literature (Reis and Kenett 2018), and JMP (Sall et al. 2017) which is a suite of computer programs for statistical analysis providing an InfoQ add-on (Kenett and Shmueli 2016) to calculate InfoQ.

These frameworks provide important criteria to evaluate data and information quality so that one can quantitatively and/or qualitatively assess the data and information quality conforming to a specific contextualized computation task. In the era of Industry 4.0, the data quality issue is much more demanding than ever before, since modern sensing systems are capable of collecting an enormous amount of data over a long period of time from distributed heterogeneous manufacturing systems (Cai and Zhu 2015; Luthra and Mangla 2018). While such a sensing system can collect a massive amount of data for machine learning, it may also include a lot of poorly integrated or redundant information which results in so-called "information-poor" datasets.

Information-poor datasets lead to the following limitations in cybermanufacturing (Cai and Zhu 2015; Li et al. 2016; Dai et al. 2020; Ismail et al. 2019). First, machine learning methods may be misled by the redundant information (i.e., low signal-to-noise ratio), which significantly limits the effectiveness and efficiency of manufacturing modeling and analysis. Moreover, due to the tremendous size of datasets, it is difficult to assess the data quality in a reasonable amount of time. Second, since cybermanufacturing involves the collection of data from heterogeneous manufacturing systems, it brings a wide variety of data types and heterogeneous data structures. This leads to unbalanced data, thus makes it difficult to integrate data for efficient manufacturing modeling and analysis. Lastly, in the era of Industry 4.0, a single type of data involves multiple contextualized computation tasks. For example, the data related to equipment (e.g., equipment vibration) may be used for fault diagnosis, preventive maintenance scheduling, and quality prediction at the same time. However, as information quality depends on the objective of a contextualized computation task, a single dataset may or may not conform to the information quality standards required in the multiple contextualized computation tasks. This makes it challenging to ensure the information quality from a dataset.

Computation complexity

Cybermanufacturing focuses on the convergence of the physical entities (e.g., manufacturing equipment) and the cyber entities (e.g., simulated equipment), such that the dynamic changes of the physical entities can be predicted and analyzed through the corresponding cyber entities (Qi and Tao 2018). Therefore, the existing efforts aimed at achieving realistic cyber entities via accurate computer simulation models and machine learning methods (Qi and Tao 2018; Störmer et al. 2018). An example, in the context of decision tree, predictive analytics is the work of Ben-Gal et al. (2014). In that paper, the authors presented the dual information distance (DID) method, which selects features by considering both immediate contributions to the classification as well as their future potential effects. Specifically, the DID method constructs classification trees by finding the shortest paths over a graph of partitions that are defined by the selected features. The authors demonstrated that the method takes into account both the orthogonality between the selected partitions as well as the reduction of uncertainty on the class partition, such that it outperforms popular classifiers in terms of average depth and classification accuracy. However, computation complexity is becoming one of the most challenging issues in realizing realistic cyber entities. First, high-fidelity computer simulations, which enable accurate prediction of the behavior of physical entities via the cyber entities, are typically too time-consuming. This issue becomes more challenging when the manufacturing process involves computationally-intensive multiphysics computer simulations (Dbouk 2017). For example, recent work indicates that a single run of computer simulation for the fused deposition modeling process takes more than 20 hours (Li et al. 2018), which could result in a significant delay in updating the cyber entities. Second, as the advancement of sensing and communication systems enables the collection of enormous data from distributed manufacturing systems, machine learning methods to build the cyber entities require significantly higher computation cost and/or communication bandwidth across cybermanufacturing infrastructure.

The issue of computation complexity leads to the following limitations in cybermanufacturing (Singh et al. 2018; Modoni et al. 2019; Bevilacqua et al. 2020; Rasheed et al. 2020). First, computation complexity involving modeling complex behaviors of heterogeneous manufacturing systems and their interactions delays the manufacturing modeling and analysis for personalized demands, which limits the timeliness of personalization. This is an important issue, since cybermanufacturing aims at achieving efficient personalization of products by utilizing heterogeneous manufacturing systems connected. Second, computational complexity also affects the ability to generate data quality (Kenett and Shmueli 2016). Insufficient storage or computation power to

handle data with adequate resolution can negatively impact the delivery of outputs from analytical work to the right persons, in the right way, at the right time, thereby reducing information quality.

Incomplete engineering knowledge

Manufacturing processes involve complex physical mechanisms. Therefore, underlying engineering knowledge, such as cause-effect relationships, first principle models, and computer simulation models (e.g., finite element analysis, FEA) and design rules, may be incomplete. Here, incomplete knowledge can be due to 1) incomplete understanding of the underlying physical mechanism (e.g., first principle model, material properties), 2) incomplete information about model parameters, and 3) stochastic behavior or uncertainties associated with the system or numerical algorithms. To address the issue, computer model calibration (Kennedy and O'Hagan 2001; Higdon et al. 2008, 2013; Wong et al. 2017) has been continuously studied to compensate the incomplete knowledge with observational data, and design for variation (Reinman et al. 2012) has been studied to reduce the variation of products under the incomplete knowledge and uncertainties. However, in the era of Industry 4.0, it is becoming more challenging to address the issue for the following reasons. First, while the core of cybermanufacturing is to take advantage of multi-modal manufacturing data, the data can mislead the decision-making processes if incomplete knowledge (e.g., invalid assumption in modeling a manufacturing process) involves the interpretation of the data. Second, since the existing efforts typically assume a single or only a few manufacturing systems (Feng et al. 2017), existing computer experimental design, modeling and calibration are not easily scalable to the scenario where a lot of heterogeneous manufacturing systems connected in cybermanufacturing. In these cases, the dependability and effectiveness of cybermanufacturing may be questioned, especially in case a delicate and complex situation requires decision-making in a real-time manner (Broy et al. 2012). These issues are more common to the newly introduced additive manufacturing (Babu and Goodridge 2015; Yang et al. 2017; Dilberoglu et al. 2017; Jared et al. 2017; Li et al. 2018; Mahmoudi et al. 2018; Sabbaghi et al. 2018; Kenett et al. 2019).

In summary, incomplete engineering knowledge leads to limitations in cybermanufacturing. First, the manufacturing design process becomes inefficient when underlying engineering knowledge is incomplete. Specifically, it has been speculated that 75% of the cost involving product development is committed early in the engineering design process when the knowledge of the product is unclear and/or incomplete (Chandrasegaran et al. 2013). Second, when computer simulation or data-driven models for a manufacturing process are limited in scope, the models cannot provide adequate predictions for prognostics and health management in cybermanufacturing (Weiss et al. 2015). This can result in inefficient planning, maintenance and logistics due to the inaccurate prediction of equipment status (Davis et al. 2012; Edgar and Pistikopoulos 2018).

Lack of systematic understanding on adequacy of machine learning methods

Cybermanufacturing focuses on personalization and customized production, which will generate a wide variety of heterogeneous data (Thoben et al. 2017). In the meantime, the adequacy of a machine learning method to such heterogeneous data may be significantly different due to the underlying statistical characteristics (e.g., the distribution of data) and/or contextualized computation tasks (e.g., fault diagnosis or quality control a specific manufacturing process) (Chen and Jin 2018). Here, we call the different adequacy as the "border" of the machine learning methods. Thus, it is important to match a specific dataset/contextualized computation task with a proper machine learning method within the border to ensure the efficiency and effectiveness of manufacturing modeling and analysis.

In current practice, a typical paradigm to identify which machine learning method for use is often heuristic based on domain knowledge of a specific contextualized computation task and/or data scientist's personal experience in data analysis. Clearly, such a heuristic manner could require a large number of trial-and-errors for identifying an efficient and effective machine learning method under a given contextualized computation task. It calls for a systematic methodology to understand the border among different machine learning methods, especially in the field of manufacturing modeling and analysis.

There are several challenges in Industry 4.0 due to the lack of systematic understanding of the borders among different machine learning methods. We list some of them below. First, considering the heterogeneous manufacturing systems connected in cybermanufacturing, it will require considerable lead time for identifying a proper machine learning method for each manufacturing system and computation task. For example, for a thermal spray coating process using heterogeneous spray guns, it is reported that linear regression model worked well for one spray gun was not applicable to the other spray guns, due to the violation of the assumption of samples from the same underlying distribution (Chen and Jin 2018). Second, it is known that manufacturing processes and systems are likely to be dynamic in model relationship, due to a number of factors from raw materials, equipment status, and environment. For example, it is reported that the model parameters for crystal growth process should be adjusted based on the degradation level of the equipment (Jin et al. 2019). However, most of the machine learning methods can-

not generate dynamic models. Therefore, it will be beneficial to efficiently match the optimal machine learning method with the degradation levels. Third, cybermanufacturing often requires different accuracy of machine learning methods with the consideration of computational cost and utility costs. A lack of understanding of borders among machine learning methods could increase efforts to select the methods not only be adequate, but also be reliable and responsive. As shown in Kang et al. (2021), a tradeoff between the computational cost of designs and the accuracy of surrogate models could facilitate the identification of the feasible design region, which is crucial in the timeliness of personalized product realization in Industry 4.0. Meanwhile, the time latency in machine learning training process, and the unreliable computation due to computation node failure or loss of communication to Cloud will prohibit the use of advanced, but computation intensive algorithms.

A lack of reproducibility

When machine learning methods are employed, different researchers or practitioners tend to choose different configurations (e.g., splitting of the samples for training and testing), even when they analyze the same dataset (Botvinik-Nezer et al. 2020). This flexibility leads to difficulties in the reproducibility of machine learning methods, and needs to be accounted for and controlled in manufacturing modeling and analysis. However, even though there have been consistent efforts to address the issue in science (Kenett and Rubinstein 2017; Botvinik-Nezer et al. 2020), the manufacturing industry is less concerned with the issue of reproducibility of analytical studies (Kenett 2020). In other words, companies tend to overlook the experimental works designed to improve processes and products for reproducibility using adequate statistical criteria. In the meantime, fierce competition in the era of Industry 4.0 allows only short-term opportunities to try out new products and/or new process setups, which calls for ensuring the reproducibility of machine learning methods in a contextualized computation task.

A lack of reproducibility leads to the following challenges in cybermanufacturing. First, in the manufacturing industry, a lack of reproducibility of machine learning methods can result in misleading decision-making, which is very costly and time-consuming. For example, it is reported that around 50% of the costs incurred in new product development tend to be spent on unnecessary design iteration (Schütze and Störmer 2012), which can be avoided by accurate and reliable predictions. Second, since cybermanufacturing involves efficient utilization of heterogeneous manufacturing systems connected to cybermanufacturing network (Lee et al. 2015; Jeschke et al. 2017; Wang et al. 2020a), reproducibility should be ensured such that consistent product quality can be achieved across the cybermanufacturing network. Lastly,

a lack of reproducibility of machine learning methods can result in increased product variation, which can deteriorate customer satisfaction (Luo et al. 2005; Dharmesti and Nugroho 2013). It is an important issue, since improved customer satisfaction is one of the most important goals to be achieved in the context of Industry 4.0 (Bortolini et al. 2017; de Man and Strandhagen 2017; Bär et al. 2018).

A promising direction: computation pipelines in cybermanufacturing

In the post-COVID-19 industry, it is very important to automate the identification of an optimal configuration for machine learning methods and match the methods to the contextualized computation tasks (e.g., variation analysis and anomaly detection), including accuracy, responsiveness, reliability, interpretability, etc. (Wang et al. 2020b). In the literature, this matching problem has been formulated as a computation pipeline recommendation problem (Chen and Jin 2018). In the rest of this section, Sect. 4.1 will review the general concept of computation pipelines for machine learning. Then, Sect. 4.2 will review the state-of-art on the computation pipeline recommendation.

Concept of computation pipelines

The concept of computation pipelines for machine learning is suggested from the software engineering community to systematically organize a sequence of method options, including data collection, data preprocessing, data filtering, feature selection (optional), data-fusion/machine learning methods, computation, and post-processing (Chen and Jin 2018). For example, Scikit-learn, which is a machine learning library for Python, proposed a computation pipeline to assemble several steps that can be cross-validated together with different setting parameters (Pedregosa et al. 2011). Similarly, Google® Tensorflow (Abadi et al. 2016) and Pytorch (Paszke et al. 2019), which are the widely used deep learning platforms, suggested the idea of computational graph to organize computation pipeline for deep learning. These computation pipelines enhance the readiness and traceability to use machine learning and deep learning method options.

In the manufacturing industry, most relevant works involving computation pipelines have focused on constructing an autonomous framework to tune a specific computation pipeline or only a limited number of method options. Examples of such works include the application of computation pipelines for preventive maintenance operation (O'Donovan et al. 2015), fault prognostics (Kozjek et al. 2017), and production planning (Wang et al. 2018b). While the aforementioned works are applicable to a specific contextualized computation task, they cannot be adequate in different tasks
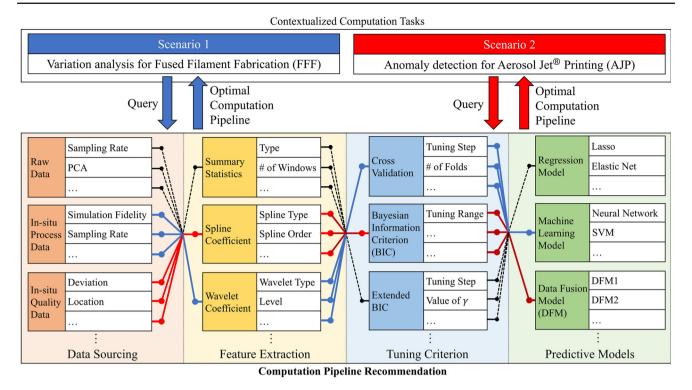
Fig. 3 An illustration of computation pipeline recommendation providing optimal computation pipelines for two different contextualized computation tasks (scenario 1: variation analysis for Fused Filament Fabrication (FFF), marked as the blue pipeline; and scenario 2: anomaly detection for Aerosol Jet® Printing (AJP), marked as the red pipeline) (Color figure online)

when modeling assumptions are violated (e.g., the underlying distribution of data is different). In other words, to ensure the effectiveness of manufacturing modeling and analysis, one should efficiently switch to a proper computation pipeline from a number of alternatives that fits well with the scenario. However, the current practice relies on trial-and-errors according to domain knowledge and experiences, which is too time-consuming to identify the optimal method options for a proper computation pipeline.

Computation pipeline recommendation

Computation pipeline recommendation is a relatively new research field aiming at identifying a proper computation pipeline from a number of alternative configurations. Figure 3 shows an example of computation pipeline recommendation. Given different contextualized computation tasks, computation pipeline recommendation efficiently explores available computation pipelines (shown as the connections between the method options in Fig. 3) and provides the optimal pipeline for each of the contextualized computation tasks. This example illustrates providing different optimal computation pipelines for scenario 1 (marked as red computation pipeline in Fig. 3) and scenario 2 (marked as blue computation pipeline in Fig. 3).

Sparks et al. (2017) and Chen and Jin (2018) are the pioneers of computation pipeline recommendation. Sparks et al. (2017) presented a system called KeystoneML, which enables the development of end-to-end computation pipelines for machine learning. Specifically, by capturing the end-to-end machine learning application, KeystoneML could optimize the whole-pipeline such that it can automatically adapt to changes in data, hardware, and other environmental factors. On the other hand, Chen and Jin (2018) proposed a *learning-to-rank* method to rank the pipelines with *Top-N* prediction accuracy to identify proper pipelines for quality prediction. They demonstrated the effectiveness and efficiency of the proposed method via multiple datasets from thermal spray coating, aerosol jet® printing, and fused deposition modeling. For the datasets, their method could ease the use of data-fusion and machine learning methods and effectively avoid unnecessary computation workloads involving exhaustive exploration of all the available computation pipeline. Similar works relevant to computation pipeline recommendation include the work called "data analytics as a service (Ardagna et al. 2018)". Specifically, the approach aims at identifying proper computation pipeline conforming to a specific set of requirements and/or platforms, by utilizing a model-based approach founded upon a domain ontology on big data services.

Computation pipeline recommendation has been adopted not only to improve the quality prediction, but also to improve other important sectors of Industry 4.0: informative visualization and efficient human-machine collaboration. In the following-up research of Chen et al. (2021), they extended the concept of computation pipeline recommendation to the personalization of a visualization system, which is called Personalized Recommender System for Information visualization Methods via Extended matrix completion (PRIME) (Chen et al. 2021). The main improvement of PRIME over the previous method lies in incorporating the wearable sensor data for pipeline recommendation, such that pipeline recommendation can be adopted to improve human-computer interaction to acquire insights from complex datasets. Specifically, PRIME quantitatively models covariates (i.e., wearable sensor data) to predict recommendation scores (e.g., perceived complexity, mental workload, etc.) for users to adapt the visualization specific to the contextualized computation task. In addition, PRIME can make accurate recommendations even for new users or new contextualized computation tasks based on historical wearable sensor signals and recommendation scores. Chen et al. (2021) demonstrated that PRIME could achieve satisfactory recommendation accuracy for adapting visualization, even when there are limited historical datasets. This capability contributes to designing a new generation of visualization systems that adapt to users' real-time status. PRIME can support researchers in reducing the sample size requirements to quantify individual differences, and practitioners in adapting visualizations according to user states and contextualized computation tasks in a real-time manner.

Computation pipeline recommendation is expected to be generalized such that it can stimulate the broader adoption of machine learning in the manufacturing industry, in the context of cybermanufacturing. For example, Chen and Jin (2021) proposed an improved computation pipeline recommendation framework, named as Adaptive computation Pipelines (AdaPipe) (Chen and Jin 2021). The main improvement of AdaPipe over the previous methods lies in considering the similarities of computation pipelines from crowdsourcing and word embedding, and features of contextualized computation tasks (i.e., process types, manufacturing settings, and quality specifications). In addition, AdaPipe can be adopted to recommend computation pipelines for a wide variety of contextualized computation tasks, including the lowest prediction error, lowest time latency, etc.. They demonstrated the effectiveness and efficiency of AdaPipe via 60 bootstrapped data sets from thermal spray coating, printed electronics, and additive manufacturing processes, and 27 computation pipelines. The generalization and further advancement of computation pipeline recommendation will provide a systematic and efficient methodology to explore a number of method options without time-consuming trial-

and-error. We envision that this will contribute to addressing the key challenges reviewed in this paper by providing a proper computation pipeline if a contextualized computation task is susceptible to either of the key challenges. To achieve the goal, the future direction for computation pipeline recommendation lies in incorporating the issues related to the challenges (e.g., information quality, computation complexity, available engineering knowledge, etc.) in its recommendation procedure.

Conclusion

In the era of Industry 4.0, the integration of cybermanufacturing and state-of-the-art machine learning methods is expected to greatly improve product quality, product realization efficiency, personalization, and domain knowledge discovery. In this work, we have focused on reviewing several key challenges in modeling and analysis in the transition into cybermanufacturing. These include required learning times in heterogeneous manufacturing systems, matching of optimal machine learning methods with equipment degradation levels, and the adaptation of machine learning methods with considerations of computational cost and utility costs. The overall goal is to optimize system reliability and responsiveness. Aligned with the challenges, we have provided a promising direction of using computation pipeline recommendation as an enabling technology of cybermanufacturing. The generalization of computation pipeline for all algorithms will facilitate the broader adoption of machine learning in the manufacturing industry by efficiently providing an optimal configuration of machine learning methods without extensive trial-and-error. Note that efficient deployment of machine learning methods in cybermanufacturing becomes more important than ever before, the use of computation pipeline can be a driving force to flexible and resilient manufacturing operations in the post-COVID-19 industry. Areas for further work include (1) investigating the impact of information quality in computation, (2) better understanding the reproducibility and repeatability of computational flows, (3) integrating analytic recommendations with constraints from communication, computation, and control, and (4) further extension of computation pipeline to improve the broader area of human-machine interaction (e.g., mutual learning between human and machine). This paper provides an initial mapping of such challenges.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., & Irving, G., Isard, M., et al. (2016) Tensorflow: A system for large-scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp 265–283

Ansari, F., Erol, S., & Sihn, W. (2018). Rethinking human-machine learning in industry 4.0: How does the paradigm shift treat the role of human learning? *Procedia Manufacturing, 23*, 117–122.

Ardagna, C. A., Bellandi, V., Bezzi, M., Ceravolo, P., Damiani, E., & Hebert, C. (2018). Model-based big data analytics-as-a-service: Take big data to the next level. *IEEE Transactions on Services Computing, 14*(2), 516–529.

Arestova, A., Martin, M., Hielscher, K. S. J., & German, R. (2021). A service-oriented real-time communication scheme for AUTOSAR adaptive using OPC UA and time-sensitive networking. *Sensors, 21*(7), 2337.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Babu, S., & Goodridge, R. (2015). *Additive manufacturing*. Taylor & Francis.

Bahrin, M. A. K., Othman, M. F., Azli, N. N., & Talib, M. F. (2016). Industry 4.0: A review on industrial automation and robotic. *Jurnal Teknologi, 78*(6–13), 137–143.

Bär, K., Herbert-Hansen, Z. N. L., & Khalid, W. (2018). Considering Industry 4.0 aspects in the supply chain for an SME. *Production Engineering, 12*(6), 747–758.

Ben-Gal, I., Dana, A., Shkolnik, N., & Singer, G. (2014). Efficient construction of decision trees by the dual information distance method. *Quality Technology and Quantitative Management, 11*(1), 133–147.

Bevilacqua, M., Bottani, E., Ciarapica, F. E., Costantino, F., Di Donato, L., Ferraro, A., Mazzuto, G., Monteriù, A., Nardini, G., Ortenzi, M., et al. (2020). Digital twin reference model development to prevent operators risk in process plants. *Sustainability, 12*(3), 1088.

Bortolini, M., Ferrari, E., Gamberi, M., Pilati, F., & Faccio, M. (2017). Assembly system design in the Industry 4.0 era: A general framework. *IFAC-PapersOnLine, 50*(1), 5700–5705.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature, 582*(7810), 84–88.

Box, G.E., Hunter, J.S., Hunter, W.G., Bins, R., Kirlin IV, K., Carroll, D., (2005) Statistics for experimenters: design, innovation, and discovery, vol 2. Wiley-Interscience

Broy, M., Cengarle, M. V., Geisberger, E., & (2012). Cyber-physical systems: imminent challenges. In *Monterey Workshop* (pp. 1–28). Springer

Cadavid, J. P. U., Lamouri, S., Grabot, B., Pellerin, R., & Fortin, A. (2020). Machine learning applied in production planning and control: A state-of-the-art in the era of Industry 4.0. *Journal of Intelligent Manufacturing, 31*, 1531–1558.

Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal, 14*(2), 1–10.

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 832.

Chandrasegaran, S. K., Ramani, K., Sriram, R. D., Horváth, I., Bernard, A., Harik, R. F., & Gao, W. (2013). The evolution, challenges, and future of knowledge representation in product design systems. *Computer-aided Design, 45*(2), 204–228.

Chen, X., & Jin, R. (2017). Statistical modeling for visualization evaluation through data fusion. *Applied Ergonomics, 65*, 551–561.

Chen, X., & Jin, R.,(2018) Data fusion pipelines for autonomous smart manufacturing. In: 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE), IEEE, pp 1203–1208, https://doi.org/10.1109/COASE.2018.8560567

Chen, X., & Jin, R. (2021). Adapipe: A recommender system for adaptive computation pipelines in cyber-manufacturing computation services. *IEEE Transactions on Industrial Informatics, 17*(9), 6221–6229. https://doi.org/10.1109/TII.2020.3035524

Chen, X., Wang, L., Wang, C., Jin, R., (2018) Predictive offloading in mobile-fog-cloud enabled cyber-manufacturing systems. In *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, (pp. 167–172), https://doi.org/10.1109/ICPHYS.2018.8387654

Chen, X., Lau, N., & Jin, R. (2021). PRIME: A personalized recommender system for information visualization methods via extended matrix completion. *ACM Transactions on Interactive Intelligent Systems, 11*(1), 1–30.

Cheng, J., Chen, W., Tao, F., & Lin, C. L. (2018). Industrial IoT in 5G environment towards smart manufacturing. *Journal of Industrial Information Integration, 10*, 10–19.

Cisco (2019) Leading tools manufacturer transforms operations with iot. https://www.cisco.com/c/dam/en_us/solutions/industries/docs/manufacturing/c36-732293-00-stanley-cs.pdf, Accessed: 2021-07-17

Coatanéa, E., Tsarkov, V., Modi, S., Wu, D., Wang, G.G., Jafarian, H., (2018) Knowledge-based artificial neural network (KB-ANN) in engineering: Associating functional architecture modeling, dimensional analysis and causal graphs to produce optimized topologies for KB-ANNs. In: ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, p V01BT02A020, https://doi.org/10.1115/DETC2018-85895, https://doi.org/10.1115/DETC2018-85895

Dagli, C.H., (2012) Artificial neural networks for intelligent manufacturing. Springer Science & Business Media

Dai, H. N., Wang, H., Xu, G., Wan, J., & Imran, M. (2020). Big data analytics for manufacturing internet of things: Opportunities, challenges and enabling technologies. *Enterprise Information Systems, 14*(9–10), 1279–1303.

Dalla Valle, L., & Kenett, R. (2018). Social media big data integration: A new approach based on calibration. *Expert Systems with Applications, 111*, 76–90.

Davis, J., Edgar, T., Porter, J., Bernaden, J., & Sarli, M. (2012). Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Computers and Chemical Engineering, 47*, 145–156.

Dbouk, T. (2017). A review about the engineering design of optimal heat transfer systems using topology optimization. *Applied Thermal Engineering, 112*, 841–854.

Deng, X., & Jin, R. (2015). QQ models: Joint modeling for quantitative and qualitative quality responses in manufacturing systems. *Technometrics, 57*(3), 320–331.

Derigent, W., Cardin, O., & Trentesaux, D. (2020). Industry 4.0: Contributions of holonic manufacturing control architectures and future challenges. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-020-01532-x

Dharmesti, M. D. D., & Nugroho, S. S. (2013). The antecedents of online customer satisfaction and customer loyalty. *Journal of Business and Retail Management Research, 7*(2)

Dilberoglu, U. M., Gharehpapagh, B., Yaman, U., & Dolen, M. (2017). The role of additive manufacturing in the era of Industry 4.0. *Procedia Manufacturing, 11*, 545–554.

Duffie, N. A., & Malmberg, S. (1987). Error diagnosis and compensation using kinematic models and position error data. *CIRP Annals, 36*(1), 355–358.

Edgar, T. F., & Pistikopoulos, E. N. (2018). Smart manufacturing and energy systems. *Computers and Chemical Engineering, 114*, 130–144.

Feng, W., Wang, C., & Shen, Z. J. M. (2017). Process flexibility design in heterogeneous and unbalanced networks: A stochastic programming approach. *IISE Transactions, 49*(8), 781–799.

Fisher, R. A. (1919). XV.–The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh, 52*(2), 399–433.

Fragapane, G., Ivanov, D., Peron, M., Sgarbossa, F., & Strandhagen, J. O. (2020). Increasing flexibility and productivity in Industry 4.0 production networks with autonomous mobile robots and smart intralogistics. *Annals of Operations Research.* https://doi.org/10.1007/s10479-020-03526-7

Gorecky, D., Schmitt, M., Loskyll, M,. Zühlke, D.,(2014) Human-machine-interaction in the Industry 4.0 era. In *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, (pp. 289–294), https://doi.org/10.1109/INDIN.2014.6945523

Hartmann, B., King, W.P., Narayanan, S., (2015) Digital manufacturing: The revolution will be virtualized. https://www.mckinsey.com/business-functions/operations/our-insights/digital-manufacturing-the-revolution-will-be-virtualized, Accessed: 2021-07-17

Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association, 103*(482), 570–583.

Higdon, D., Gattiker, J., Lawrence, E., Jackson, C., Tobis, M., Pratola, M., Habib, S., Heitmann, K., & Price, S. (2013). Computer model calibration using the ensemble kalman filter. *Technometrics, 55*(4), 488–500.

Huysamen, K., Bosch, T., de Looze, M., Stadler, K. S., Graf, E., & O'Sullivan, L. W. (2018). Evaluation of a passive exoskeleton for static upper limb activities. *Applied Ergonomics, 70*, 148–155.

Ismail, A., Truong, H. L., & Kastner, W. (2019). Manufacturing process data analysis pipelines: A requirements analysis and survey. *Journal of Big Data, 6*(1), 1–26.

Ivanov, D., Dolgui ,A. (2020) Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. a position paper motivated by COVID-19 outbreak. International Journal of Production Research 58(10):2904–2915

Jaloudi, S. (2019). Communication protocols of an industrial internet of things environment: A comparative study. *Future Internet, 11*(3), 66–83.

Jared, B. H., Aguilo, M. A., Beghini, L. L., Boyce, B. L., Clark, B. W., Cook, A., Kaehr, B. J., & Robbins, J. (2017). Additive manufacturing: Toward holistic design. *Scripta Materialia, 135*, 141–147.

Jeschke, S., Brecher, C., Meisen, T., Özdemir, D., & Eschert, T. (2017). Industrial Internet of Things and Cyber Manufacturing Systems. *Industrial internet of things* (pp. 3–19). Springer.

Jin, R., & Deng, X. (2015). Ensemble modeling for data fusion in manufacturing process scale-up. *IIE Transactions, 47*(3), 203–214.

Jin, R., Deng, X., Chen, X., Zhu, L., & Zhang, J. (2019). Dynamic quality-process model in consideration of equipment degradation. *Journal of Quality Technology, 51*(3), 217–229.

Kang, S., Deng, X., & Jin, R. (2021). A cost-efficient data-driven approach to design space exploration for personalized geometric design in additive manufacturing. *Journal of Computing and Information Science in Engineering, 21*(6), 061008.

Kendrik, Y., Hong, L., Pai, Z., & Chun-Hsien, C. (2020). A state-of-the-art survey of digital twin: Techniques, engineering product lifecycle management and business innovation perspectives. *Journal of Intelligent Manufacturing, 31*, 1313–1337.

Kenett RS (2020) Reviewing of applied research with an Industry 4.0 perspective Available at SSRN 3591808. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3591808

Kenett, R.S., Rubinstein, A. (2017) Generalizing research findings for enhanced reproducibility: A translational medicine case study Available at SSRN 3035070. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3035070

Kenett, R.S., Shmueli, G., (2016) Information quality: The potential of data and analytics to generate knowledge. John Wiley & Sons

Kenett, R. S., Zacks, S., & Amberti, D. (2013). *Modern Industrial Statistics: with applications in R*. MINITAB and JMP: John Wiley & Sons.

Kenett, R. S., Zonnenshain, A., & Fortuna, G. (2018). A road map for applied data sciences supporting sustainability in advanced manufacturing: The information quality dimensions. *Procedia Manufacturing, 21*, 141–148.

Kenett, R. S., Swarz, R. S., & Zonnenshain, A. (2019). *Systems engineering in the fourth industrial revolution: Big data, novel technologies, and modern systems engineering*. John Wiley & Sons.

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(3), 425–464.

Kozjek, D., Kralj, D., Butala, P., et al. (2017). A data-driven holistic approach to fault prognostics in a cyclic manufacturing process. *Procedia CIRP, 63*, 664–669.

Kuo, C. J., Ting, K. C., Chen, Y. C., Yang, D. L., & Chen, H. M. (2017). Automatic machine status prediction in the era of Industry 4.0: Case study of machines in a spring factory. *Journal of Systems Architecture, 81*, 44–53.

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters, 3*, 18–23.

Li, J., Jin, R., & Hang, Z. Y. (2018). Integration of physically-based and data-driven approaches for thermal field prediction in additive manufacturing. *Materials & Design, 139*, 473–485.

Li, Z., Wang, K., He, Y., (2016) Industry 4.0-potentials for predictive maintenance. In *International Workshop of Advanced Manufacturing and Automation (IWAMA 2016)*. (pp. 42–46), Atlantis Press.

Loayza, N., Pennings, S.M., (2020) Macroeconomic policy in the time of covid-19: A primer for developing countries. World Bank Research and Policy Briefs No 147291

Luo, L., Kannan, P., Besharati, B., & Azarm, S. (2005). Design of robust new products under variability: Marketing meets design. *Journal of Product Innovation Management, 22*(2), 177–192.

Luthra, S., & Mangla, S. K. (2018). Evaluating challenges to Industry 4.0 initiatives for supply chain sustainability in emerging economies. *Process Safety and Environmental Protection, 117*, 168–179.

Mahmoudi, M., Tapia, G., Karayagiz, K., Franco, B., Ma, J., Arroyave, R., Karaman, I., & Elwany, A. (2018). Multivariate calibration and experimental validation of a 3D finite element thermal model for laser powder bed fusion metal additive manufacturing. *Integrating Materials and Manufacturing Innovation, 7*(3), 116–135.

Malhi, A., Yan, R., & Gao, R. X. (2011). Prognosis of defect propagation based on recurrent neural networks. *IEEE Transactions on Instrumentation and Measurement, 60*(3), 703–711.

de Man, J. C., & Strandhagen, J. O. (2017). An Industry 4.0 research agenda for sustainable business models. *Procedia CIRP, 63*, 721–726.

Meguid, S., Shagal, G., Stranart, J., & Daly, J. (1999). Three-dimensional dynamic finite element analysis of shot-peening induced residual stresses. *Finite Elements in Analysis and Design, 31*(3), 179–191.

Mehami, J., Nawi, M., & Zhong, R. Y. (2018). Smart automated guided vehicles for manufacturing in the context of Industry 4.0. *Procedia Manufacturing, 26*, 1077–1086.

Modoni, G. E., Caldarola, E. G., Sacco, M., & Terkaj, W. (2019). Synchronizing physical and digital factory: Benefits and technical challenges. *Procedia CIRP, 79*, 472–477.

Monostori, L., & Prohaszka, J. (1993). A step towards intelligent manufacturing: Modelling and monitoring of manufacturing processes through artificial neural networks. *CIRP Annals, 42*(1), 485–488.

Nardo, M. D., Forino, D., & Murino, T. (2020). The evolution of man-machine interaction: The role of human in industry 4.0 paradigm. *Production and Manufacturing Research, 8*(1), 20–34.

Oakley, J., & O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika, 89*(4), 769–784.

Oliff, H., & Liu, Y. (2017). Towards Industry 4.0 utilizing data-mining techniques: A case study on quality improvement. *Procedia CIRP, 63*, 167–172.

O'Donavan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. (2015). An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. *Journal of Big Data, 2*(1), 25–50.

Paelke V (2014) Augmented reality in the smart factory: Supporting workers in an Industry 4.0. environment. In *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)* (pp. 1–4), https://doi.org/10.1109/ETFA.2014.7005252

Pagani, L., Grasso, M., Scott, P. J., & Colosimo, B. M. (2020). Automated layerwise detection of geometrical distortions in laser powder bed fusion. *Additive Manufacturing, 36*, 101435.

Park, J. K., Kwon, B. K., Park, J. H., & Kang, D. J. (2016). Machine learning-based imaging system for surface defect inspection. *International Journal of Precision Engineering and Manufacturing-Green Technology, 3*(3), 303–310.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S. (2019) Pytorch: An imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R (eds) Advances in Neural Information Processing Systems 32, Curran Associates, Inc., pp 8024–8035

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, Édouard. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research, 12*(85), 2825–2830.

Prabhu, V. V., & Duffie, N. A. (1995). Modelling and analysis of nonlinear dynamics in autonomous heterarchical manufacturing systems control. *CIRP Annals, 44*(1), 425–428.

Profanter, S., Tekat, A., Dorofeev, K., Rickert, M., Knoll, A., (2019) Opc ua versus ros, dds, and mqtt: performance evaluation of industry 4.0 protocols. In *2019 IEEE International Conference on Industrial Technology (ICIT)* (pp. 955–962). IEEE.

Qi, Q., & Tao, F. (2018). Digital twin and big data towards smart manufacturing and Industry 4.0: 360 degree comparison. *IEEE Access, 6*, 3585–3593.

Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access, 8*, 21980–22012.

Reinman, G., Ayer, T., Davan, T., Devore, M., Finley, S., Glanovsky, J., Gray, L., Hall, B., Jones, C., Learned, A., et al. (2012). Design for variation. *Quality Engineering, 24*(2), 317–345.

Reis, M. S., & Kenett, R. (2018). Assessing the value of information of data-centric activities in the chemical processing Industry 4.0. *AIChE Journal, 64*(11), 3868–3881.

Roberts, I. A., Wang, C., Esterlein, R., Stanford, M., & Mynors, D. (2009). A three-dimensional finite element analysis of the temperature field during laser melting of metal powders in additive

layer manufacturing. *International Journal of Machine Tools and Manufacture, 49*(12–13), 916–923.

Sabbaghi, A., Huang, Q., & Dasgupta, T. (2018). Bayesian model building from small samples of disparate data for capturing in-plane deviation in additive manufacturing. *Technometrics, 60*(4), 532–544.

Sall, J., Stephens, M.L., Lehman, A., Loring, S., (2017) JMP start statistics: a guide to statistics and data analysis using JMP. Sas Institute

Santner, T. J., Williams, B. J., Notz, W. I., & Williams, B. J. (2003). *The design and analysis of computer experiments,* (Vol. 1). Springer.

Schluse, M., Priggemeyer, M., Atorf, L., & Rossmann, J. (2018). Experimentable digital twins-streamlining simulation-based systems engineering for Industry 4.0. *IEEE Transactions on industrial informatics, 14*(4), 1722–1731.

Schulze, A., & Störmer, T. (2012). Lean product development-enabling management factors for waste elimination. *International Journal of Technology Management, 57*(1/2/3), 71–91.

Schütze, A., Helwig, N., & Schneider, T. (2018). Sensors 4.0-smart sensors and measurement technology enable Industry 4.0. *Journal of Sensors and Sensor systems, 7*(1), 359–371.

Shi, J. (2006). *Stream of variation modeling and analysis for multistage manufacturing processes*. CRC Press.

Siemans (2019) Smart manufacturing in the u.s. https://www.siemens.com/innovation/en/home/pictures-of-the-future/industry-andautomation/digital-factory-smart-manufacturing-in-the-us.html, accessed: 2021-07-17

Singh, S., Shehab, E., Higgins, N., Fowler, K., Tomiyama, T., & Fowler, C. (2018). *Challenges of digital twin in high value manufacturing*. SAE Technical Paper: Tech. rep.

Sparks, E.R., Venkataraman, S., Kaftan, T., Franklin, M.J., Recht, B. (2017) KeystoneML: Optimizing pipelines for large-scale advanced analytics. In *2017 IEEE 33rd international conference on data engineering (ICDE)* (pp. 535–546), IEEE. https://doi.org/10.1109/ICDE.2017.109

Stojanovic, L., Dinic, M., Stojanovic, N., Stojadinovic, A., (2016) Big-data-driven anomaly detection in industry (4.0): An approach and a case study. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 1647–1652), IEEE. https://doi.org/10.1109/BigData.2016.7840777

Strandhagen, J. W., Alfnes, E., Strandhagen, J. O., & Vallandingham, L. R. (2017). The fit of Industry 4.0 applications in manufacturing logistics: A multiple case study. *Advances in Manufacturing, 5*(4), 344–358.

Sun, H., Huang, S., & Jin, R. (2017). Functional graphical models for manufacturing process modeling. *IEEE Transactions on Automation Science and Engineering, 14*(4), 1612–1621.

Tao, F., Qi, Q., Wang, L., & Nee, A. (2019). Digital twins and cyber-physical systems toward smart manufacturing and Industry 4.0: correlation and comparison. *Engineering, 5*(4), 653–661.

Thoben, K. D., Wiesner, S., & Wuest, T. (2017). Industrie 4.0 and smart manufacturing-a review of research issues and application examples. *International Journal of Automation Technology, 11*(1), 4–16.

Van Luttervelt, C., Childs, T., Jawahir, I., Klocke, F., Venuvinod, P., Altintas, Y., Armarego, E., Dornfeld, D., Grabec, I., Leopold, J., et al. (1998). Present situation and future trends in modelling of machining operations progress report of the CIRP working group 'modelling of machining operations. *CIRP Annals, 47*(2), 587–626.

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018a). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems, 48*, 144–156.

Wang, J., Yang, J., Zhang, J., Wang, X., & Zhang, W. (2018b). Big data driven cycle time parallel prediction for production planning in wafer manufacturing. *Enterprise Information Systems, 12*(6), 714–732.

Wang, J., Xu, C., Zhang, J., Bao, J., & Zhong, R. (2020a). A collaborative architecture of the industrial internet platform for manufacturing systems. *Robotics and Computer-Integrated Manufacturing, 61*, 101854.

Wang, L., Zhang, Y., Chen, X., & Jin, R. (2020b) Online computation performance analysis for distributed machine learning pipelines in fog manufacturing. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)* (pp. 1628–1633), IEEE. https://doi.org/10.1109/CASE48305.2020.9216979

Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering, 7*(4), 623–640.

Weimer, D., Scholz-Reiter, B., & Shpitalni, M. (2016). Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Annals, 65*(1), 417–420.

Weiss, B.A., Vogl, G., Helu, M., Qiao, G., Pellegrino, J., Justiniano, M., Raghunathan, A. (2015) Measurement science for prognostics and health management for smart manufacturing systems: key findings from a roadmapping workshop. In: Proceedings of the Annual Conference of the Prognostics and Health Management Society. Prognostics and Health Management Society. Conference, NIH public Access, vol 6, pp 46–63

Wen, Y., Yue, X., Hunt, J. H., & Shi, J. (2018). Feasibility analysis of composite fuselage shape control via finite element analysis. *Journal of Manufacturing Systems, 46*, 272–281.

Wong, R. K., Storlie, C. B., & Lee, T. C. (2017). A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79*(2), 635–648.

Wu, C. J., & Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization*. John Wiley & Sons.

Xiong, G., Zhu, F., Liu, X., Dong, X., Huang, W., Chen, S., & Zhao, K. (2015). Cyber-physical-social system in intelligent transportation. *IEEE/CAA Journal of Automatica Sinica, 2*(3), 320–333.

Yan, H., Paynabar, K., & Shi, J. (2020). AKM^{2}D: An adaptive framework for online sensing and anomaly quantification. *IISE Transactions, 52*(9), 1032–1046.

Yang, H., & Ni, J. (2005). Dynamic neural network modeling for nonlinear, nonstationary machine tool thermally induced error. *International Journal of Machine Tools and Manufacture, 45*(4–5), 455–465.

Yang, J., Yuan, J., & Ni, J. (1999). Thermal error mode analysis and robust modeling for error compensation on a CNC turning center. *International Journal of Machine Tools and Manufacture, 39*(9), 1367–1381.

Yang, Z., Eddy, D., Krishnamurty, S., Grosse, I., Denno, P., Lu, Y., & Witherell, P. (2017) Investigating grey-box modeling for predictive analytics in smart manufacturing. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, p V02BT03A024

Zhang, Y., Wang, L., Chen, X., & Jin, R. (2019) Fog computing for distributed family learning in cyber-manufacturing modeling. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)* (pp. 88–93), https://doi.org/10.1109/ICPHYS.2019.8780264

Zhao, H., Jin, R., Wu, S., & Shi, J. (2011). PDE-constrained gaussian process model on material removal rate of wire saw slicing process. *Journal of Manufacturing Science and Engineering, 133*(2), 021012.

Zhao, R., Wang, D., Yan, R., Mao, K., Shen, F., & Wang, J. (2018). Machine health monitoring using local feature-based gated recurrent unit networks. *IEEE Transactions on Industrial Electronics, 65*(2), 1539–1548.