A state-of-the-art on production planning in Industry 4.0

Dan Luo, Simon Thevenin, Alexandre Dolgui

To cite this version:


HAL Id: hal-03793675
https://hal.science/hal-03793675
Submitted on 1 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A state-of-the-art on production planning in Industry 4.0

Dan Luo, Simon Thevenin, and Alexandre Dolgui

IMT Atlantique, LS2N-CNRS, La Chantrerie, 4 Rue Alfred Kastler, B.P. 20722, 44307 Nantes, France

ABSTRACT

The Industry 4.0 revolution is changing the manufacturing landscape. A broad set of new technologies emerged (including software and connected equipment) that digitize manufacturing systems. These technologies bring new vitality and opportunities to the manufacturing industry, but they also bring new challenges. This paper focuses on the impact of Industry 4.0 on production planning approaches and software. We first propose a digital twin framework that integrates production planning systems and frontier technologies. The frontier technologies that may impact production planning software are the internet of things, cloud manufacturing, blockchain, and big data analytics. Second, we provide a state-of-the-art on the application of each technology in the production planning, as well as a detailed analysis of the benefit and application status. Finally, this paper discusses the future research and application directions in the production planning. We conclude that Industry 4.0 will lead to the construction of data-driven models for production planning software. These tools will include models built accurately from data, account for uncertainty, and partially actuate the decision autonomously.

KEYWORDS

Production planning; Industry 4.0; Digital twin; Cloud manufacturing; Blockchain; Big data analytics

1. Introduction

With the Industry 4.0 revolution, the manufacturing shop floors are digitizing at a high pace, with more IoT (internet of things) devices, software, and interconnection with the external environment (suppliers, customers). The technologies of Industry 4.0 develop rapidly, and they include the digital twin (DT)/cyber-physical systems (CPS), internet of things (IoT), big data analytics (BDA)/artificial intelligence (AI), cloud manufacturing (CMg), and blockchain (BC) (Ivanov and Dolgui 2020; Ivanov, Sokolov, and Dolgui 2020). This new manufacturing landscape calls for a change in the production planning tools. To realize its full potential, production planning software must take advantage of the massive amount of data generated on the shop floor, integrate easily, take advantage of new technologies fostered by Industry 4.0, and adjust automatically to the constant changes on the shop floor.

The resulting tools will have a strong impact on the manufacturing industry. Despite the short return on investment of prescriptive analytic tools, most manufacturers
are not using these tools due to the high initial investment or the lack of knowledge. According to a recent survey BARC (2016), 74% of companies still use Excel for production planning, and 33% rely solely on Excel to plan their production. The implementation of prescriptive analytic tools requires high consulting costs to adapt software. Big data analytics not only allows us to forecast the value of unknown parameters accurately, but it also allows us to incorporate uncertainties of these forecasts in the models. Adaptive stochastic/robust optimization provides decisions (production planning) that are not only robust to various uncertainties but select the states (resource utilization, inventory level) to react appropriately when unknown parameters unfold. In addition, machine learning tools can help automatically learn the production capacity from the data or simulation. Automated planning model creation from data will reduce the costs of the production planning software since the software will automatically adjust to the requirements of the shop floor. As a result, prescriptive analytics will be widely used in manufacturing systems. The resulting tools will lead to production plans with the right level of agility, which is crucial in the current production context with high complexity, high flexibility, mass customization, dynamic decisions, and volatile markets.

This paper focuses on production planning in Industry 4.0. We identify the challenges related with research and application of Industry 4.0 keywords, including internet of internet, cloud manufacturing, blockchain, big data analysis, machine learning, digital twin, cyber-physical system. The main challenges for the application of frontier technologies in production planning are listed as follows:

(1) The integration of data, software, and decisions remains a complex challenge. This integration concerns the relations within the physical systems, the relations within virtual systems, and the relations between physical and virtual systems.

(2) Massive data open both new possibilities and difficulties for developing an effective production plan using cutting-edge technologies.

(3) These cutting-edge technologies may give managers dynamic and automatic supports of production planning. The challenge is to develop tools that can react in real-time and interact properly with the shop floor managers and the workers.

A framework is proposed in this study for an intelligent digital production planning twin. Such a digital twin integrates the current trends in production planning: the use of IoT data, big data analytics, cloud manufacturing, advanced decision aid techniques based on stochastic and robust optimization, and hybrid simulation-optimization planning approach.

For each of these research trends, we provide a state-of-the-art. Note that we are not attempting to give an exhaustive bibliography based on a systematic review. Instead, we select the papers for their quality and their relevance, considering the following key dimensions: journal quality, number of citations, innovation, practical applications, and reference. Finally, we give a new vision of the intelligent digital twin for production planning that integrates all Industry 4.0 technologies to facilitate production planning decision in manufacturing.

The present paper differs strongly from existing reviews. Bueno, Godinho Filho, and Frank (2020) provide a systematic review on the use of Industry 4.0 keywords in the production planning and control (PPC) papers. The authors show that most of work focuses on scheduling. On the contrary, this paper provides a vision of the future trends of production planning in the Industry 4.0 context, and we explain the benefits in this context of research that do not explicitly mention Industry 4.0 (e.g., papers on
simulations, stochastic optimization, ...). Cadavid et al. (2020) provide a systematic review on machine learning for PPC. Our review paper deals with a broader spectrum of Industry 4.0 technologies and methods. Ivanov and Dolgui (2020); Lu et al. (2020); Rossit, Tohme, and Frutos (2019) propose frameworks or architectures of the supply chain or manufacturing system in the context of Industry 4.0. Zhang, Zhang, and Yan (2019); Agostino et al. (2020) provide DT frameworks that focus on scheduling. However, the authors do not give clear information on the use of the emerging technologies of Industry 4.0 to support production planning decisions. Besides, Kasten (2020); Leng et al. (2020); Fosso Wamba et al. (2020); Li et al. (2021) present systematic reviews on blockchain for supply chain or manufacturing industry, but they do not discuss production planning issues in detail. Moreover, most of the existing literature review papers have focused on presenting what technologies are available for implementing Industry 4.0 rather than how Industry 4.0 factories make their decisions and manage operations. This paper fills in this gap.

Section 2 introduces classical functions of production planning, and Section 3 presents the main concepts of the intelligent digital twin for production planning. We then provide a stat of the art on the key elements of intelligent digital twins for production planning; IoT, cloud manufacturing, and blockchain (Section 4); big data analytics (Section 5); simulation and optimization (Section 6).

2. Definition, structure, and research scope for the production planning

2.1. Production planning and control (PPC) system

Production planning and control (PPC) systems help companies match manufacturing performance with customer demands (Bonney 2000). PPC is a value-adding process (Wiendahl, Von Cieminski, and Wiendahl 2005) that encompasses all tasks related to the management of the value creation processes in a company (Bendul and Blunck 2005). PPC is a function determining the global production quantities (production plan) for a given planning horizon to satisfy the commercial plan and to meet the profitability, productivity, and delivery time objectives (Lolli et al. 2019). PPC also includes the control of the manufacturing process for real-time resource synchronization and product customization. (Moeuf et al. 2018). Scholars often use hierarchical frameworks to describe the process of PPC at different levels and planning horizons (Oluyisola, Sgarbossa, and Strandhagen 2020). Although the details and terms for the framework of PPC systems are different in different studies, the core content remains the same. Existing research often describes the PPC framework at the long-term, medium-term and short-term (Bonney 2000; Oluyisola, Sgarbossa, and Strandhagen 2020; Garetti and Taisch 1999; Jacobs et al. 2011). Figure 1 depicts such PPC frameworks.

The decision process in PPC includes multiple sub-processes (production planning, capacity planning, and rough-cut capacity planning, etc). This decomposition was defined even before computers allowing humans to plan by hand. The first software for PPC, e.g., MRP, followed this historical decomposition, and they provide a set of functionality, where each functionality corresponds to one of these sub-processes. As this decomposition is sub-optimal and inconvenient, the literature suggests integrating these decisions (e.g., sales and operations planning), and the software followed (e.g., enterprise resource planning (ERP) fosters the integration of procurement, production, and capacity planning). With the increase of computation power and the development
of optimization approaches, decision support tools for production planning tend to integrate all the decisions and data at a given planning level.

At the strategic level, manufacturing operations are viewed in a long-term, aggregated manner. (Oluyisola, Sgarbossa, and Strandhagen 2020). Strategic decisions begin with sales and operations planning (S&OP) or aggregate planning. The tactical level considers the medium-term planning, which is called materials resource planning (MRP). The operational level concerns day-by-day, shift-by-shift detailed scheduling, and real-time control. The focus of this study is on long-term and medium-term production planning, and we do not discuss scheduling and real-time control.

2.2. Aggregate production planning

S&OP aims to balance the overall demand with the available resources. This process is dedicated to unifying plans traditionally produced independently by different departments related to production, distribution, procurement, and sales (Pereira, Oliveira, and Carravilla 2020). S&OP is performed monthly, at an aggregated level (based on product family), and for a planning horizon of up to a few years since S&OP decisions (buying new machine, hiring workers) must be taken long before implementation (Noroozi and Wikner 2017). The input of S&OP includes demand data (volumes per product family per planning period), metadata (such as forecast uncertainty) from demand management (DM), as well as future available aggregate capacity from resource planning (RP) (Oluyisola, Sgarbossa, and Strandhagen 2020; Jacobs et al. 2011).

The S&OP process gathers people from different functional areas, to balance the demand and the capacity plans. S&OP might lead to jointly deciding pricing with the production plan. S&OP is sometimes classified as a strategic process since it might lead to capacity extension, but most of the literature considers it a tactical process.

2.3. Master production scheduling

While S&OP considers product families, the master production schedule (MPS) generates the production target for each end-item by period typically monthly. In recent planning systems, MPS integrates rough-cut capacity planning (RCCP) (Rossi et al. 2017), where planners check that the capacity of critical resources (bottleneck, labor, critical materials) is sufficient to meet the production target. If this is not the case, the
2.4. **Materials requirements planning**

MRP combines the MPS records with the bill of materials (BOM) data and inventory data to obtain the requirements of components and parts. Using the results of MPS as the input, MRP makes recommendations on the release replenishment orders for materials. Based on the production capabilities and lead times which dictate the capacity requirements planning (CRP), MRP releases, typically weekly, detailed material replenishment and capacity plans for a planning horizon of a few months (Oluyisola, Sgarbossa, and Strandhagen 2020). These plans are often updated, and the output of MRP are the input for the operational level (Dolgui and Prodhon 2007).

3. **Intelligent digital twin for the production planning and structure of the state-of-the-art**

In recent years, the growing requirement for customized products and the extension of supply chains to all the globe led to various uncertainties in the supply chain, like delays in deliveries and unpredictable demands. Therefore, the supply chain is characterized by high complexity, high flexibility, mass customization, dynamic conditions, and volatile markets (Bonney 2000). In Industry 4.0, the fast changes in the industrial environment motivate the evolution and integration of supply chain management (Bueno, Godinho Filho, and Frank 2020). Industry 4.0 leads to a fast digitalization of the shop floors, and this provides new perspectives for production planning methods and software. Figure 2 shows the key elements and their relationships. As manufacturing digital twin integrates most digital advances fostered by Industry 4.0, we explain below the concept of a DT for production planning, and this concept guides the rest of this literature review.
3.1. **Definition and characteristic of the DT in the production planning**

Shafto et al. (2010) published one of the first public definitions of a DT in 2010. While the essence of digital twins is simulation models, DT is very different from the traditional simulation model. The DT is multi-physics, multi-scale, data-driven, and ultra-fidel. DT reflects the state of a corresponding twin in a timely manner based on the historical data, real-time sensor data, and physical models (Glaessgen and Stargel 2012). With the development of Industry 4.0, the concept of DTs has been expanded. Nowadays, the DT includes not only the simulation model but also the mathematical and data models.

There are many frameworks for the DT and CPS, but they share the same core elements shown in Figure 3. A classical digital twin requires 5 elements: (1) a physical object, (2) a virtual model, (3) data, (4) data connections, (5) services provided to the end-users. In addition, a digital twin usually provides the following characteristic:

- The data are collected from the physical object, and send to the model automatically.
- The computer model stay in synchronisation with a physical object. That is, any change in the physical object must be passed on to the virtual models.
- The model is able to pass instruction to the physical object.
- The model accounts for uncertainties. On the one hand, the model must account for uncertainties in the environment of the physical object since it include some parameters that can never be forecasted perfectly. On the other hand, any model merely a rough approximation of the complex real world. The model should be robust enough to provide valid decisions despite these approximations.

This definition is broad enough to encompass any physical object, and any type of virtual model (simulation, mathematical models, data model, ...). This generality explains the rising interest among researchers and industrial on the topic. As a broad concept able to gather all technologies used in computer science for manufacturing, and with the impulsion of Industry 4.0 revolution, the DT is becoming a core concept of the Industry 4.0 revolution. This will likely lead to the design of DTs for production planning (Luo, Thevenin, and Dolgui 2021). In fact, there is a growing literature on all components of such a digital twin.
3.2. Framework and key technologies of the DT in the production planning

Figure 4 presents the concept and framework of DT for production planning. The process of implementing digital twins in real production systems requires the collaboration of multiple technologies and tools. We cluster these key technologies into four categories, comprising intelligent perception, modeling and simulation, data management, and actuation.

(1) **Intelligent perception** aims to collect accurate input from the real world, which is the key for building a high-fidelity model. Intelligent perception mainly involves measurement technology. In a manufacturing environment, the data are collected from IoT devices, various software, that constitutes a cyber-physical system (CPS). In a communication network, the CPS is a group of embedded systems that communicate and interact with each other (Geisberger and Broy 2012). The CPS is the primary data and information source of the DT. The CPS is referred to as the cyber-physical production system (CPPS) in the context of production technologies (Weyrich et al. 2017). The information and data collected from CPPS can be used to build DTs for production planning. Generally, the CPPS collects hardware, software, and real-time information (Biesinger et al. 2019). Section 4 provides a state-of-the-art on the impact of IoT data in the production planning, and it reviews the integration between various data sources.

(2) **Data management**: The data of digital twins is massive, multi-time scale, multi-dimensional, multi-source, and heterogeneous. Therefore, data management is essential for the implementation of DTs. The domain model serves as a link between the physical and virtual systems. This domain model combines data from a variety of sources such as MES (manufacturing execution system),
ERP, and IoT devices. It also gives a rich data structure for the user to interpret. New paradigms emerge in the framework of Industry 4.0 for collecting and storing huge volumes of data in real time and across productive and logistical activities, enabling the development of the DT concept and associated techniques (Agostino et al. 2020). Following the digital twin perspective, the digital model must be as accurate as possible, and the progress in big data analytics (BDA) helps to provide a good prediction of the planning parameters. Section 5 provides a state-of-the-art on BDA methods in the production planning.

(3) **Modeling and simulation**: Mathematical and simulation models are the most used quantitative approaches for decision-making in the production planning. These models convert physical entities into virtual objects that can be analyzed with computers. Mathematical models provide a systematic way of expression for further analysis and optimization. The correctness and accuracy of these models directly affect production planning. The simulation models help the user validate a production plan by providing a precise execution of the plan at a detailed level (with each machine, employee, transport between machines, etc.). The simulation gives a clear understanding of the performance of a production plan since it can compute various KPIs relevant to the user. The simulation is also a valuable tool to enrich optimization models. Section 6 provide the state-of-the-art on data-based simulation in the production planning, simulation-optimization approach, and optimization under uncertainties.

(4) **Actuation**: An important aspect of the DT is the ability of the model to act on the physical object. Digital twins that do not provide this feature are sometimes called digital shadows. An automated actuation of some production planning decisions (e.g., number of workers to hire) is not possible. Nevertheless, the cloud manufacturing research trends provides a paradigm that allows manufacturers to share their production capacity in real time on the cloud. Besides, the blockchain leads to smart contract that enable real-time acceptance and tracking of production order. Section 4 provides a state-of-the-art on CMg and BC in the production planning.

(5) **Interconnection**: The main purpose of interconnection is to obtain effective and accurate data in the real physical world. The sharing of information and data can include interaction between different information systems, virtual system and physical system, and man-machine interface. Interconnection is an essential element for production planning, because the production plan involves data and information of the entire production system (supply chain). Moreover, only when the system interaction is efficient, the production plan can be implemented in real production.

In Appendix, Table A1 summarizes the key technologies and corresponding tools (Qi et al. 2019) for each category with pointers to the literature for the interested readers. Digital twins can support decision-making in every stage or at each level of production planning systems. For aggregate planning, DTs can achieve multi-level data sharing, traceable data flows, as well as the integration with demand forecast, inventory control, MES, and ERP system (Yu et al. 2018). DTs provide capabilities in real-time and dynamic production planning, with distributed and collaborative decision-making through MES, MPS/ERP, and CPS integration (Rossit, Tohme, and Frutos 2019). The DT model and CPS assist MRP in the automatic forecast, optimization, and re-planning (Lin, Wong, and Ge 2018), as well as expand MRP with real-time calculations, early reporting, traceability, and visibility (Shao and Helu 2017).
In the initial stage of the research about DTs, researchers mainly proposed digital twin frameworks for the entire supply chain management issues. With the deepening of research, researchers began to focus on more precise realisations dedicated to PPC systems. However, as scheduling is more sensitive to real-time data, most works on digital twins for PPC concern scheduling problems, and few studies discuss mid-term and long-term planning. Furthermore, with the published digital twin frameworks, there are few quantitative analyses and case studies. In Appendix, Table A2 reports papers that propose DT frameworks, and provides the author’s viewpoint, core methods, and considered applications. Additionally, a series of remarkable studies that have emerged recently is the digital twin-enabled Graduation Intelligent Manufacturing System (GiMS) proposed by Guo et al. (2019). This series of studies not only proposes a detailed implementable digital twin framework, but also investigates how the planning and scheduling are executed under the framework (Lin et al. 2019; Guo et al. 2020a,b,c; Li and Huang 2021). Their research is very timely, intriguing, and worthy of further study.

4. Frontier technologies for the data collection and sharing in the production planning

Data are the source and foundation of the production planning, and the essential of any system in the digital twin framework. Therefore, we first introduce the data sources commonly used in the production planning, and then discuss the current status of application and research of frontier technologies for data collection and sharing in the production planning one by one, in the order of IoT, cloud manufacturing, and blockchain.

4.1. Relevant data sources for the production planning

The key of BDA technologies is massive data, which is used to gain autonomous computer knowledge (Sharp, Ak, and Hedberg Jr 2018). When it comes to training a machine learning (ML) model, the selection of the data source is crucial since the final results depend largely on the quality of the data. We introduce five data sources that are very important for intelligent production planning (Lu 2014; Tao et al. 2018b; Cadavid et al. 2020), and we explain the importance of this data for production planning.

(1) Management data are the historical data collected from enterprise information systems, including the ERP, MES, etc. Besides providing basic parameters for production planning, the management data also include the historical production plans and execution results of production plans. Analyzing these historical production plans and execution results provides knowledge to improve future plans and not repeat mistakes.

(2) Equipment data are collected from IoT devices. The equipment data helps to estimate the resource capacity in the production planning. The production resource includes machines, humans, space, and even containers.

(3) Consumer data are collected from e-commerce platforms or other social media about consumers, who are the users of products. These data can be used to train machine learning models, which can provide support for demand forecasts.

(4) User data are system feedback given by workers or experts, who are the user of production planning tools. This type of users data is usually obtained through
interviews or questionnaires. These data can be used to optimize system performance.

(5) **Product data** originated from products or services either during the production process or during their use by the final consumer. The production planning mainly concerns on the production data during the production process, including the BOM, process step, etc. This data help to estimate the production yield.

Management data are the most used data sources. Due to commercial reasons, the data of the enterprise is often confidential. Because it is difficult to access data coming from companies, researchers often use simulated data and public data to train the machine learning model. However, the result is often different from the real life situation.

IoT technologies motivate the BDA applications with equipment and product data (Correa et al. 2020; Hajjaji et al. 2021). Nevertheless, accessing IoT data in the PPC system remains a challenge. The use of DTs could tackle this challenge, by collecting IoT data scattered in various systems, and automatically cleaning and integrating the data. While various studies provided tools and methods to create the digital twin (Tao et al. 2018a; Zheng, Yang, and Cheng 2019; Lu et al. 2020), this still represents a research issue. Companies need to build general domain models to integrate interactive platforms, as well as to realize the data connection between the physical and virtual systems.

4.2. **Internet of things**

The IoT originated from the radio frequency identification devices (RFIDs) system proposed by MIT Auto-ID Labs in 1999 (Ashton et al. 2009). The international telecommunications union (ITU) defined IoT as the intelligent connectivity for anything at any time and anywhere (Atzori, Iera, and Morabito 2010). The internet of things (IoT) is the critical component of the CMIg, DT, and BDA (Hwang, Kim, and Rho 2016).

The core function of IoT is to acquire real-time data from the shop floor and its environment. With the IoT technology, a product can be equipped with a uniquely identifiable code. Through uniquely identifiable code, we can monitor and track this product throughout its entire life cycle by sensors and wireless sensor networks (Fang et al. 2016). The key technologies of IoT are RFID and wireless communication technologies. The RFID enables tracking and distinguishing every single product. The wireless communication technologies embedded in intelligent devices enable real-time access to data on the status of products. Finally, the IoT collects various data (e.g., the information of sound, light, heat, electricity, mechanics, chemistry, biology, and location) by global position systems (GPS), infrared sensors, laser scanners, gas sensors, and other devices (Tao et al. 2014a).

The IoT is exploited industrially at several different levels of production and logistics systems, such as the inventory management, assembly processes, and after-sales services (Fang et al. 2016). As shown in Figure 5, the IoT increases the accuracy and flexibility of production planning by providing up to data from physical systems (Bueno, Godinho Filho, and Frank 2020; Rauch, Dallasega, and Matt 2018). For instance, Tao et al. (2017); Zuo, Tao, and Nee (2018) find that RFID reduces inventory shrinkages due to damage and thieves. Typically, the data gathered by IoT devices help production planner to know the demand of customers, the inventory levels of materials, the capacity of the workshop, and the status of suppliers. With the accurate collection of data in real-time, IoT helps production planning become more automatic.
and intelligent. As a result, production planning can respond quickly to various events such as machine breakdown, and urgent incoming customer orders, a late material delivery.

In Industry 4.0, one important task of IoT is the integration of information systems, such as the ERP systems and MES, to enable information exchange and cooperation (Fang et al. 2016). Most IoT research focuses on the collection of real-time data and its use in scheduling (Zhang et al. 2018). However, little research concerns the application of IoT in the production planning (Wang et al. 2020; Zhong et al. 2016). In Appendix, Table A3 summarizes the IoT literature about production planning. Thus various production planning issues still need to be addressed. These issues include the integration of information systems while minimizing their complexity, the development of methods to take advantage of IoT in data-driven and dynamic planning, the development of tools for distributed and collaborative planning among different workshops.

4.3. Cloud manufacturing

CMg is a new paradigm that require real time actuation of production planning decisions. Cloud manufacturing relies on IoT, cloud computing, virtualization, service-oriented architectures, and advanced computing technologies (Wu et al. 2013a). CMg aims to package as services the production resources and capabilities of all manufacturers in the supply chain. The supply chain becomes a cloud of manufacturing services that provide on-demand, self-service, and agile commercial manufacturing resources. As a result, the production resources of an enterprise are shared (as manufacturing services) not only to major downstream distributors of the supply chain but also to provide customized manufacturing services for customers. Meanwhile, a manufacturing enterprise can outsource its resources to other manufacturers, and it can use the production resources of other enterprises for an efficient and low-cost production.

The cloud manufacturing creates a challenge in the production planning. On the one hand, in the CMg environment, enterprises can schedule and integrate various re-
sources within the enterprise to improve resource utilization and reduce costs. On the other hand, the service-oriented CMg paradigm makes production patterns and application scenarios more diversified and complex. Therefore, the difficulty of production planning under the CMg environment will increase sharply. Although, decision-makers can obtain more information about the whole supply chain to optimize production plans under the cloud manufacturing paradigm. However, how to integrate production resources in the supply chain, how to reduce production costs in all aspects, and how to increase the speed corresponding to customer needs to achieve agile manufacturing is still challenging for production planning and deserves scholars’ attention.

CMg application in the production planning includes the applications in the enterprise and among enterprises. Figure 6 illustrates the interconnection within an enterprise and the connection between enterprises. The application of CMg in the enterprise promotes the integration of the information related to production, product, and other business management information, as well as the integration of the IoT-based workshop and other enterprise information subsystems. The application of CMg among enterprises can address the information integration, storage, retrieval, analysis, use, data security, and other issues during these ubiquitous service management and application process among different enterprises. With the support of CMg, the production planning can obtain more valuable data from various sources to improve the practicality of plans. Finally, cloud computing facilitates the storage and interaction of massive data, and it can speed up the optimization of planning.

Many papers discuss the architecture and application of CMg from a macroscopic point of view (Ning et al. 2011; Hasan and Starly 2020). However, few studies have focused on PPC in CMg. Most research on PPC for cloud manufacturing focus on scheduling (Erol and Sihn 2017; Yu et al. 2018; Arunarani, Manjula, and Sugumaran 2019; Liu et al. 2019b), and few works consider production planning. This is surprising, because CMg requires careful management of production resources and capacities of service providers, and the possible subcontracting through CMg must be included in the planning tool of the manufacturers. Therefore, the production planning under the CMg paradigm deserves in-depth research by scholars. In Appendix, Table A4 presents the CMg literature about the production planning.

The application of CMg within a company started gradually (Yu et al. 2018; Wang et al. 2019b). However, the application of CMg among enterprises is difficult (especially in the entire supply chain) because of commercial confidentiality, data security, and access to heterogeneous data sources. To solve this problem, blockchain (BC), an emerging technology, has captured the interest of academics. We describe the definition and application of the blockchain in detail in the next section.

4.4. **Blockchain**

Blockchain (BC) is an emerging technology that protects security and privacy through a new type of safe and reliable peer-to-peer communication platform (Lakhani and Ian-siti 2017). While the academic community does not provide a uniform and strict definition for blockchain (Tan et al. 2021), the following definition is commonly accepted. The blockchain is a decentralized and collaborative database, where all members (or nodes) of a network can equally shares, verifies and maintains stored data (Li, Barenji, and Huang 2018). The BC has no centralized node, and thus no third parties (Zhu et al. 2019). This approach enhances the trust between nodes. The BC provides a stable and reliable way of data storage (Vatankhah Barenji et al. 2020), because the
Figure 6. Caption: CMg in the enterprise and among enterprises for the production planning.

Figure 6. Alt Text: This figure illustrates the interconnection within an enterprise and the connection between enterprises in CMg environment. The objects of their services, as well as the data source, are different.
data stored in BC can only be added, not deleted, or modified. Therefore, the BC permanently records all operations on data. This guarantees the traceability of history, and the cost of doing evil becomes very high.

BC is also a trading platform that executes "smart contracts" (Hofmann, Strewe, and Bosia 2018). Smart contracts, also called chain codes (Tsai et al. 2017), are digital agreements between nodes. Programs created in high-level programming languages form these smart contracts, which are stored and copied in the form of a BC. Smart contracts can be automatically executed once meeting specific conditions (Dolgui et al. 2020).

Although the current typical BC applications focus on the cryptocurrency. There is growing attention to the applications of BC in manufacturing contexts, such as in supply chain management, BDA, CPS, DT, IoT, CMg. In Appendix, Table A5 lists the relevant literature about the BC applied in Industry 4.0. The current research is more concerned with the architecture of the application, and there remain some limitations. Nowadays, supply chains are global, but the BC has not formed a unified international standard yet. Before extensive and large-scale application, a series of issues such as transaction mechanism, credit mechanism, compatibility, and connectivity still need to be resolved.

Blockchain makes it possible to have new service models for production planning. BC enables smart contracts that can automate the order acceptance process, as shown in Figure 7. Consumers submit their demands (product types, quantities, expected price, and various personalized needs). Suppliers submit their available resources (product inventory, production resources). If the customer demand matches the supplier resources, a digital signature generates the smart contract. BC also enables the continuous monitoring of the fulfillment of the smart contract. Suppliers can update idle resources in real-time, and customers may respond dynamically. This process involves flexible and reactive production plans, and this enhances the utilization efficiency of production resources. Moreover, through the credit and reward and punishment mechanism, the BC system can reward or punish suppliers or consumers based on the fulfillment of the smart contract.

To effectively deal with smart contracts, manufacturers need production planning software with an accurate description of the production capacity. In addition, the production plan must account for uncertain demand, and to modify the plan when firm orders arrive. The application research of BC in the manufacturing industry mainly focuses on the macro-system level (Yu et al. 2020; Vu, Ghadge, and Bourlakis 2021). Up to now, few researchers have considered the implications of BC in the production planning (Rahmanzadeh, Pishvaee, and Rasouli 2020; Herrgoß et al. 2020).

4.5. Limitation and future direction

IoT and cloud manufacturing enable the digitization of the entire supply chain and its environment. IoT can provide valuable data. CMg and blockchain can enhance collaborations between companies. While IoT, CMg, and blockchain can support production planning by providing and sharing valuable data, studies on the application of IoT, CMg, and blockchain in the production planning remain scarce. The major limitations and future directions on these topics are summarized as follows:

(1) IoT collects a large amount of data, and it interconnects the virtual and real systems. The resulting information systems are often large and complex, with heavy memory load and slow calculation. Reducing the complexity of the result-
(2) To get the full value of IoT data, more research should be conducted on event-based and data-driven planning. The goals are to improve the representation of the shop floor in the model thanks to data, to integrate the variability of data in the models, and to react to events efficiently without creating nervousness.

(3) The integration of information from different systems remains a challenge because it requires reconciling data from heterogeneous sources. Other difficulties include the use of different standards in information systems and data interaction. Solutions to overcome this integration issue include service-oriented architectures (Niknejad et al. 2020) and blockchain (Korpela, Hallikas, and Dahlberg 2017) for the flexibility and security of data transmission, as well as ontologies (Kumar et al. 2019) for the mapping of different data models. Nevertheless, future work is required to ease the integration of the information collected from IoT devices, software, and between information systems from different shop floors. This requires the development of the standard for data format, protocols for system interaction, and the data management procedure that ensures safety and reliability. Future works also include the development of tools to automatically clean the data, and to detect and fix incoherent information (e.g., the level of inventory in the ERP and computed from RFID).

(4) Collaborative planning (between different firms) reduces the delivery lead times uncertainty and leads to better production capacity usage. Nevertheless, the contradiction between sharing information and protection of privacy and core technology is a barrier to the adoption of collaborative planning. Blockchain and cloud manufacturing are enablers to distributed and collaborative planning. Blockchain technology may be one of the potential solutions for creating a secure communication protocol for collaborative planning in the cloud (Vatankhah Barenji et al. 2020; Li, Barenji, and Huang 2018).

(5) Research is required to foster the application of cloud computing in the production planning and speed up the calculation. In particular, researchers must focus on the development of parallel algorithms to solve large-scale lot-sizing problems.
5. Big data analytics applied in the production planning

With the development of the industrial internet, a variety of sensors have been installed in the plant to track the state of equipment and product quality (Sun et al. 2019). These sensors along with the growing number of software systems in the factory collect a massive amount of data from physical systems that support the decision-making for production planning. The recent development in big data analytics/artificial intelligence led to better consideration of uncertainties in the production planning (Bonney 2000). Based on the vast amount of data gathered by IoT, BDA/AI tools can forecast the distribution of input parameters required for production planning, such as demand, production yield, supply/product lead times, process duration, and production capacity (Gonzalez-Vidal, Jimenez, and Gomez-Skarmeta 2019; Lolli et al. 2019). The development of BDA/AI tools and the increasing amount of data leads to better accuracy of the forecast. As a forecast will never be correct, these tools allow computing the variability of the parameters to account for the uncertainty. As a result, production planning models represent more precisely the production process on the workshop. Accounting for uncertainty leads to plans that are more often implementable in practice. In addition, some approaches incorporate the dynamic of the decision process, where the plan can change over the time, and this leads to adaptable planning.

Even though manufacturing generates huge data sets, and despite the growing interest in BDA/AI in the production planning, the exploitation of big data in the production planning is still immature compared with other fields like IT, finance, and e-commerce (Lamba and Singh 2017). Figure 8 shows how to use the BDA method in the production planning. The application of big data analytics requires a combination of understanding and knowledge about the domain and the right BDA algorithms. Therefore, companies have collected massive data, but they cannot currently get the best value of this data. This section analyzes research on big data analytics applied in the production planning in the context of Industry 4.0.

The processes for implementing BDA/AI tools in the production planning include data collection and cleaning, predictive models, model training, validation, and testing (Cheng et al. 2018). We review below the literature on BDA/AI tools for demand forecast, before surveying the works on time estimation, as shown in Figure 8.
5.1. **Demand forecast with BDA/AI tools for the production planning**

For manufacturing organizations, demand forecasting is critical because it serves as a foundation for production planning. Demand forecasting is challenging, however, since consumer demands frequently shift due to a variety of factors, such as policy, economic trends, market competition (Kück and Freitag 2021). On the one hand, in the framework of the digital twin, companies can collect huge amounts of data for demand forecasting, which brings new and unlimited opportunities for profitability. On the other hand, the current state of application and research shows that demand forecasting errors are persistent and their results are frustrating and costly. Demand forecasting has long been stuck in a backwards-looking perspective. In fact, demand forecasting based on only a few years of order information makes little meaning for long-term production planning. The focus of demand forecasting should be to explain the changing context and factors that influence each crucial turning point, but this is difficult.

Compared with traditional methods, machine learning methods, such as artificial neural networks (ANN) (Kourentzes 2013; Kourentzes, Barrow, and Crone 2014), support-vector machines (SVM) (Lu 2014; Villegas, Pedregal, and Trapero 2018), Bayesian networks (BNs), random forest, have shown promising results in current studies, and have surpassed traditional methods in precision and performance. Although these ML approaches have exploded in popularity in recent years for time series forecasting in a variety of fields, including banking, power generation, and water resources (Dudek 2020; Salinas et al. 2020). But these forecasting approaches are still not widely used in the production planning, and the demand forecasting and production planning are still mostly based on the planner’s expertise (Lorente-Leyva and Alemany 2020). One of the main reasons behind this is that research in the area of big data applied to prediction is not mature. Apart from the immaturity of demand forecasting models, how to generate meaningful demand forecasts for production planning based on big data inputs is also a problem to be solved. Currently, demand forecasting and production planning are often studied in isolation, but the coupling between them is a key issue to be considered.

5.2. **Time estimation with BDA/AI tools for the production planning**

BDA-based time estimation is promising to adjust different time-related parameters to current working conditions. The time estimation includes the prediction of lead time, cycle time, production time, and even the yield (it is also related to the time). In Appendix, Table A6 summarizes the literature on BDA-based methods to predict time-related parameters in the production planning.

Only a few works consider lead time prediction in the research community (Cadavid et al. 2020). Lingritz et al. (2018), on the one hand, compare the performance of several ML methods for predicting lead times. They, on the other hand, do not examine high variance processes and do not require process mining-based information. Meidan et al. (2011) also evaluate several ML methods, however they only take into account the waiting time. In a flow-shop setting, Mori and Mahalec (2015) focus on product characteristics to anticipate lead times, but disregard the complexity of the processes involved. Öztürk, Kayalgiligil, and Özdemiür (2006) compare the accuracy of predicted lead times, and employ only simulated data as the input for their models. Alenezi, Moses, and Trafalis (2008) demonstrate that support vector machines outperform neural networks for order flow time predictions. However, the authors use
data from computer simulation rather than real-world data from the work floor. Wang et al. (2018) compute the probability distribution based on operating conditions, but the authors concentrates mostly on the difficulties of working with binarized variables. Schuh et al. (2019) provide a methodology as well as a research based on real-world data to illustrate how ML algorithms may be used to anticipate the transition time, which consists of post-process waiting time, transport time, and pre-process waiting time. Despite the fact that their case study is highly process-oriented, data mining is not employed to improve features.

5.3. Limitation and future direction

The majority of currently available research concentrates on demand forecasts, and they just seek to forecast a single parameter. Few publications consider ML approaches to predict the joint distribution of multiple parameters, whereas production planning parameters may be related (e.g., the demand and lead time). Furthermore, the volume of data available and its use vary widely from one manufacturer to the other. ML approaches available for predictive analysis are various (Kusiak 2017, 2019). Hence, trying to develop a general big data cleaning and prediction method for MRP systems may be a new research trend.

There are many discussions about the possible advantages of BDA technologies in the supply chain. In recent years, enterprises and researchers pay more and more attention to this research field. However, research focusing on the application of BDA in the production planning is very scarce. We identify the following avenues for future works on BDA in the production planning:

(1) The research of BDA in manufacturing systems is still at the preliminary stage. Some researchers study how to use BDA in supply chains. But they only test different BDA methods, do not provide a breakthrough in forecasting models. More research is required to provide the best way to apply generic machine learning tools in the production planning context.

(2) In terms of applications, companies have collected massive amounts of data, but have not sufficiently exploited them to support decision-making in the production planning. Some studies are required to validate the use of BDA in real use cases, by integrating BDA methods in the production planning of enterprises. In addition, the actual application of BDA would require focusing research on data cleaning, domain model, and predictive models for ERP systems.

(3) Research on the planning optimisation models based on BDA. Few papers further consider the impact of the forecast on the final production plans. There exist no comparison between BDA-based production planning and traditional production planning methods.

6. Simulation, optimization for the production planning in the Industry 4.0 era

Supported by IoT technologies, the DT provides a real-time picture of the factory. Based on historical data, BDA can predict the future value of planning parameters and estimate their variability. This section discusses prescriptive analytics methods, that combine machine learning, simulation, and optimization to prescribes the best course of actions to optimize the future plan. This section successively discusses data-
driven simulation, cloud simulation, simulation-optimization, and optimization under uncertainty approaches for production planning.

6.1. **Data-driven automatic modeling and simulation technology**

Manufacturing systems are very different from a company to the next, and it is not possible to create a generic simulation model for manufacturing. The construction of simulation models for large-scale production systems requires knowledge from business experts, and it is time-consuming. To shorten the time it takes to develop simulation models, researchers have proposed a data-driven method to automatically build simulation models (Liu et al. 2019a; Zhang, Zhang, and Yan 2019), which is named data-driven automatic modeling and simulation (DDAMS). These tools can reduce the total modeling time from several months to several weeks (Wang et al. 2021; Wy et al. 2011) and they reduce errors in the modeling process.

Figure 9 shows how the data-driven modeling and simulation technology works. First, we extract original information from the data centre of information systems (MES, ERP, APS (Advanced Planning & Scheduling System), etc.), and standardise these data. Second, we further classify and associate the data to build a structured data model. Third, based on original simulation objects, we personalise the internal logic and attributes of objects to build the general simulation model library, which meets the needs of the particular industry. Fourth, with the help of resource generation engine, we can use the objects in the general model library to generate the model layout quickly and automatically based on the layout data. Lastly, driven by real-time data, we can obtain a specific simulation operation model, and run it to get simulation results.

In addition, in the simulation operation phase, when production demands or production layout change, the traditional offline simulation will take several hours to update the data and adjust the model manually. The data-driven modeling and simulation method can update the data and adjust the model automatically and quickly. Therefore, with the data-driven automatic modeling and simulation technology, the production planning can respond to the uncertain parameters quickly. The data-driven modeling and simulation technology is also one of the important technologies in the DT (Zhang et al. 2019a; Wang et al. 2021; Luo et al. 2021). At the same time, the implementation of this technology is very dependent on the deployment of IoT and CMg in the system. This is because there are high requirements for collecting, storing
and sharing data from the information systems when performing automated modeling and simulation. Within the scope of our knowledge, there is little literature on the use of data-driven automatic modeling and simulation technology in the production planning. When we consider the data-driven automatic modeling and simulation for production planning, multi-level and multi-fidelity modeling is a research trend (Zhang et al. 2022). Because the scale of the model affects the efficiency of the model and the accuracy of the analysis, it is necessary to model and simulate the system at different levels and fidelity for different planning periods.

6.2. **Cloud simulation technology**

The modeling and simulation cycle of the manufacturing system is long and requires considerable expertise, time, and effort. Moreover, the real production system requires a scalable simulation solution that can be supported by cloud computing systems when they are expanded. Therefore, cloud-based simulation services have been proposed recently. Cloud simulation (CS), i.e. cloud-based factory simulation, uses cloud resources and services to simulate factories. In the framework of the digital twin, cloud simulation is no longer isolated from the actual production system. The cloud simulation platform is connected to the production physical system and can update the cloud simulation model in real-time with real-time data collected by IoT devices, which has higher requirements for platform security, transmission speed, and integration than merely storing the simulation model in the cloud.

At present, cloud simulation construction across various manufacturing fields has been studied to some extent (Zhou et al. 2019). When companies apply cloud simulation to the real system, they still face many challenges. Because the cloud simulation model is not general and reconfigurable, building a large-scale simulation model in the cloud is very difficult and time-consuming. The existing cloud simulation models cannot reconfigure and update automatically according to the changes in systems (Yu, Cao, and Schniederjans 2017). In some proposed cloud systems, the client cannot upload other models for simulation, and the user interface operability is poor (Chi, Pepper, and Spedding 2004). When conducting factory visualization and large-scale simulation in the cloud, the largest problem is the running speed of systems (Lindskog et al. 2012). For cloud simulation technology, the technology, which is used in the distributed simulation to transfer simulation components and add nodes to distributed architecture during running, can not be directly applied to cloud-based simulations (D’Angelo and Marzolla 2014). The differences of existing factory simulation systems in input format, processing logic, and data structure also hinder the smooth running of the cloud simulation system during operation (Chen and Lin 2017). Besides, the construction of the digital twin also needs to improve the cloud simulation technology (Coronado et al. 2018). For cloud simulation in the production planning, our focus should be on the coupling between different cloud simulation models. For example, the coupling between long-term and short-term planning cloud simulation models, the integration of planning cloud simulation models between different workshops in the same company, and the interaction between planning cloud simulation models of different customers and suppliers.
6.3. **Optimization model for the production planning**

Optimization models for production planning problems involve replenishment planning and lot-sizing problems. The target of the lot-sizing problem is to obtain production and procurement quantities and their timing (Yano and Lee 1995). Since the beginning of the twentieth century, researchers have solved some expansions of the lot-sizing problem, and have proposed numerous modeling approaches and algorithms (Buschkühl et al. 2010). With the deepening of research, the focus of research on the lot-sizing problem gradually changed (Louly and Dolgui 2013; Hnaien, Dolgui, and Wu 2016; Schemeleva, Delorme, and Dolgui 2018; Tavaghof-Gigloo and Minner 2020) from single-product single-period single-machine systems to complex multi-product multi-period multi-machine systems (Cunha et al. 2018). One of the most generic versions for the lot-sizing problem in the production planning is the multi-echelon multi-item capacitated lot-sizing problem (MMCLP). This problem’s target is to determine when to produce as well as the size of production lots to minimize the expected total cost, based on the demand, the BOM, the production capacity, and the lead time. The total cost comprises inventory holding costs, backlog costs, setup costs, production costs, and extra capacity costs. For the MMCLP, the mathematical optimization is the best instrument at present. In fact, the operation research community has put much effort into lot-sizing models, and has proposed several reformulations, cuts, and solution algorithms such as Lagrangian relaxation and cutting planes. Tempelmeier and Helber (1994); Tempelmeier (2006); Helber (1995); Helber and Sahling (2010) have done a series of studies about the decomposition approaches and Lagrangian relaxation based heuristic algorithms for the multi-level capacitated lot-sizing problem. These solution approaches offer opportunities for the improvement of large problem instances. Table A8 in Appendix gives the literature review about stochastic and distributionally robust optimization for MMCLP.

Furthermore, the new paradigm of an intelligent digital twin for production planning changes the optimization tools. Although the main mathematical model will remain mostly the same, its parameters can be better anticipated through BDA and ML. Another main change comes from constraint learning, which can make the model more accurate.

6.4. **Simulation-optimization approaches**

Simulation methods mainly include discrete event simulation (DES), agent-based simulation (ABS), and system dynamic simulation (SDS). These methods are commonly used for facility resource planning, capacity planning, and job planning. Simulation can provide a detailed representation of the production process, and can simulate the execution of a policy. Most simulation-optimization approaches use optimization methods (e.g., local search, gradient descent, genetic algorithms, . . . ) to optimize the input parameter of the simulation. In this context, the simulation is embedded in the optimization approach to evaluate the costs associated with the input parameters. For instance, Lim, Alpan, and Penz (2017) simulate the use of a dynamic inventory control policy under various sources of uncertainties, and optimize the parameters of the policy with a local search. Similarly, Liu et al. (2011) use a genetic algorithm that evaluates the expected cost of a production plan through a simulation. A major drawback of such approaches is the time-consuming solution evaluation by simulation, especially when multiple replicates are required to approximate the expected cost in an uncertain environment, or when the simulation is very detailed. An approach to
circumvent this issue is to build surrogate models (e.g., Osorio and Bierlaire 2013) to approximate the expected cost evaluated with the simulation. These surrogate models are learned with machine learning from past simulation, and they are used to reduce the number of solutions evaluated through simulation.

The state-of-the-art optimization approaches for lot-sizing models commonly encountered in the production planning rely on mathematical models solved with commercial solvers. This approach was also used in combination with simulation. In a simple framework, the simulation is only used to complete the decisions made by the analytical optimization model. For instance, Lim et al. (2006) use an optimization approach to set the capacity in the factory and a simulation model to compute the production plan. A more advanced setting is the recursive optimization-simulation approach, where the mathematical model is improved iteratively with the result of the simulation. For instance, Jung et al. (2004) solves a deterministic lot-sizing problem and iteratively adjusts the safety stock after evaluation in simulation that accounts for uncertain demand. This iterative approach was also recently applied for planning in a collaborative assembly line (Vieira et al. 2021), and for planning in a wafer fabrication production plant (Kim and Lee 2016).

For more information on simulation-optimization approaches, the interested reader is referred to Figueira and Almada-Lobo (2014). In the context of Industry 4.0, there is a new trend in the research and application of simulation-optimization methods. The real-time data collected by IoT devices can help simulation models simulate production systems more accurately. This means that simulation-optimization methods can solve more complex and large-scale problems. Then this creates a new challenge for the speed of finding the optimal solution for simulation-optimization methods. How to use algorithms to enhance the speed of finding the optimal solution is a problem to be solved. Furthermore, the generality and reusability of the algorithm development module coupled with the simulation model is also a concern. Overall, there is growing attention toward the simulation-optimization approaches, but their applications in the production planning remain scarce. We believe that such approaches must be investigated, since a detailed simulation complement the optimization approaches, and ensure that the computed production plan is implementable on the shop floor. Stochastic optimization can be seen as an integration of simulation and optimization since it directly incorporates scenarios to describe possible realizations of uncertain parameters in the optimization model.

6.5. Uncertainty

While the first studies on lot-sizing considered that all parameters are known, in practice, none of the planning parameters can be forecasted perfectly. Uncertainty may be defined as the difference between the amount of information required to perform a task and the amount of information already possessed (Galbraith 1973). Over the years, many researchers tried to formalize and model uncertainties in production systems (Sethi et al. 2002; Yano and Lee 1995). The production planning literature provides various approaches and models that consider a variety of uncertainties. The main four uncertain parameters in the production planning are demand, lead time, capacity, and yield.

(1) Demand uncertainty is critical for production planning, particularly for manufacturers with long production lead times (Aouam et al. 2018). Demand uncertainty has various forms, such as the order size and due date. For example,
customers submit a demand signal (a prediction of what their orders will be) long in advance of the due date in the semiconductor production system. They progressively change their orders as time passes until a firm order is secured. However, customers still want orders to be fulfilled on schedule, regardless of the extent of changes between the demand signal and firm order (Higle and Kempf 2010). In the context of digital manufacturing, manufacturers can expand the number of finished products, which leads to production upgrades of mass customization and mass individualization. Nevertheless, a new problem arises, that is, it becomes more difficult to forecast the demand for each product. On the one hand, the product has a shorter life cycle, and the demand varies faster over its life cycle. On the other hand, thanks to the amount of data collected, a better forecast is possible, which diminished the demand uncertainty.

2) **Lead time** refers to the number of periods between the placement of an order and its arrival. In production planning, we may distinguish between delivery lead time and production lead time. The first refers to the time required by suppliers to deliver components, whereas the second refers to the time between the release of an order to the shop floor and its shipping date. Delivery lead time uncertainty is common in practice and it is due to issues at the supplier production level or transport (Hnaien, Dolgui, and Wu 2016). The reason production lead times are uncertain involves several factors, such as inaccurate capacity constraints modeling when building the production plans, machine breakdowns, stochastic variations on the operation processing time (Aghezzaf, Sitompul, and Najid 2010). Some studies suggest modeling uncertain lead time with discrete support probability distribution built based on statistical data (Ben-Ammar and Dolgui 2018). In the context of mass manufacturing, more finished products mean more components are needed in the production process. This leads to an increase in the number of manufacturers and suppliers throughout the supply chain, resulting in a more complex overall supply chain. It also leads to an increased risk of late deliveries. With the DT, the lead time can be effectively shortened and predicted through real-time control and data traceability, while the interaction of data between upstream and downstream of the supply chain can also effectively reduce the risk of delivery delays.

3) **Production capacity** uncertainty refers to issues to ensure the shop floor can satisfy the required production load. There may be uncertainty about the available resource capacity due to machine breakdown or employee absenteeism, and uncertainty in the capacity consumption for an operation due to variable process duration, or product quality if the shop floor reworks or redoes bad quality parts. Another major source of problems is that the optimization models for planning only approximates the capacity roughly to produce a feasible plan. Note that the lead time uncertainty is often related to the capacity uncertainty. The capacity uncertainty is also related with workload, i.e. the demand from other clients or customers at the same time. In practice, even when a good scheduling tool is used, the resources may have idle times. In addition, in flexible production plants, it is difficult to estimate which resource will perform each task before doing the production schedule. While capacity uncertainty leads to infeasible plans, very few works consider planning under capacity uncertainty, when compared with the cases of demand and lead time uncertainty.

In Industry 4.0 manufacturing systems, we can monitor machine breakdowns in time so that repairs can be made or production schedules can be adjusted promptly. Through DT-based scheduling and control, the uncertainty of produc-
tion capacity can be greatly reduced. Furthermore, the information collected on the status of the machine can help to make maintenance forecasts and decrease the uncertainty of production capacity by scheduling more reasonable machine maintenance, which can also improve machine utilization.

4) **Yield uncertainty** occurs when bad quality parts cannot be re-worked or replaced by a new one. This situation occurs for operation with long processing time such as aluminum casting, or in multi-echelon systems, where producing an additional part is impossible when the components are not available. Yield uncertainty is also common in the disassembly of end-of-life items since the quality of components is only observed once the item is disassembled (Ben-Ammar, Bettayeb, and Dolgui 2020). Because the product life cycle becomes shorter, the production process lacks regularity and product quality is difficult to guarantee. The good news is that we can achieve quality control automatically through machine learning.

The classical approach computes the lot sizes under the assumption that all parameters are deterministic, whereas safety stock, safety lead times, and safety capacities are computed separately to hedge against the uncertainty. With the improvement of computation power and new development in optimization approaches, it is nowadays possible to integrate the uncertainty directly in the optimization problem with stochastic optimization (SO) approaches (Spall 2005). That is, random variables appear in the formulation of the optimization problem itself, which involves random objective functions or random constraints. Consequently, the research recently moved from the initial deterministic to non-deterministic lot-sizing model (Aloulou, Dolgui, and Kovalyov 2014; Tavaghot-Giglooo and Minner 2020). The majority of the research considers restrictive assumptions (single level, single period, and single item) to develop analytical models (Ertogral 2011; Sana 2013; Aloulou, Dolgui, and Kovalyov 2014). In particular, most studies don’t take into account the capacity constraints of manufacturing systems, when calculating lot sizes. This results in impractical production plans, long and uncertain lead times, and massive work-in-process inventories. In recent years, more scholars have studied more generic approaches able to cope with the complex multi-level/multi-periods/multi-item lot-sizing problems (Li, Tao, and Wang 2012; Thevenin, Adulyasak, and Cordeau 2021; Meistering and Stadtler 2019). Many studies consider a single uncertainty parameter (Yano and Lee 1995; Zikopoulos 2017; Kroer et al. 2018; Afsar et al. 2020), but more scholars have paid attention to the consideration of multiple uncertain parameters in recent years. For instance, demand and lead time are sometimes considered together (Tang et al. 2019; Köchel and Thiem 2011; Song and Dinwoodie 2008). Considering multi uncertain parameters in the stochastic optimization model to describe the production system more accurately is a future research trend, and it will also be a challenge. Finally, a large variety of methods were proposed to solve lot-sizing problems, such as fuzzy logic, scenario-based stochastic optimization, robust optimization, and game theory (Su 2017; Cunha et al. 2018; Carvalho et al. 2018; Simon Thevenin 2021; Zarei, Rasti-Barzoki, and Hejazi 2021).

6.6. **Limitation and future direction**

1) We find it difficult to solve the complicated lot-sizing problem under uncertainty, particularly in the dynamic decision framework. Because once new information is coming, production settings will be updated. Existing research only considers
small-scale cases in a basic setting (Thevenin, Adulyasak, and Cordeau 2020). When considering large-scale instances with multi-echelon BOM in a long planning horizon, we have to provide more effective heuristic algorithms. For instance, Thevenin, Adulyasak, and Cordeau (2021) demonstrate that the two-stage approximation is a useful heuristic algorithm for solving the lot-sizing problem with uncertain demand in the static-dynamic decision framework. More research is needed, however, to tackle a large-scale problem in a long time horizon, and one of the future directions is the fix-and-optimize method. Furthermore, we must build approaches to handle the problem in the dynamic decision framework.

(2) While most approaches assume the probability is known, this will never be true in practice, and the distribution can only be estimated. Distributionally robust optimization is an interesting class of approaches that optimize for the expected cost of the worst case distribution (Zhang, Shen, and Mathieu 2016), and its application to production planning must be further explored.

(3) The development of frontier technologies provides a better platform for data collection and sharing for stochastic optimization of production planning, and also puts forward new requirements for the solution speed and quality. More research is required to link optimization approaches with frontier technologies from Industry 4.0, and to validate these work in a realistic environment.

7. Conclusion and further research perspectives

In this study, we give a literature review and assessment on production planning in Industry 4.0. The paper focuses on how to apply the Internet of Things, cloud manufacturing, big data analytics, digital twins, simulation-optimization, and stochastic optimization in the production planning.

We will perform and apply these cutting-edge technologies in a real company for future research purposes, and will present a complete framework that covers not just production planning but also scheduling and connectivity protocols in detail. Furthermore, improving the heuristic algorithm and machine learning approach is also necessary for the MMCLP. Finally, another intriguing research work is to find ways to improve efficiency while minimizing the complexity of the integration system when it is integrated with other systems under the CPS environment in Industry 4.0.

8. Acknowledgement

The present work was conducted in the project ASSISTANT (https://assistant-project.eu/) that is funded by the European Commission, under grant agreement number 101000165, H2020 – ICT-38-2020, Artificial intelligence for manufacturing.

9. Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.
References


Guo, Daqiang, Mingxing Li, Ray Zhong, and George Q Huang. 2020a. “Graduation intelligent manufacturing system (GiMS): an Industry 4.0 paradigm for production and operations management.” Industrial Management & Data Systems.


Hasan, Mahmud, and Binil Starly. 2020. “Decentralized cloud manufacturing-as-a-service (CMaaS) platform architecture with configurable digital assets.” Journal of Manufactur-
on system sciences.


Liu, Wing Kam, George Karniadakis, Shaoqiang Tang, and Julien Yvonnet. 2019a. “A computational mechanics special issue on: Data-driven modeling and simulation—theory, methods, and applications.”


Song, Dong-Ping, and John Dinwoodie. 2008. “Quantifying the effectiveness of VMI and integrated inventory management in a supply chain with uncertain lead-times and uncertain demands.” *Production Planning and Control* 19 (6): 590–600.


Tao, Fei, Yongping Zhang, Ying Cheng, Jiawei Ren, Dongxu Wang, Qinglin Qi, and Pei Li. 2020. “Digital twin and blockchain enhanced smart manufacturing service collaboration and management.” Journal of Manufacturing Systems .


**Appendix A. Supplementary tables**
Table A1. Literature review about key techniques and tools for the implementation of a digital twin-driven production planning system.

<table>
<thead>
<tr>
<th>Category</th>
<th>Key technique</th>
<th>Content</th>
<th>Examples of tools</th>
<th>Related papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligent perception</td>
<td>Measurement</td>
<td>Laser measurement, image recognition measurement, conversion measurement, and micro/nano precision measurement</td>
<td>Micro-sensor, RFID, light detection and ranging system, depth camera, global positioning system</td>
<td>Donges and Noll (2016); Nagato et al. (2017); Jacob and Thiemann (2017); Dachyar, Zagloel, and Saraghi (2019); Tavana, Hajipour, and Oveisi (2020)</td>
</tr>
<tr>
<td>Data management</td>
<td>Data collection, transmission, storage, processing, fusion and visualization</td>
<td>Wire and wireless transmissions, database, interpretable- operable- traceable heterogeneous data fusion, data cleaning, data compression, data smoothing, data reduction, data clustering storage</td>
<td>Aspera, HBase, Spark, Echarts, Spyder</td>
<td>Lei (2018); Cupek et al. (2019); Ge et al. (2020); Liu et al. (2020); Xiao et al. (2021)</td>
</tr>
<tr>
<td>Modeling and simulation</td>
<td>Modeling of mathematical and simulation model</td>
<td>Mixed integer programming, Data-driven modeling and simulation, virtual reality, and augmented reality technology</td>
<td>Cplex, Pulp, Flexsim, Solid-Works, Anylogic</td>
<td>Shapiro (1993); Wu et al. (2013b); Pochet and Wolsey (2006); Salah et al. (2019); Luo et al. (2021)</td>
</tr>
<tr>
<td>Actuation</td>
<td>Cloud Manufacturing, blockchain</td>
<td>Cloud computing, smart contracts</td>
<td>Fernández-Caramés and Fraga-Lamas (2018); Wang et al. (2019b); Hasan and Starly (2020); Fosso Wamba et al. (2020); Rožman, Diaci, and Corn (2021)</td>
<td>Ray (2016); Berg and Vance (2017); Heidari (2019); Wang et al. (2019a); Wang and Luo (2021)</td>
</tr>
<tr>
<td>Interconnection</td>
<td>Virtual-real interaction</td>
<td>Heterogeneous resources real-time perception and access technology, multi-source/modal data fusion and encapsulation technology, multi-source data communication and distribution technology</td>
<td>MindSphere of Siemens, Jasper Control Center of Cisco, Jasper, Thingworx of PTC</td>
<td></td>
</tr>
</tbody>
</table>
Table A2. Literature review about DT frameworks.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Application</th>
<th>Viewpoint</th>
<th>Core methods/focus</th>
<th>Case study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tao et al. (2018a, 2019)</td>
<td>Product design</td>
<td>Product</td>
<td>BDA and CPS</td>
<td>The power transformer and bicycle, no data</td>
</tr>
<tr>
<td>Ivanov et al. (2019); Ivanov and Dolgui (2020)</td>
<td>Digital supply chain twins</td>
<td>Supply chain</td>
<td>Additive Manufacturing, BC, and BDA</td>
<td>No</td>
</tr>
<tr>
<td>Qi et al. (2019)</td>
<td>Digital supply chain twins</td>
<td>Supply chain</td>
<td>Five-dimension model, enabling technologies, enabling tools</td>
<td>No</td>
</tr>
<tr>
<td>Tao et al. (2018b)</td>
<td>Smart manufacturing</td>
<td>Manufacturing system</td>
<td>Lifecycle of manufacturing data, framework</td>
<td>Silicon wafer production line, figures of implementation interface</td>
</tr>
<tr>
<td>Lu et al. (2020)</td>
<td>Smart manufacturing</td>
<td>Manufacturing system</td>
<td>Review, connotation, reference model, applications, and research issues</td>
<td>No</td>
</tr>
<tr>
<td>Rossit, Tohme, and Frutos (2019)</td>
<td>Smart Manufacturing</td>
<td>PPC</td>
<td>Review in CPS</td>
<td>No</td>
</tr>
<tr>
<td>Agostino et al. (2020)</td>
<td>Smart shop</td>
<td>PPC</td>
<td>CPS</td>
<td>Scheduling in a job shop of a Brazilian supplier for the automotive industry</td>
</tr>
<tr>
<td>Zhang, Zhang, and Yan (2019)</td>
<td>Smart shop-floor</td>
<td>Workshop</td>
<td>CPS</td>
<td>Scheduling of the blisk machining, data</td>
</tr>
<tr>
<td>Ding et al. (2019)</td>
<td>Smart shop-floor</td>
<td>Workshop</td>
<td>CPS, operations control</td>
<td>Interface of operations control, no data</td>
</tr>
<tr>
<td>Guo et al. (2020c)</td>
<td>Fixed-position assembly islands</td>
<td>Graduation Intelligent Manufacturing System</td>
<td>The decision making mechanism with by IoT, cloud-based services and industrial wearable technologies</td>
<td>Laser equipment manufacturer</td>
</tr>
<tr>
<td>Li and Huang (2021)</td>
<td>Flexible assembly lines</td>
<td>GiMS</td>
<td>Production-intralogistics processes</td>
<td>Air conditioner manufacturer</td>
</tr>
</tbody>
</table>
### Table A3. IoT literature about production planning.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Focus</th>
<th>The degree of attention to the production planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tao et al. (2014a)</td>
<td>Cloud manufacturing service system</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Zhong et al. (2016)</td>
<td>Shop floor logistics</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Fang et al. (2016)</td>
<td>Production system (the product life cycle includes procurement, production and product recovery, and acquisition)</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Tao et al. (2017)</td>
<td>Inventory control policy</td>
<td>Focus on local issues</td>
</tr>
<tr>
<td>Wang et al. (2018)</td>
<td>Production planning and control</td>
<td>One of several concerns</td>
</tr>
<tr>
<td>Zuo, Tao, and Nee (2018)</td>
<td>Capacity consumption evaluation and analysis</td>
<td>Focus on local issues</td>
</tr>
<tr>
<td>Wang et al. (2020)</td>
<td>Shop floor material management</td>
<td>Focus on local issues</td>
</tr>
<tr>
<td>Bueno, Godinho Filho, and Frank (2020)</td>
<td>Smart production planning and control</td>
<td>One of several concerns</td>
</tr>
</tbody>
</table>

### Table A4. CMg literature about the production planning.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Focus</th>
<th>The degree of attention to the production planning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ning et al. (2011)</td>
<td>Architecture and key technologies</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Wu et al. (2013a)</td>
<td>Strategic vision</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Tao et al. (2014b)</td>
<td>Manufacturing service system</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Erol and Sihn (2017)</td>
<td>Intelligent production planning and control</td>
<td>One of several concerns</td>
</tr>
<tr>
<td>Ren et al. (2017)</td>
<td>Key characteristics and applications</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Yu et al. (2018)</td>
<td>Multi-level aggregate service planning</td>
<td>Focus on</td>
</tr>
<tr>
<td>Henzel and Herzwurm (2018)</td>
<td>Literature review</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Wang et al. (2019b)</td>
<td>Additive manufacturing</td>
<td>Focus on</td>
</tr>
<tr>
<td>Li et al. (2019)</td>
<td>Multibjective optimization</td>
<td>Focus on</td>
</tr>
<tr>
<td>Suginoouchi and Mizuyama (2021)</td>
<td>Production planning and revenue allocation</td>
<td>One of two concerns</td>
</tr>
<tr>
<td>Paper</td>
<td>Focus</td>
<td>The degree of attention to the production planning</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>Fosso Wamba et al. (2020); Leng et al. (2020); Li et al. (2021); Vu, Ghadge, and Bourlakis (2021)</td>
<td>Literature review about the supply chain/manufacturing system</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Herrgoß et al. (2020)</td>
<td>PPC in the semiconductor industry</td>
<td>One of several concerns</td>
</tr>
<tr>
<td>Rahmanzadeh, Pishvae, and Rasouli (2020)</td>
<td>Integrated innovative product design and supply chain tactical planning</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Christidis and Devetsikiotis (2016)</td>
<td>Smart contracts for IoT</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Zhang et al. (2019c,b); Pal et al. (2020)</td>
<td>IoT in in supply chain or smart manufacturing</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Kaynak, Kaynak, and Uygun (2019)</td>
<td>CMg architecture</td>
<td>Related</td>
</tr>
<tr>
<td>Hasan and Starly (2020)</td>
<td>Contemporary CMg-as-a-Service platforms including smart contract</td>
<td>Related</td>
</tr>
<tr>
<td>Kumar et al. (2020); Tan et al. (2021)</td>
<td>Smart contract for CMg</td>
<td>Related</td>
</tr>
<tr>
<td>Shahbazi and Byun (2021a)</td>
<td>Integration framework (BC, IoT and ML) for smart manufacturing</td>
<td>Weak related</td>
</tr>
<tr>
<td>Zhang et al. (2021)</td>
<td>Service power calculation of high-performance blockchain consensus for CMg in smart manufacturing</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Yu et al. (2020); Song and Moon (2019)</td>
<td>Framework for CPS</td>
<td>Mentioned</td>
</tr>
<tr>
<td>Vatankhah Barenji et al. (2020)</td>
<td>Ubiquitous manufacturing architecture for CPS</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Tao et al. (2020)</td>
<td>Smart manufacturing service collaboration and management in DT</td>
<td>Related</td>
</tr>
<tr>
<td>Deepa et al. (2020)</td>
<td>Approaches, opportunities, and future directions for BDA</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Shahbazi and Byun (2021b)</td>
<td>Smart Manufacturing Real-Time Analysis using ML method</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Paper</td>
<td>Application</td>
<td>Parameter</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Garre, Ruiz, and Hontoria (2020)</td>
<td>Food industries</td>
<td>Food industry</td>
</tr>
<tr>
<td>Meidan et al. (2011)</td>
<td>Semiconductor Cycle time manufacturing</td>
<td>Cycle time</td>
</tr>
<tr>
<td>Wang et al. (2018)</td>
<td>Semiconductor Cycle time fabrication systems (SWFS)</td>
<td>Cycle time</td>
</tr>
<tr>
<td>Mori and Mahalec (2015)</td>
<td>Eyeglasses Lead time (a flowshop manufacturing environment)</td>
<td>Lead time</td>
</tr>
<tr>
<td>Gyulai et al. (2018)</td>
<td>Steel production time</td>
<td>Production time</td>
</tr>
<tr>
<td>Lingitz et al. (2018)</td>
<td>Semiconductor Lead time manufacturing</td>
<td>Lead time</td>
</tr>
<tr>
<td>Öztürk, Kayalıgil, and Özdemirel (2006)</td>
<td>Hypothetical Lead time manufacturing environment (Simulation)</td>
<td>Lead time</td>
</tr>
<tr>
<td>Schuh et al. (2019)</td>
<td>Demonstration Factory time</td>
<td>Transition time</td>
</tr>
<tr>
<td>Paper</td>
<td>Key simulation technology</td>
<td>Application</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Wy et al. (2011)</td>
<td>DDAMS</td>
<td>Logistics-embedded assembly manufacturing lines</td>
</tr>
<tr>
<td>Liu et al. (2019a)</td>
<td>DDAMS</td>
<td>Many disciplines (physical and information) of science</td>
</tr>
<tr>
<td>Zhang, Zhang, and Yan (2019)</td>
<td>DDAMS</td>
<td>CPPS towards smart shop-floor</td>
</tr>
<tr>
<td>Zhang et al. (2019a)</td>
<td>DDAMS</td>
<td>Digital twin manufacturing cell</td>
</tr>
<tr>
<td>Wang et al. (2021)</td>
<td>DDAMS</td>
<td>In digital twin for the design, production, operation, and service of elevators</td>
</tr>
<tr>
<td>Zhou et al. (2019)</td>
<td>DDAMS and CS</td>
<td>Numerical control machining in cloud manufacturing</td>
</tr>
<tr>
<td>Luo et al. (2021)</td>
<td>DDAMS and CS</td>
<td>Automated flexible production lines in real smart factories</td>
</tr>
<tr>
<td>Chi, Pepper, and Spedding (2004)</td>
<td>CS</td>
<td>Production lines of automotive components</td>
</tr>
<tr>
<td>Lindskog et al. (2012)</td>
<td>CS</td>
<td>Discrete event simulation using 3D scans</td>
</tr>
<tr>
<td>D’Angelo and Marzolla (2014)</td>
<td>CS</td>
<td>A new simulation middleware and generic adaptive interaction architecture</td>
</tr>
<tr>
<td>Chen and Lin (2017)</td>
<td>CS</td>
<td>Model conversion among various simulation systems and the digital equipment identifier system</td>
</tr>
<tr>
<td>Yu, Cao, and Schniederjans (2017)</td>
<td>CS</td>
<td>Multi-agent simulation for supply chain</td>
</tr>
<tr>
<td>Paper</td>
<td>Focus</td>
<td>Uncertainty</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Quezada et al. (2020)</td>
<td>Production planning in remanufacturing system</td>
<td>Production capacity, demand, and costs</td>
</tr>
<tr>
<td>Behnamian et al. (2017)</td>
<td>Multi-level production planning</td>
<td>Levels</td>
</tr>
<tr>
<td>Haque et al. (2021)</td>
<td>Multi-stage decentralized supply chain</td>
<td>No</td>
</tr>
<tr>
<td>Thevenin, Adulyasak, and Cordeau (2021)</td>
<td>Static-static and static-dynamic decision frameworks</td>
<td>Demand</td>
</tr>
<tr>
<td>Meistering and Stadtler (2019)</td>
<td>Production planning in rolling schedules</td>
<td>Demand</td>
</tr>
<tr>
<td>Li, Tao, and Wang (2012)</td>
<td>Production planning</td>
<td>Demand</td>
</tr>
</tbody>
</table>

Table A8. Literature review about stochastic and distributionaly robust optimization for multi-echelon, multi-period, capacitated lot-sizing.