Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

MSD: Continuous Pharmaceutical Manufacturing Data for the 2024 MSOM Data-Driven Research Challenge

Tugce Martagan, Marc Baaijens, Coen Dirckx, James Holman, Robert Meyer, Oscar Repping, Bram van Ravenstein

To cite this article:
Tugce Martagan, Marc Baaijens, Coen Dirckx, James Holman, Robert Meyer, Oscar Repping, Bram van Ravenstein (2024) MSD: Continuous Pharmaceutical Manufacturing Data for the 2024 MSOM Data-Driven Research Challenge. Manufacturing & Service Operations Management 26(5):1587-1604. https://doi.org/10.1287/msom.2024.0860

https://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 26, No. 5, September–October 2024, pp. 1587–1604
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Data-Driven Challenge

MSD: Continuous Pharmaceutical Manufacturing Data for the 2024 MSOM Data-Driven Research Challenge

Tugce Martagan,[a,*] Marc Baaijens,[b] Coen Dirckx,[a,c] James Holman,[d] Robert Meyer,[e] Oscar Repping,[b] Bram van Ravenstein[c]

[a] Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands; [b] Ceva Santé Animale, 33500 Libourne, France; [c] MSD, 5831 AN Boxmeer, Netherlands; [d] Technology Management, GEA Group, Eastleigh SO53 4DG, United Kingdom; [e] Merck & Co., Inc., Rahway, New Jersey 07065

*Corresponding author

Contact: T.G.Martagan@tue.nl, https://orcid.org/0000-0002-4396-7102 (TM); marc.baaijens@ceva.com (MB); coen.dirckx@merck.com (CD); james.holman@gea.com (JH); robert_meyer2@merck.com (RM); oscar.repping@ceva.com (OR); bram.van.ravenstein@merck.com (BvR)

Received: February 23, 2024
Accepted: March 5, 2024
Published Online in Articles in Advance: April 29, 2024

https://doi.org/10.1287/msom.2024.0860

Abstract. To support the 2024 MSOM Data-Driven Research Challenge, Merck & Co., Inc., Rahway, New Jersey (hereafter "MSD"), provides pharmaceutical manufacturing data from a continuous tablet production setting. The data set contains approximately 300 million data points related to around 75 process parameters monitored over 120 hours. In this paper, we present the data set and share our vision to inspire and facilitate new applications of operations management (OM) methodologies in pharmaceutical manufacturing. We begin with an introduction to pharmaceutical manufacturing for OM researchers and then elaborate on emerging technologies, common industry challenges, and research opportunities. We explain the data set and propose a roadmap for future research directions. Researchers are welcome to examine the proposed research questions or analyze other research questions using the data set.

History: This paper has been accepted as part of the 2024 MSOM Data-Driven Research Challenge.

Funding: This work was supported by The Dutch Research Council - NWO VIDI Grant.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2024.0860

Keywords: OM practice • data challenge • pharmaceutical industry • biomanufacturing

1. Introduction

This 2024 MSOM data-driven research challenge is a joint effort of MSD's Human Health and Animal Health divisions, in collaboration with one of their equipment manufacturers, GEA. Our main objective in this initiative is to inspire and stimulate new research on the applications of data analytics and operations management (OM) methodologies to complement our expertise in life sciences.

Until recently, the competitive advantage in the (bio)-pharmaceutical industry has been driven primarily by life sciences (i.e., our knowledge of the biological and chemical processes to produce medicines). However, with increasing demand and competition, the industry is becoming more sensitive to costs and production footprints. In addition, the COVID-19 pandemic has highlighted the importance of rapid response and resilience in (bio)pharmaceutical manufacturing and supply chains. The competitive advantage in the industry is currently shifting toward cost efficiency and resilience. We are under constant pressure to make our manufacturing operations smarter and reduce costs and lead times to produce critical medicines for our patients.

The industry's transition from "science labs" to "smart operations" is not an easy process, and it requires a vision and a transformation toward a proactive use of data analytics and OM to inform daily decisions. As Oscar Repping, Executive Director at MSD Animal Health, states, *"The first thing I was taught was the belief that biological processes are highly unpredictable, and guess what, I believed it. Over time, I learned a 'tribal' knowledge of how to set up and influence these processes. These processes remained highly variable, and during my development within the biotechnology field, I came to the conclusion that we should use our data differently than we ever did before. Working towards a data-driven decision-making model instead of the best guess […]."* (Repping et al. 2023). To facilitate this transition, the industry would

benefit from an interdisciplinary approach, working in close collaboration with the INFORMS community, to harness the power of cutting-edge OM with data analytics and life sciences.

Our scope includes two emerging technologies, also referred to as *key enabling technologies*: continuous manufacturing and real-time release testing. These technologies are relevant to both the pharmaceutical and biopharmaceutical industries. The use of these technologies in commercial-scale production is relatively new in the industry, but it also represents a significant opportunity to transform manufacturing practices in the near future. The data we are sharing with the MSOM community come from a continuous tablet-manufacturing process. It captures around 300 million data points collected over 120 hours of commercial-scale production. Although the data set is related to pharmaceutical tablet manufacturing (e.g., small molecules), we believe the models and insights generated from this initiative can be generalized to more complex settings in the biopharmaceutical industry (e.g., large molecules) and pave the way for successful implementations in the future.

At MSD Animal Health in The Netherlands, we have been on a long journey of OM implementations in collaboration with the Eindhoven University of Technology. Through our work presented in the 2022 Franz Edelman competition, we have demonstrated that linking OM to pharmaceutical manufacturing drives sustainable and substantial improvements (Martagan et al. 2023). To share our experience with a broader community, we have also participated in the seminar series organized by the MSOM practice platform (Repping et al. 2023). In this 2024 MSOM data-driven research challenge initiative, we aim to share our vision and future research directions with the OM community. For this purpose, we begin with an introduction of products and production processes in Section 2. We then introduce the concept of continuous manufacturing and real-time release testing in Section 3. We present the data in Section 4 and propose future research directions in Section 5. We finish with concluding remarks in Section 6.

2. Introduction to (Bio)pharmaceutical Manufacturing

We provide a short introduction to drug manufacturing for OM researchers. Our scope includes both pharmaceutical and biopharmaceutical drugs. We begin with an overview of the products (Section 2.1) and production processes (Section 2.2). We also elaborate on the human and animal health industries (Section 2.3) and discuss common industry challenges (Section 2.4).

2.1. Overview of Products

Drugs produced using biomanufacturing technologies are called *biopharmaceuticals* (or biologics). One of the most well-known examples of biopharmaceuticals is COVID-19 vaccines produced via mRNA technology (see Online Appendix A.1 for an overview of different types of biopharmaceuticals).

Biopharmaceuticals are fundamentally different from pharmaceutical drugs. We summarize the main differences in Table 1. The primary difference is related to the method of synthesis. More specifically, pharmaceuticals are obtained from chemical reactions, whereas biopharmaceuticals are biologically synthesized by using genetically engineered living organisms (e.g., bacteria). In addition, the molecular structure of biopharmaceuticals is larger and more complex compared with pharmaceuticals. For example, an aspirin (acetylsalicylic acid) molecule consists of only 21 atoms, whereas an immunoglobulin molecule (also known as IgG1, a common example of biopharmaceuticals) includes more than 20,000 atoms (Globerman 2016). The bonds between the atoms of biopharmaceuticals are also more complex. Their complex structure allows us to equip biopharmaceuticals with smart capabilities; that is, these drugs can search for a tumor inside the body without damaging healthy cells. On the other hand, biopharmaceuticals are more susceptible to handling and storage conditions. For example, most biopharmaceuticals require cold storage during production and supply chain logistics to preserve their molecular integrity and quality.

Biopharmaceuticals and pharmaceuticals are commonly referred to as *large* and *small* molecules, respectively. However, we note that not all "small" molecules are chemically synthesized. For example, antibiotics are known as small molecules, although they are biologically synthesized from fungi. In this paper, we broadly use the term biopharmaceutical to refer to large or small molecules that are biologically synthesized from living organisms.

2.2. Overview of Production Processes

All (bio)pharmaceutical operations are under the heavy influence of regulations and targeted expectations on

Table 1. Main Differences Between Pharmaceuticals and Biopharmaceuticals

	Pharmaceuticals	Biopharmaceuticals
Method of synthesis	Chemical	Biological
Molecular size and structure	Small and relatively simple	Large and highly complex
Sensitivity to handling and storage	Low	High

both the quality and safety of the drugs produced. These expectations are described in the international prerequisites, such as the current Good Manufacturing Practice guidelines (cGMP) of the European Medicines Agency (EMA) and the Food and Drug Administration (FDA), as well as the product-specific dossiers and Process Analytical Technologies (PAT) guidelines. To provide contextual information on the 2024 MSOM data-driven research challenge, we present a brief overview of the production process of biopharmaceuticals (Section 2.2.1) and pharmaceuticals (Section 2.2.2).

2.2.1. Production of Biopharmaceuticals. Biomanufacturing operations can be roughly categorized into two steps: upstream and downstream operations. Upstream operations focus on the production of the desired active ingredients via fermentation. The fermentation process is typically conducted in a stainless-steel vessel called a bioreactor and makes use of living organisms (e.g., bacteria) to produce the desired active ingredients. The output of fermentation is a liquid mixture consisting of the desired active ingredients along with unwanted impurities. These impurities consist of various types of metabolic byproducts such as dead cells, ammonia, and lactate. The accumulation of the desired active ingredients along with unwanted impurities during fermentation also leads to an interesting trade-off in harvesting decisions (see, e.g., Martagan et al. 2020, Koca et al. 2023). Subsequent downstream operations focus on eliminating unwanted impurities through a series of purification steps (e.g., centrifugation, chromatography, filtration) in order to achieve certain prespecified standards on drug quality and safety. Downstream operations also include fill-and-finish operations where the products are packaged and labeled into their final form. We refer to Online Appendix A.3 for more information on biomanufacturing processes and technologies.

2.2.2. Production of Pharmaceuticals. The data provided for the 2024 MSOM data-driven research challenge is related to the continuous manufacturing of a (small molecule) pharmaceutical drug in the form of a tablet. Therefore, we focus on this particular pharmaceutical production process to introduce our problem setting. Our production setting consists of four major steps: (1) formulation, (2) blending, (3) compression, and (4) coating. During the formulation step, active pharmaceutical ingredients are chemically synthesized, and other required raw materials are prepared. Next, all the individual raw materials (also including active ingredients) are mixed into a uniform blend. The output of the blending process is typically a powder mixture. Next, a tablet press is used to compress the powder mixture coming from the blender into tablets of uniform size and weight. As a final step, the tablets are coated with a thin polymer-based film. In a traditional pharmaceutical manufacturing setting, each of these production steps is a separate batch process with relatively long setups and manual interventions. However, our problem setting adopts an emerging technology for continuous blending, compression, and coating operations (we elaborate on the concept of continuous manufacturing in Section 3).

2.3. Human vs. Animal Health Industries

The data challenge is a collaborative effort of MSD's human and animal health divisions. Therefore, we briefly elaborate on the similarities and differences between these two industries. The drug development processes (e.g., clinical trials and regulatory approvals for market entry) may be different between the human and animal health industries because they target different species. However, the manufacturing operations are similar in both human and animal health applications; that is, both of these industries use the same equipment technologies and production processes and adopt similar cGMPs. Moreover, both industries are subject to stringent regulatory requirements. Therefore, they share similar manufacturing and operational challenges.

The animal health industry faces additional pressures that make it critical to improve manufacturing efficiency. First, most regulatory requirements between human and animal health are equally stringent, but animal health products may require additional studies to ensure food safety for human consumption. Second, unlike human drugs that are often subsidized, animal owners are responsible for paying full price for these drugs. For the manufacturers, this leads to smaller margins to cover the risks and costs associated with research, development, and production. Lastly, the animal health industry serves several different species as opposed to only one species in human health. The large portfolio of products with different manufacturing protocols leads to additional operational complexities.

2.4. Common Industry Challenges

We briefly discuss common industry challenges related to production processes, factory dynamics, and market needs, as summarized in Table 2.

2.4.1. Production Processes. A common industry challenge in both biopharmaceutical and pharmaceutical manufacturing is process uncertainty (e.g., batch-to-batch variability in quality, yield, processing times, and costs). In biopharmaceutical manufacturing, the process uncertainty can be attributed mostly to the use of living organisms and biologically synthesized materials during production. In both biopharmaceutical and pharmaceutical manufacturing, variability in raw materials is another common factor leading to variability in

Table 2. Common Industry Challenges in (Bio)pharmaceutical Manufacturing

Category	Manufacturing challenge
Production processes	Process uncertainty and batch-to-batch variability
	Optimal control and data analytics to achieve robustness
Factory dynamics	Coordination of multiple interdependent tasks under process uncertainty and production constraints
Market needs	Increasing market demand
	Increasing pressure to reduce costs and improve process efficiencies

process outcomes. Other factors include the inherent complexity of the biological and chemical processes and the limitations in our scientific understanding of these processes. Moreover, it may be difficult to optimally control these processes in real-time because of limitations in sensors or data analytics platforms. Furthermore, the production system includes a high number of exogenous input parameters that cannot be controlled, such as media composition (or other raw materials), which is often made by the suppliers.

The manufacturing of (bio)pharmaceuticals follows a predefined "recipe" describing the production protocols. For example, the recipe of fermentation describes the critical process parameters (CPP), quality attributes (CQA), and their control limits. In (bio)pharmaceutical manufacturing, we use the term "golden profile" to represent the desired trajectory of CPPs and CQAs over time. Therefore, the golden profile represents the "ideal" process profile and the boundaries (e.g., statistical control limits) within which the CPPs and CQAs mimic an optimum outcome for a particular production process. A critical challenge is to define an optimal course of action to bring the process back to its golden profile when it starts to deviate from it. Current industry practices typically define these control policies based on domain knowledge (i.e., based on the underlying biological and chemical dynamics). However, we believe that an interdisciplinary approach combining life sciences, data sciences, and OM is needed to achieve robustness and cost-effectiveness. In Section 5, we elaborate on a portfolio of optimal control problems to improve the performance of these production systems.

Another common challenge concerns how to make the best use of process data to achieve robustness. The industry is encouraged to measure, analyze, and control the CPP and CQAs over time. Often, these parameters and their respective boundaries in relation to efficacy and safety are known. However, the cross-functional interactions between the process parameters (i.e., how they influence each other in a positive and negative correlation toward the golden profile) are not fully understood at all times. A large number of process data can already be collected on, in, and at line. In addition, recent developments provide an impressive number of new sensors (e.g., Raman probes) that might generate additional important data. However, the industry currently has a limited understanding of which additional data need to be collected to achieve the best performance from these manufacturing systems. We need to close the gap between process measurements, data analysis, and optimal control to achieve robustness.

2.4.2. Factory Dynamics. The dynamics of a (bio)pharmaceutical manufacturing factory can be very complex. For example, when we performed an end-to-end mapping of all production activities in our Boxmeer facility in The Netherlands, we found that our manufacturing system involved more than 8,000 interdependent production steps (e.g., inspections, documentation, preparation of raw materials, setup of equipment, etc.). In addition, we have several production constraints that add additional complexity to planning and scheduling (see Martagan et al. (2023) for an extended list):

- Zero wait-time constraints. Active ingredients of biopharmaceuticals are fragile, and their properties can easily deteriorate if they wait in between production steps. Hence, each batch needs to smoothly flow throughout our 8,000 production steps with zero waiting time in between steps. There are only a few points during production where the batch can be frozen to wait for a subsequent step (e.g., after fermentation and hence, before starting purification).
- Coordination of multiple interdependent activities under uncertainty. (Bio)pharmaceutical processes involve variability in processing times, costs, production yields, and quality. Moreover, variability propagates, resources are limited, and operational decisions made at an earlier step have a magnifying impact on subsequent processes.

2.4.3. Market Needs. The market demand for both human and animal health products is increasing. It is anticipated that the global animal health market will grow at a compound annual growth rate (CAGR) of 8.8% during the forecast period 2023 – 2030, whereas the global pharmaceutical market size for human health will grow at a CAGR of 5.9% in the same period (Grand View Research 2022, Skyquest 2023). With increasing demand and competition, the industry is experiencing higher pressures to reduce costs and lead times. There is an increasing need for innovative applications of data

sciences and OM to facilitate data-driven, automated decisions to improve manufacturing efficiency.

3. Key Enabling Technologies and Trends

Several technology roadmaps have been developed by the industry, regulators, and public-private partnerships (see, e.g., BioPhorum Operations Group 2022, ICH Q13 2023, NIIMBL 2023). In this section, we highlight two emerging technologies: continuous manufacturing (Section 3.1) and real-time release testing (Section 3.2). We reflect on common industry challenges and opportunities in these technologies (Section 3.3) and elaborate on the scope of the data challenge (Section 3.4).

3.1. Continuous Manufacturing

In contrast to batch manufacturing with multiple discrete steps and long setup times in between steps, continuous manufacturing (CM) relies on a continuous (e.g., nonstop) flow of materials during production. Consider the fermentation process described in Section 2. In classical batch fermentation, the medium is added only at the start of fermentation, and the produced active ingredients are collected only at the harvest. In contrast, raw materials are continuously fed in a continuous fermentation system while the produced solution (containing the active ingredients) is continuously discharged. This way, the fermentation process maintains a steady state of control with continuous feeding and discharge of materials and runs for a longer period compared with batch processes. CM is a key enabling technology for both pharmaceutical and biopharmaceutical manufacturing and can be implemented at various degrees, for example, only at selected unit operations or the entire end-to-end manufacturing system.

CM offers several advantages, including smaller manufacturing footprints, reduced challenges with scale-ups, shorter lead times, and higher productivity with process intensification. It is an emerging technology to produce larger volumes of high-quality products at lower costs (Balfour 2021). Despite its benefits, the use of CM remains limited in the industry for several reasons. First, transitioning to CM may require high investment costs on equipment and infrastructure, and hence, the business case with risk-benefit trade-offs may not always be clear for new technologies (Lee 2017). Moreover, CM implementation requires advanced approaches to automation, data analytics, real-time monitoring, and optimal control. In some specific application settings, sensors or equipment technologies may not be mature enough to support the commercial use of CM. There are also regulatory uncertainties related to all new technologies. Nevertheless, we believe that the future of CM is highly promising, as supported by two recent initiatives. First, the U.S. Congress passed a bill in 2021 supporting advanced and continuous pharmaceutical manufacturing (Giaquinto and Auchincloss 2022). This bill supports the development, review, and approval of products manufactured using CM. Second, the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) published a new guideline, ICH Q13, on the continuous manufacturing of drugs substances and drug products. The ICH Q13 guideline is an important step forward in describing "scientific and regulatory considerations for the development, implementation, operation, and lifecycle management of continuous manufacturing" (ICH Q13 2023).

We highlight a few scientific and regulatory considerations from ICH Q13, with a particular focus on OM-driven aspects.

- **Input materials.** Variability in raw materials can occur for several reasons in practice, for example, aging of materials, batch-to-batch variability, and changes made by suppliers. Therefore, it is important to understand how the attributes of raw materials affect CM and the resulting product quality and also incorporate these input-output relationships into control strategies.
- **State of control.** ICH guidelines define the concept of state of control as "a condition that provides assurance of continued process performance and product quality" (ICH Q10 2015, and ICH Q13 2023). Creating control strategies that meet the specific needs of CM is essential. The right combination of descriptive, predictive, and prescriptive analytics can help identify the target set points and control limits, detect process disturbances in real time, and define optimal control policies to bring the process back to its golden profile when there are drifts. Other aspects related to the state of control include model maintenance (e.g., regular validation and improvement of the models used), data collection and monitoring strategy (e.g., identification of the critical process parameters and their monitoring frequency), and sampling strategy (e.g., the amount, location, and frequency of samples to be collected).
- **System dynamics.** CM systems should be flexible to handle disturbances while maintaining product quality. Therefore, it is important to define control strategies from a system perspective to ensure a continuous flow of materials and achieve resilience (e.g., consider interdependencies between production steps, and plan for lead time and capacity buffers). It is also crucial to have well-defined contingency plans. These include determining how far to trace back in case of quality issues or disturbances (*traceability strategy*), and when to divert products to scrap or other unit operations for rework (*diversion strategy*).

We further elaborate on the scientific and regulatory aspects described above when we propose future research directions in Section 5.

3.2. Real-Time Release Testing

Medical drugs must comply with stringent regulatory requirements on safety, efficacy, and quality. To ensure compliance with these requirements, the industry conducts various types of end-product release tests. However, these tests are typically offline (or at line), expensive, labor intensive, and time consuming. As a main premise, the concept of real-time release testing (RTRT) offers a significant opportunity for reducing these tests by replacing them with advanced analytics and process control. The ICH guidelines define the concept of real-time release testing as "the ability to evaluate and ensure the quality of in-process and/or final product based on process data, which typically includes a valid combination of measured material attributes and process controls." (ICH Q8 2017). Other advantages of RTRT include an enhanced understanding of processes based on a large set of data, real-time control and optimization, and timely interventions to ensure quality. The concept of RTRT is relevant to a wide range of settings in (bio)pharmaceutical manufacturing, for example, chemical or biological synthesis, batch processing, and CM. Although RTRT is not a regulatory requirement for CM, both RTRT and CM require similar initiatives on analytics and automation.

RTRT focuses on the monitoring and control of the process parameters to predict quality attributes. RTRT recognizes that a systematic combination of descriptive, predictive, and prescriptive analytics, together with the PAT tools, may provide a greater assurance of product quality to replace end-product release tests (European Medicines Agency 2012). In this setting, it is critical to have a firm understanding of the processes and the relations between the process parameters, in-process material properties, and product attributes. When communicating with regulatory authorities, it is important to clearly explain and justify the connection between end-product testing, material attributes, process monitoring, and acceptance criteria. Moreover, predictive models should be explained, justified, and verified at the commercial site (European Medicines Agency 2012). Depending on the application context, these models can be based on simple mechanistic relationships from life sciences, multivariate statistical models, or more complex machine-learning models. Most often, RTRT aims to develop real-time surrogate methods for replacing product release tests. The right selection of sensors and their locations are also important for RTRT. The decision-makers should consider the risks and trade-offs related to measurement speed, accuracy, robustness, repeatability, and costs. Moreover, contingency plans (e.g., when there is missing data or equipment failure) should be specified on alternative testing and monitoring strategies. Technology infrastructure is another key element for RTRT to facilitate automated sampling and systematic data storage.

In summary, successful RTRT implementations require a collaboration of multiple disciplines, including life sciences (e.g., mechanistic models), process technologies (e.g., sensors, cyber-physical systems), data science (e.g., machine learning, predictive analytics), and operations management (e.g., stochastic optimization, new regulatory outlooks). Currently, there are only a few documented examples of industry-scale RTRT implementations in the pharmaceutical industry (see, e.g., Markl et al. 2020). To the best of our knowledge, RTRT has not been implemented in commercial-scale biologics production yet. Nevertheless, we believe that the data and insights obtained from relatively simpler systems (e.g., small molecules) will pave the way for successful results in more complex systems (e.g., large molecules).

3.3. Challenges and Opportunities in CM and RTRT

We discuss the current industry challenges and opportunities related to CM and RTRT. We elaborate on three main aspects: (1) technology, (2) operations, and (3) regulations. Our discussion is relevant to both pharmaceutical and biopharmaceutical production.

3.3.1. Technology. Moving from batch-based to continuous systems and/or from offline tests to real-time control requires a substantial shift in technology and mindsets. Many of the gold-standard techniques, which were demonstrated to be robust and widely accepted, may not necessarily be suitable for CM or RTRT. For example, in vitro cell culture is a gold-standard test to detect contaminants, yet it is an offline test with 28 days of incubation time (Jiang et al. 2017). Nevertheless, several sensors and other promising technologies are currently being developed to attain new gold standards for CM and RTRT. On the other hand, new technology adoption decisions may be challenging because of resource limitations, potential risks with scaleups, and regulatory uncertainties. New collaborative approaches between regulators, equipment manufacturers, and (bio)pharmaceutical manufacturers may facilitate the development and adoption of new technologies.

3.3.2. Operations. A successful implementation of CM and RTRT requires knowledge from multiple disciplines (e.g., artificial intelligence, OM, life sciences). As a common challenge, the industry needs more experts with multidisciplinary proficiency to develop and implement these models in practice. This may also lead to difficulties when transferring knowledge from one factory to another.

The industry has a regulatory commitment to keep the models accurate and valid. In a CM and RTRT setting, it is critical to establish effective approaches for model maintenance, validation, and verification. When

necessary, models should be revised, reevaluated, and compared with their previous versions, and the regulatory authorities should be informed. However, the overall process of model maintenance and validation has its own challenges. For example, frequent updates may be needed because of the inherent variability of chemical and biological systems, fluctuations in raw materials, changes made by the suppliers, changes in the hardware, etc. In this setting, it is also important to establish mechanisms for timely and efficient communication with the regulators.

Another operational challenge is to maintain the processes in a continuous *state of control* and also to demonstrate (and validate) this state of control with data and models. Especially in the earlier phases of CM and RTRT implementations, there may be high levels of uncertainties because of the limited number of process data collected. However, as the system generates more data over time, there will be significant opportunities for exploiting data-driven decision-making approaches. For example, Bayesian statistics or other machine-learning approaches could help train and improve these models and control algorithms over time. Moreover, we may need to collect less data or sample less frequently to maintain the state of control over time. If we can demonstrate a consistent and successful use of data- and OM-driven approaches in CM and RTRT, this may help reduce the amount of data required by regulators.

3.3.3. Regulations. The adoption of CM and RTRT may come with some regulatory uncertainties and changes to established standards. For example, standard concepts such as the definition of a "batch" have been questioned and revisited for CM. Another implementation challenge could be a lack of precedence in the industry. For example, companies may hesitate to be the first when approaching regulators. Nevertheless, regulatory authorities have been taking initiatives to encourage open dialogue with the industry. For example, the FDA's Emerging Technology Program provides access to written guidance, meetings, and preoperational visits. Finally, we note that implementations of AI and OM methodologies in (bio)pharmaceutical manufacturing are still very limited compared with other industries. However, the industry and regulators are showing increasing interest and a more open mindset, as demonstrated by new guidelines and discussion papers on this domain (see, e.g., Center for Drug Evaluation and Research 2023 and Food and Drug Administration 2019, 2023).

3.4. Scope of the MSOM Data-Driven Research Challenge

We provide production data related to the continuous manufacturing of tablets. These data are collected in collaboration with GEA, one of the world's largest systems suppliers for the food, beverage, and pharmaceutical sectors. The data set can help generate insights to support the broader use of CM and RTRT in the pharmaceutical and biopharmaceutical industries. Our data set can be used to support data-driven approaches, but we also welcome model-driven approaches, such as stylized models that use the data set to support modeling assumptions. We believe that new approaches and insights generated by the OM community will pave the way for successful CM and RTRT implementations in the industry.

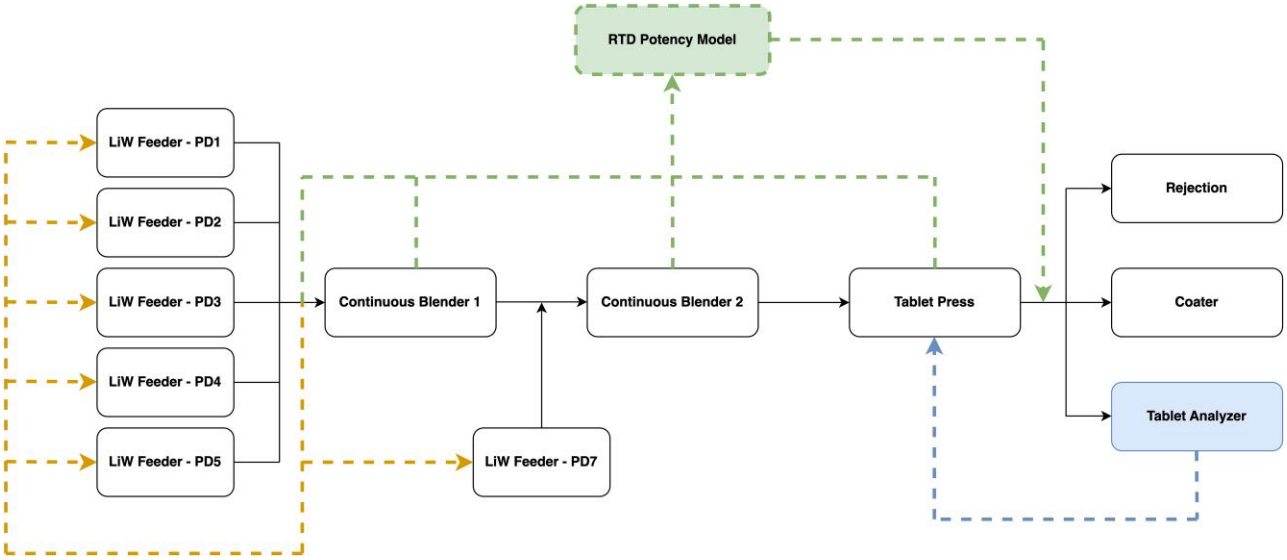
4. Description of Data

The data set presented in this section is based on our previous work (Holman et al. 2021), in which MSD and GEA demonstrated the robustness of a CM line for direct compression and coating of a pharmaceutical drug. We first describe the problem setting (Section 4.1) and then present a high-level overview of our data set (Section 4.2). The data set can be divided into two categories: machine data consisting of inline process parameters (Section 4.3) and contextual quality data on raw materials, tablet specifications, and content uniformity (Section 4.4). We propose research questions that can be addressed using the data set (Section 4.5).

4.1. Problem Setting

Figure 1 shows a schematic overview of our problem setting with continuous blending, compression, and coating operations (see Holman et al. 2021 for details). In this setting, raw materials are fed into benders via a so-called loss-in-weight (LiW) feeder. The LiW feeders are designed to feed the blender with a predetermined amount of raw materials at the correct ratio and the right pace. It is a gravimetric feeder that directly measures the raw materials' weight to maintain their flow at a predetermined rate (i.e., see GEA Pharma and Healthcare 2021 for a video on LiW feeders). At the start of the process, we have five LiW feeders (PD1–PD5 in Figure 1) to deliver a controlled flow of individual raw materials to the first continuous blender (Blender 1). Then, Blender 1 is used to create a homogeneous mixture using a high level of holdup and shear. Next, LiW feeder PD7 adds a lubricant to the mixture at the second continuous Blender 2 (note that LiW feeder PD6 does not exist because of internal naming conventions). Finally, a rotary tablet press is used to press the final blend into core tablets. After the tablet press operation, three possible routes exist. Depending on whether the tablet meets the required quality checks, it is either accepted and sent to the coater or diverted (rejected) based on prespecified quality attributes. As part of routine offline samples, tablets can be also sampled and sent to the tablet analyzer for additional checks on tablet specifications such as weight, thickness, and

Figure 1 (Color online) Schematic Overview of the Problem Setting (Solid Arrows Represent Product Routings, Dashed Arrows Denote Feedback Controls)



hardness. In addition, the quality attributes include blend potency (i.e., a measure of the concentration of the active ingredients contained in a tablet), and the portion of the tablets that have been subjected to an out-of-specification (OOS) alarm.

Common industry practices in CM work with the residence time distribution (RTD), a probability distribution function describing the amount of time a single particle spends in the system (herein, the term "system" could represent a unit operation such as blending or the entire end-to-end CM system). In our previous work, we used mathematical models to capture the RTD at the continuous Blenders 1 and 2 and the continuous tablet press operation (Holman et al. 2021). RTD is an important measure to facilitate potency predictions and traceability decisions. For example, we used the RTD information to predict potency after continuous Blender 1 and 2. In case of quality failures, the RTD information is used to support decisions on how far to trace back and scrap the produced (defective) materials.

Our CM system is equipped with several control mechanisms as depicted in Figure 1:

- **Local feedback controls at the LiW feeders.** The LiW feeders collect data related to the input materials (e.g., net weight of materials flowing per unit time) and the LiW feeder-related operating parameters (e.g., screw speed, feed factor, and motor speed of LiW feeders). Control mechanisms, indicated by orange dashed lines in Figure 1, are built into LiW feeders to adjust operating parameters (e.g., adjust screw speed as a function of the mass of materials flowing per unit time) and provide a target flow rate of materials. Moreover, the LiW feeder data can be used to generate a better understanding of raw material variability and the impact of raw material properties on product quality.
- **Local feedback controls at the tablet press.** A tablet press control mechanism (indicated by the blue dashed line in Figure 1) adjusts the compression force and fill depth to meet the predetermined tablet specifications on weight, thickness, and hardness. Moreover, the tablet compression system is capable of rejecting tablets that do not meet these predetermined tablet specifications.
- **Alarms and process traceability.** Alarm limits are set throughout the process based on predefined requirements (see Holman et al. 2021 for details on alarm triggers and the associated specification limits). When an alarm is triggered at a unit operation, 100% of the materials in that unit operation are considered to be OOS for the duration of the alarm. Subsequently, the RTD model can be used to track the OOS material throughout the system. In this case, all produced tablets are rejected until the OOS signal switches off and the process is back within the specified limits.
- **Potency prediction and rejection.** Information collected at each LiW feeder is used in the RTD models to predict potency. The potency estimations at the blenders and tablet press are used to decide whether to accept or reject tablets based on their estimated potency. The LiW data are not directly used to make rejection decisions but can trigger OOS alarms. The control mechanism used for tracking OOS and predicting blend potency by the RTD model is indicated by the green dashed lines in Figure 1.
- **Routine offline samples.** Tablets were sampled from the output of the table press for additional checks

Table 3. High-Level Overview of the Data Set

File name	Type of data	Additional information
LiW Feeders 1.csv	Machine data (Section 4.3)	Measured every second
LiW Feeders 2.csv	Machine data (Section 4.3)	Measured every second
Blenders.csv	Machine data (Section 4.3)	Measured every second
Tablet Press.csv	Machine data (Section 4.3)	Measured every second
Humidity.csv	Machine data (Section 4.3)	Measured as relative humidity percentage every 15 minutes
Temperature.csv	Machine data (Section 4.3)	Measured in degrees Celsius every 15 minutes
RM Tablet Properties and Drum Change.xlsx	Contextual quality data (Section 4.4)	Averaged over ten results
RM Content Uniformity.xlsx	Contextual quality data (Section 4.4)	Measured every 30 minutes and averaged over 10 results
RM Material Properties.xlsx	Contextual quality data (Section 4.4)	Triplicate measurements for each material sample tested
Logbook Long Run Days.xlsx	Event logbook (optional file as an appendix)	Events are recorded in 30-minute intervals

on their weight, thickness, and hardness. Additional tablet samples were taken every 30 minutes to check their specifications and measure their pharmaceutical ingredient concentration at line. Moreover, powder samples were taken from each raw material lot to measure their properties.

4.2. Overview of Data Set

The data set consists of several CSV files and Excel spreadsheets. Table 3 presents a high-level summary of the different files to help navigate the reader.

4.3. Machine Data

For 121 hours and 18 minutes at time intervals of one second, inline process parameters corresponding to the LiW feeders, blenders, and tablet press are collected. Table 4 provides an overview of the process parameters measured for the LiW feeders. We refer to Online Appendix 3.3 for background information on LiW feeders and their operating mechanisms.

First, for each of the feeders, the feed factor is determined (Feed Factor PD1–PD7). The feed factor is the theoretical maximum amount of powder that the feeder

Table 4. Inline Process Parameters, File '*LiW Feeders 1.csv' and 'LiW Feeders 2.csv*'

Column name	Corresponding columns (file name)	Description	Unit of measure
Feed Factor PD1–PD7	B-G (LiW Feeders 1.csv)	Theoretical maximum amount per screw rotation LiW feeders	Gram/screw revolution
Screw RPM PD1–PD7	H-M (LiW Feeders 1.csv)	LiW feeder screw speed	Rotations per minute
VolMode PD1–PD7	N-S (LiW Feeders 1.csv)	Binary variable indicating whether the LiW feeders are operating on volumetric (1) or gravitational mode (0)	{0,1}
Massflow PD1–PD7	T-Y (LiW Feeders 1.csv)	Mass flow rate LiW feeders	Kilogram/hour
%PD1 to %PD3	Z-AB (LiW Feeders 1.csv)	Compositional mass flow rate LiW feeders PD1–PD3	Percentage
%PD4 to %PD7	A-C (LiW Feeders 2.csv)	Compositional mass flow rate LiW feeders PD4–PD5 and PD7	
Estimated weight IBC PS1–PS7	D-I (LiW Feeders 2.csv)	Estimated weight intermediate bulk containers LiW feeders PD1–PD5 and PD7	Kilogram
RefAct PD1–PD7	J-O (LiW Feeders 2.csv)	Binary variable indicating whether refill IBC is in progress at the current moment (1) or not (0)	{0,1}
Net Weight PD1–PD7	P-U (LiW Feeders 2.csv)	Weight of the material inside the LiW feeder	Kilogram
Totalizer PD1–PD7	V-AA (LiW Feeders 2.csv)	Cumulative amount of material fed over time	Kilogram

Table 5. Target Composition and Alarm Limits of Feeders

Feeder	PD1	PD2	PD3	PD4	PD5	PD7
Target composition (%)	28.50	29.00	10.00	28.50	3.00	1.00
Allowed relative deviation (+/− %)	10.00	10.00	15.00	10.00	5.00	5.00

could deliver per screw rotation. Combined with the actual screw speed (Screw RPM PD1−PD7), it enables the system to calculate the theoretical mass flow of the component delivered to the system for each of the feeders[1].

We note that LiW feeders can operate in either gravitational or volumetric mode. The default in our setting is gravitational mode; that is, the LiW feeder is placed on a scale (see Online Appendix 3.3), and hence, the mass flow is determined by measuring the decrease in material weight over time. However, sometimes, the scale could be thrown off because of physical disturbances such as vibrations or external circumstances. In this case, the feeder switches from gravitational mode to volumetric mode, indicated by VolMode PD1–PD7 in Table 4. When switched to the volumetric mode, the theoretical feed rate is calculated using the feed factor multiplied by the screw speed.

The actual mass flow rate of materials at the individual LiW feeders is measured as the loss in weight over time (Massflow PD1–PD7). The percentage compositional mass flow (%PD1 to %PD7) is calculated by dividing the actual mass flow of each feeder by the total mass flow from all feeders. If this value falls outside the specified maximum compositional deviation, an alarm is triggered. See Table 5 for the target composition and corresponding alarm limits (i.e., allowed relative deviations from the target) for each feeder. Recall that any critical process alarm triggered in the system results in a 100% OOS for the duration of the alarm.

The estimated weight IBC represents the estimated weight of the intermediate bulk containers (IBC). These are containers placed on top of the feeders containing the bulk amount of the raw materials to be dispensed by the LiW feeders. They act as a buffer, and their estimated weight is measured to ensure that the bulk of the raw materials is replenished in time. During a refill sequence, the corresponding binary variable RefAct for PD1–PD7 changes from zero to one. Net Weight PD1–PD7 represents the weight of the material in the hopper and the screw itself (i.e., the total mass of the system minus the mass of the equipment). The cumulative amount being fed over time is recorded for each feeder as the Totalizer PD1–PD7.

Table 6 shows the parameters for the blending operations at Blender 1 and 2. Massflow Blender 1 and 2 describes the amount of material passing through the blender per time unit. In steady-state operation (i.e., after start-up and before shutdown), the mass flow of Blender 1 should be equal to the total input of LiW feeders PD1–PD5, and the mass flow of Blender 2 should be equal to the mass outflow of Blender 1 plus the inflow of feeder PD7. In addition, the blend potency and the OOS concentration for both blenders are measured. The blend potency is measured as the relative percentage of the target concentration of the active component in the blend. Therefore, a value of 100% means that the amount of active components in the blend is exactly at the desired setpoint concentration, as specified in Table 5.

The process parameters corresponding to the rotary tablet press are presented in Table 7. Online Appendix 3.4 provides additional information on the tablet press principles and naming conventions. The tablet press machine data consist mainly of parameters related to precompression and main compression. During precompression, the bottom punch moves while the upper punch remains in a fixed position to compress the incoming powder mixture into a tablet. Therefore, the parameter "precompression height bottom" (and "main compression height bottom") measures the height of the bottom punch during the precompression (and main compression, respectively). The top dwell time (i.e., see precompression and main compression top dwell times

Table 6. Inline Process Parameters of Continuous Blenders B1 and B2, File 'Blenders.csv'

Column name	Corresponding columns	Description	Unit of measure
Massflow Blender 1 and Blender 2	B-C	Mass flow rate through the blender	Kilogram/hour
Blend Potency Blender 1 and Blender 2	D-E	Blend potency measured by RTD	Relative percentage
OOS Concentration at Blender 1 and Blender 2 inlet	F-G	Fraction of the blend containing material that has been subjected to an OOS	Percentage

Table 7. Inline Process Parameter Tablet Press, File 'Tablet Press.csv'

Column name	Corresponding columns	Description	Unit of measure
Precompression height bottom	B	Height of bottom punch during precompressions	Millimeter
Precompression top dwell time	C	Time powder/tablet is held under compression	Millisecond
Precompression force	D	Average force tablet is under during precompression	Kilonewton
Precompression displacement top sigma	E	Variability in precompression displacement	Percentage
Main compression height bottom	F	Height of bottom punch during compressions	Millimeter
Main compression top dwell time	G	Time powder/tablet is held under compression.	Millisecond
Main compression force	H	Average force tablet is under during the main compression	Kilonewton
Compression cycle fill depth	I	Fill depth of the punch when allowing powder into the dies of the tablet press	Millimeter
Filling Shoe M20M13 speed	J	Feeder paddle 1 speed	Rotations per minute
Filling Shoe M20M23 speed	K	Feeder paddle 2 speed	Rotations per minute
Material inlet: Hopper level detection	L	Height of powder in level sensor assembly above tablet press	Percentage
Ejection force tablet	M	Ejection force of tablet from the die after compression	Kilonewton

in Table 7) denotes the time the powder mixture is under the compression force while being transformed into a tablet. The parameter "precompression force" (and "main compression force") denotes the average force applied to the tablet during precompression (and main compression, respectively). The parameter "precompression displacement top sigma" represents the variability in precompression displacement (i.e., precompression compared with its starting position) and relates to the variability of tablet weight via the height of the compression roll.

The remaining parameters in Table 7 are related to the materials entering and leaving the rotary tablet press. The parameter named "compression cycle fill depth" measures the fill depth of the punch when powder enters the dies. A feeder paddle evenly distributes the powder into the die cavities to form tablets. It is located in the feed frame section and consists of a rotating paddle to uniformly distribute the material from the hopper to the feed frame. Hence, the speed of the feeder paddles 1 and 2 is captured by "Filling Shoe M20M13" and "Filling Shoe M20M23" speed. The fill depth and paddle speed influence the weight and weight variability of the tablets. To ensure that enough material is available for tablet compression, the fill level of the tablet press hopper is measured by "Material Inlet: Hopper Level Detection." After being compressed, tablets are freed from the die by applying an ejection force. Hereafter, tablets are either sent to coating, the tablet analyzer, or rejected based on the critical quality attributes.

In addition to the inline parameters described above, the relative humidity and the temperature were measured at the feeders' section and the tablet press (see the files Humidity.csv and Temperature.csv).

4.4. Contextual Quality Data

As part of the original 120-hour run study, we collected contextual quality data to assess the effects of raw material changes during the run. The contextual quality data consist of several Excel files, as summarized in Table 3. We first provide some background information (Section 4.4.1) and then describe the data (Section 4.4.2).

4.4.1. Background Information on Contextual Quality Data. The contextual quality data can be used to understand the impact of raw materials on production. In this context, we use the term "lot" to represent a large-scale production run (e.g., it could represent a certain amount of tablet production, such as one ton, or a certain amount of raw material usage, such as 10 tons of Avicel). We use the terms "bag" and "drum" interchangeably, which represent a smaller unit of a particular lot. Thus, a lot may be divided into several smaller bags/drums for material handling or other purposes. In our setting, the effects of raw materials occurred in two ways:

1. Intra-lot changes in raw materials: Changes between individual drums/bags of raw materials produced in the same batch by the material supplier.

2. Inter-lot changes in raw materials: Changes between different batches of the same raw material.

Table 8. Materials Used in the Study

Material identifier	Formal name	Lot no.		LiW feeder
pAPAP	Acetaminophen powder, APAP_P	85170761		PD3
L316	Spray dried lactose, Fast Flo '316'	637517L052		PD2
A102	Microcrystalline cellulose, Avicel PH102	71734C	71732C	PD1 and PD4
Ac-di-sol	Sodium Croscarmellose, Ac-di-sol SD-711	T1726C		PD5
Mag St	Magnesium stearate, Ligamed MF-2-V	31400041		PD7

Powder-grade acetaminophen ('APAP_P'; Mallinckrodt, St. Louis, MO) was selected as the model active pharmaceutical ingredient (API). Microcrystalline cellulose (Avicel PH102; FMC Biopolymer, Cork, Ireland), spray dried lactose (Fast Flo 316, Kerry, Naas, Ireland), sodium croscarmellose (Ac-di-sol SD-711; FMC Biopolymer, Cork, Ireland), and magnesium stearate (Ligamed MF-2-V; Peter Greven, Venlo, The Netherlands) made up the rest of the formulation used during the study.

To study the inter-lot effects of material variation, two different lots of Avicel PH102 were supplied, whereas for all other materials a single lot/batch of the material was used. This would allow for a detailed study of the intra-lot variation for all materials, whereas the inter-lot variation would be assessed using the Avicel PH102. The details of the individual materials used in the study are summarized in Table 8.

4.4.2. Description of the Contextual Quality Data. The file named "RM Content Uniformity" reports the drug content (i.e., the percentage of API per tablet) for each tablet sampled. We sampled 10 tablets every 30 minutes (unless there was a pause in the process). The data were collected at line upon the completion of the tablet press operation. The upper and lower control limits for the drug content can be found in Table 5 (column PD3).

The file named "RM Tablet Properties and Drum Changes" consists of two spreadsheets. The first spreadsheet (named "tablet properties") reports the average weight (milligrams), average thickness (millimeters), average hardness (kiloponds), and the corresponding standard deviations (SD) based on 10 tablets collected per sample. These samples were collected at the tablet press and represent the same samples as those reported in the RM Content Uniformity file. See Holman et al. (2021) for the corresponding control limits. The information on tablet weight, thickness, and hardness is used to make a control response in the system (i.e., see the illustration of the feedback loop between the tablet analyzer and tablet press in Figure 1). For example, if the tablet is too hard, then the system will lower the compression force.[2] The column labeled "entered into the system" denotes the time when the tablet information was entered into the system to generate a control response.

The second spreadsheet (named "raw material drum change") provides refill information associated with each raw material. The file considers six raw materials (e.g., API, magnesium stearate; see Section 4.4.1 for more information on raw materials). Each of these raw materials was fed into a specific LiW feeder (see Table 8). For each raw material, the column "date/time" denotes the specific time when the raw material was refilled at the LiW feeder, whereas the column "amount" indicates the specific refill quantity in kilograms. The estimated refill times of each raw material are also reported (see Holman et al. 2021 for details).

Table 9 provides an overview and description of the data presented in the file named "RM Material Properties." Where required, materials have been characterized using an FT4 powder rheometer (Freeman Technology 2023). The following standard experiments were run according to the standard operating procedures of the FT4: stability and variable flow rate and compressibility and shear cell at 6 kPa. Further details on the description of material characterization are recorded in the literature (Freeman Technology 2023).

Lastly, we provide an optional Excel file (named "Logbook Long Run Days") as an Online Appendix. This file consists of six spreadsheets, where each spreadsheet represents a production day starting from January 12, 2018, until January 17, 2018. For each day, the file contains the notes of the operators on important events for each of the operations (feeding, blending, compression, etc.). For example, on January 12 at 10 a.m., an alarm resulted in a system shutdown. For brevity, we omit the details because the notes are self-explanatory. This additional information may help contextualize some unexpected events or patterns that may be observed in the data.

4.5. Research Questions

In this section, we propose specific research questions that can be answered using the data set. In Section 5, we present additional research questions to support generic applications of CM and RTRT in both pharmaceutical and biopharmaceutical manufacturing.

- What real-time data/information can be gained from the LiW feeder signals that relate to changes in powder properties beyond mass flow?
- Can the system be used to monitor changes in raw materials (i.e., change of drum or lot) using the LiW feeder signals?

Table 9. Material Property Data, File *'RM Material Properties.xlsx'*

Name	Units	Description
Conditioned bulk density (CBD)	g/ml	The bulk density of the material after standard preconditioning
Basic flow energy (BFE)	MJ	A measure of the powder's flowability when forced to flow, that is, through a feeder
Compressibility (CPS) % at stated pressure	%	Change in volume of a powder form under zero pressure to the defined test pressure
Cohesion	kPa	The force between particles that hold the powder together
Unconfined yield stress (UYS)	kPa	The stress necessary to cause a material unsupported in two directions to fail in shear
Major principal stress (MPS)	kPa	The stress necessary to cause a material unsupported in two directions to fail in shear under a pre-consolidated state
Flow function (FF)	—	The ratio of MPS to UYS and is used to rank flowability. Flow Function is a parameter commonly used to rank flowability, with values below 4 denoting poor flow and above 10, good flow.
Angle of internal friction (AIF)	°	A measurement of a powder unit's resistance to shear stress is known as the internal angle of friction. When failure happens only in response to a shearing stress, the angle measured between the normal force and resultant force is the angle.

- How do changes in the raw material attributes (i.e., change of Avicel PH102 lot) impact the downstream process (i.e., compression), and can the change be predicted?
- How can the mass of data be simplified to have an easy measure of process performance during feeding with respect to top-up size, powder density, and feeding performance?
- What is the effect of a temporary disturbance (e.g., shutdowns, OOS) on the system? How to predict the disturbances in advance and take preventive actions accordingly?
- How to adjust loss-in-weight feeders to control quality and throughput? What are the optimal raw material feeding strategies to increase throughput and quality?
- What are optimal mixing and blending decisions based on our data? How fast can we run a CM system to maximize throughput while maintaining quality?
- How to develop good prediction models for RTD and use these prediction models for real-time optimal control of CM or RTRT systems?

5. Future Research Directions for OM

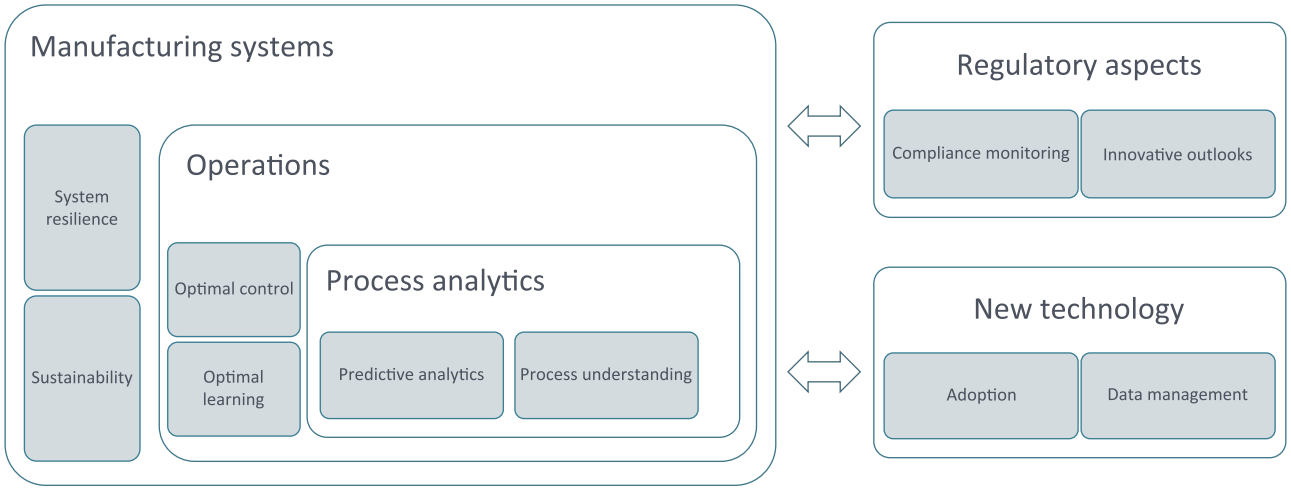
In this section, we propose several research questions for the OM community, with a specific focus on CM and RTRT. We present research questions that apply to both pharmaceutical and biopharmaceutical manufacturing. For ease of exposition, we develop a roadmap that systematically categorizes future research directions in Figure 2. Our roadmap begins with process analytics (Section 5.1). We then focus on operational aspects related to optimal control (Section 5.2) and optimal learning (Section 5.3). We consider the dynamics of manufacturing systems for resilient and sustainable production (Section 5.4). Lastly, we elaborate on regulatory aspects (Section 5.5) and new technology adoption (Section 5.6). We conclude each topic with open research questions.

5.1. Process Analytics

The OM community can help the industry develop a better understanding of biological and chemical processes based on production data. In life sciences, some process dynamics are intuitive and well-understood (e.g., cells start to die if they are underfed during fermentation). However, it is more challenging to understand the cross-functional relationships between multiple process parameters and quality attributes because these relationships cannot always be explained by domain knowledge or intuition. On the other hand, recent advances in sensors, such as Raman spectroscopy, have enabled the collection of new process data that were previously difficult to collect. When combined with data analytics and OM, these new measurements can help us develop a better understanding of the production processes. Another critical challenge for the industry is identifying the right type and amount of data to collect and translating the data into explainable models. Motivated by these industry needs, the OM community can help address several open questions on process analytics:

- What data do we need to collect to better understand and control these processes? How can we identify the key process parameters from a given data set? How can we develop new data- and model-driven approaches to determine the golden profile?

Figure 2. (Color online) An OM Research Roadmap to Support CM and RTRT Applications



• What data are sufficient for regulators to evaluate the golden profile? What data, model, or analysis is needed to statistically demonstrate the (robustness of the) golden profile?

Another line of research is related to predictive analytics. State-of-the-art approaches in life sciences typically use mechanistic models (also known as kinetic models) to model and predict process dynamics. Examples of mechanistic models include Monod equations and ordinary differential equations (McNeil and Harvey 2008, Putra and Abasaeed 2018). As a major advantage, mechanistic models are well-established in the field and are easy to explain. However, mechanistic models have limitations because they are typically deterministic and require prespecified input parameters. Therefore, they cannot fully capture the inherent stochasticity of these processes, nor can they learn from real-time data. Although general-purpose approaches such as reinforcement learning can partially alleviate this problem, these black-box models do not incorporate domain knowledge from the life sciences and may require significant amounts of training data that may not be available. As a result, black-box models may be difficult to explain to regulators and may also contain model uncertainty. Therefore, we propose the following scientific challenges to the OM community:

• How can we combine data-driven methods with mechanistic models from life sciences in such a way that the resulting predictive models are explainable yet adequately equipped to represent these complex systems?

• How can the stochastic input/output relationships between the key process parameters and quality attributes be established? What is the impact of input material attributes and their variability on continuous processing and quality? How can we discover new patterns in our data to identify parameters (or soft sensors) needed to control quality and yield?

5.2. Operations: Optimal Control

R & D studies tend to focus on the underlying biological and chemical aspects but often overlook operational risks and trade-offs. In daily manufacturing practice, however, the industry faces multiple operational risks, such as failure risks, and variability in processes and raw materials. Therefore, "R & D optimal" (defined based on biology and chemistry) may not always align with "manufacturing optimal" (defined based on operational risks and trade-offs) in practice. To bridge this gap, the OM community can help develop new approaches for optimal control mechanisms by linking the underlying process dynamics (i.e., predictive models) with operational aspects (i.e., prescriptive analytics). We believe that a systematic approach combining life sciences, data sciences, and OM is necessary to support the industry's transition toward "smart" manufacturing to reduce costs and lead times. Our Franz Edelman work provides one of the first attempts at OM applications in biomanufacturing (Martagan et al. 2023). Nevertheless, the industry offers a wide range of optimal control problems, as described below.

5.2.1. Maintain the State of Control. Real-time measurements can already help identify when processes start to deviate from their golden profile or exceed their predefined specification limits. However, there is still a need for optimal control mechanisms to bring processes back to their golden profile in real time and to maintain a state of control over long periods of time. These optimal control mechanisms can be based on data- and model-driven approaches, as described

earlier, and should be equipped to capture risk-benefit trade-offs under process uncertainty. In addition, the OM community can help identify what process data should be presented to the regulators to demonstrate and validate the state of control over lime. To summarize:

- What data need to be collected and analyzed to ensure the state of control? Where to put sensors and how to eliminate data redundancies?
- How do the lapses in data collection (e.g., recalibrating sensors) affect product quality and related decisions?
- How to maintain a state of control over long periods of time in CM and RTRT? How to bring the process back to its golden profile when it deviates from it? What are the most effective methods for real-time process monitoring and control in CM and RTRT to maintain quality and prevent deviations from the golden profile?
- How to develop new models based on data- and OM-driven approaches to detect disturbances and/or variability in the system and make real-time preventive (or corrective) interventions accordingly? How can advanced data analytics and machine-learning techniques be used to improve process robustness and reduce variability in CM and RTRT systems?

5.2.2. Optimal Release Policies. The industry aims to transition from performing time-consuming release tests to autonomous RTRT in the future. To facilitate this transition, we can work with the concept of "release by exception." This concept implies a mixed strategy where either an RTRT or an offline release test is performed based on certain prespecified process conditions (as opposed to an all-or-nothing approach to release testing). However, the concept of release by exception is relatively new, and we need a better understanding of the associated risks and operational aspects. This line of research could be built on condition-based manufacturing and maintenance optimization (De Jonge and Scarf 2020) and address the following questions:

- Which conditions should a system meet to move toward fully autonomous RTRT?
- How can predictive maintenance strategies be applied in CM and RTRT systems to minimize downtime and increase system reliability? Under which conditions (on parameters, prediction errors, costs, etc.) it is optimal to release in real-time instead of at-line/offline inspections?

5.2.3. Optimal Stopping Problems. Our previous work presents some examples of optimal stopping problems in a batch and semicontinuous biomanufacturing setting (see, e.g., Martagan et al. 2020, Koca et al. 2023, Wang et al. 2023). Future research can develop stochastic optimization models specifically tailored to the needs of CM and RTRT systems:

- What is the longest duration we can run a CM system while maintaining quality?
- When to stop and restart a specific unit operation (e.g., continuous blending) based on operational risks, trade-offs, and the stochastic evolution of the process parameters and quality attributes over time?
- What is an optimal campaign duration (e.g., when to stop a campaign in real-time) based on raw material properties, quality, cost, and risk considerations?

5.2.4. Modeling and Optimization of Continuous Fermentation Systems. Although there are a few successful implementations of continuous fermentation technologies in practice (e.g., continuous stir tank reactors), there is still significant room to optimize the performance of continuous fermentation systems. Because our data set does not involve fermentation, we do not list specific research questions on this topic to prevent confusion. However, future OM research on continuous fermentation systems can address decisions related to feeding rates (i.e., feeding "too much" or "too little" has an impact on yield and quality), medium exchange rates (i.e., the amount of fresh medium supplied to the system and the amount of medium extracted from the system have an impact on yield and quality), starting conditions (i.e., the amount of seed culture and the properties of the media used affect the yield and quality), and bleeding decisions (i.e., when to remove cells from the bioreactor).

5.2.5. Traceability Strategies. Future research can develop new approaches to facilitate process traceability in case of unforeseen disturbances and quality issues. These traceability strategies should be based on a combination of sensor measurements, prediction models such as RTD, and risk-benefit trade-offs. In addition, sampling strategies should be defined to ensure product quality and safety in CM and RTRT. In summary:

- How to capture the impact of a disturbance as it propagates in a CM/RTRT system? What is the impact of changes made at a unit operation on subsequent operations and output quality?
- How far to trace back (and scrap the produced materials) based on potential risks and costs?
- What is an optimal sampling strategy (e.g., sample location, sample size, frequency, statistical approach and criteria, and their relevance to the intended use) to ensure safety and quality in CM and RTRT?

5.2.6. Model Maintenance and Validation. The role of model maintenance and validation becomes increasingly important as the industry moves toward automated decisions. The OM community can help the industry improve, train, and validate models using real-time data, advanced analytics, and PAT tools. New OM-based approaches can

enable timely and efficient communication between the industry and the regulators:

- When and how often should the models be revised and improved based on real-time data? What is the best approach to the maintenance and validation of models?
- How to demonstrate the appropriateness and robustness of the developed models and control strategies to the regulatory authorities?

5.3. Operations: Optimal Learning

Despite advances in sensor technologies, decision making with "small" (limited) data continues to be an industry concern. In commercial production settings, challenges related to small data arise every time there is a change in equipment or raw materials. For instance, if the raw material supplier changes their formulation or new equipment is installed, the process output is significantly affected, making historical process data unreliable or obsolete. Small data can also be a concern in small-scale research and development projects. Therefore, all research questions described in this paper can be revisited to support decision-making with small data. The OM community has already developed methodologies to support decision making with small data (see, e.g., Mišić and Perakis 2020). These methodologies can be applied in CM and RTRT settings.

Bayesian learning or other machine-learning techniques can be used to help the industry make better use of newly collected data over time. The concept of learning by doing is particularly important in CM, because these systems may start with small amounts of data (because of variations in campaigns, raw materials, and lots) but can generate large data sets in a relatively short time. Furthermore, Bayesian learning can help the industry identify the best operating conditions within a certain predefined operating range (so-called design space defined by R & D). For instance, in our previous work, we employed a Bayesian design of experiment approach to effectively search for the best bioreactor operating configuration using a limited number of industry-scale experiments (Martagan et al. 2023). This project helped us identify a better bioreactor operating condition with a 50% increase in our bioreactor yield. Similar to these initiatives, future research on optimal learning applications can help the industry bridge the gap between "optimal" operating policies defined by R & D (based on biology and chemistry) and manufacturing (based on risk-benefit trade-offs) by effectively searching the feasible design space. Additionally, optimal learning methodologies can assist in identifying OOS signals or hidden failures (such as worsening spectroscopy measurements) to support preventive and corrective inferences. In summary, we propose the following research questions:

- How to make better CM decisions with small data? How to develop good prediction models to support scaleups (i.e., transition from laboratory scale to commercial scale) using small data? What are the key factors influencing the technology transfer and scalability of CM processes from laboratory to commercial scale production?
- By using Bayesian learning or other machine-learning approaches, how can we reduce the amount of data needed to maintain the state of control? How can we reduce the amount of data provided to the regulators to ensure the state of control?
- How to define a new "optimal" for manufacturing operations by effectively searching (and even expanding) the feasible design space defined by R & D?

5.4. Manufacturing Systems

Many industries such as automotive and high-tech have benefited from Industry 4.0 and 5.0 initiatives on automation, digital twins, and the Internet of Things (IoT). However, the (bio)pharmaceutical industry still presents a significant opportunity for Industry 4.0 and 5.0 initiatives to support cost-effectiveness and resilience. For a successful implementation of CM and RTRT, it is essential to ensure the system's ability to maintain an integrated flow under uncertainty. Moreover, these systems should be resilient to disruptions to ensure safety and quality. The industry also seeks to effectively integrate automated process technologies with human expertise and real-time data to achieve resilience. There is also a growing emphasis on flexibility and sustainability with the development of new and modular equipment (e.g., single-use reactors). Therefore, we foresee several opportunities to learn from other industries and implement Industry 4.0 and 5.0 initiatives in CM and RTRT settings:

- How can digital twin, IoT, and virtual reality technologies be utilized to optimize process design and operator training in continuous manufacturing?
- How can risk assessments and quality risk management tools be effectively integrated into CM and RTRT processes to ensure product safety and efficacy?
- What are the environmental sustainability benefits and challenges associated with CM/RTRT, and how can these be quantified and optimized?
- What are the economic implications of transitioning from batch manufacturing to CM in terms of capital investment, operational costs, and overall profitability?
- How much buffer on lead time and capacity is needed in CM/RTRT systems to ensure flexibility and robustness to unforeseen disturbances?
- How to develop smart scheduling and production planning approaches tailored to the specific needs of CM and RTRT (e.g., staffing decisions, multi-products, limited shelf life, changeovers, and inventory aspects)? How to move from reactive to proactive actions (i.e.,

prevent OOS signal and, consequently, diversion of products)?

5.5. Regulatory Aspects

The use of data analytics and AI is becoming increasingly important to regulatory authorities and the industry (Wuest et al. 2020, FDA 2023, Garguilo 2023). To encourage further discussion on the use of AI in drug manufacturing, regulators have recently published several white papers and requested public feedback. For example, the Food and Drug Administration published an exploratory paper and proposed a regulatory framework for AI-based software as a medical device (FDA 2019). The European Medicines Agency explored similar AI-driven initiatives (EMA 2020). More recently, in 2023, the Center for Drug Evaluation and Research published a discussion paper seeking public feedback on the use of AI in drug manufacturing (Center for Drug Evaluation and Research 2023). We believe that the input of the OM community can also play an important role in shaping new regulatory frameworks and policy developments. Another line of research can investigate new governance mechanisms to ensure the ethical use of AI in drug manufacturing. Moreover, future research can explore new approaches to facilitate compliance monitoring and collaboration between regulators and the industry. In summary, we propose the following questions:

- What are the regulatory challenges and considerations for implementing CM and RTRT in different countries and regions?
- How can quality assurance systems and validation approaches be adapted to meet the unique requirements of CM and RTRT?
- What data are sufficient for regulators to evaluate CM and RTRT systems? What data need to be provided for timely and effective communication? How can the regulators increase their efficiency by using AI and automation to analyze large volumes of data?
- How to define a "batch" in a CM setting?
- How to develop new regulatory frameworks to address the specific needs of CM and RTRT? What is the potential impact of new regulatory policies on the industry and social welfare?

5.6. New Technology Adoption

The industry is undergoing a digital transformation with the development of new sensors, manufacturing technologies, and advanced IT infrastructures. To stay ahead of the competition, manufacturers should be able to collect, analyze, and systematically store large amounts of (real-time) data and also effectively share data across different functions in their value chain. In addition, organizational leadership plays an important role in the adoption of new technologies. In current practice, flowsheet models are generally used to assess the performance of new CM technologies (see, e.g., Wang et al. 2017). As a future research direction, OM-driven approaches and new analytical models could be developed to support new technology adoption decisions under market, regulatory, and technology uncertainties. Another future line of research could focus on data collection and management strategies tailored to the needs of CM and RTRT systems:

- How can the business case be made for different levels of CM or RTRT technologies? What are the requisite conditions (i.e., market, technology, regulatory considerations) to deploy CM or RTRT technologies?
- Where should we place the sensors? How do we ensure that we are collecting the right data at the right time and location to support CM and RTRT? How to set up the right control system to make real-time adjustments based on data? Where should real-time data be stored? How should on- or inline data be evaluated during CM and RTRT?

6. Conclusions

The 2024 MSOM data-driven research challenge is a collaborative effort of MSD's Human and Animal Health divisions and GEA. We provide 120 hours of production data on a continuous pharmaceutical tablet manufacturing process. We also propose several research questions and future research directions based on common industry challenges and opportunities. The data set can be used to develop data- or model-driven approaches to support successful implementations of CM and RTRT technologies. The data can also be used to generate new teaching materials.

The applications of OM methodologies created a significant impact on industries such as automotive, retail, and semiconductor manufacturing. However, there are still a limited number of papers and written accounts of success related to OM applications in (bio)pharmaceutical manufacturing. With this data-driven research challenge initiative, we believe that new knowledge generated by the OM community will facilitate the industry's transition toward data- and OM-driven decision making.

Acknowledgments

The authors are grateful for the support of the editors of the MSOM practice platform, Jérémie Gallien, Pavithra Harsha and Ananth Raman, and the editor-in-chief, Georgia Perakis. The authors also thank our extended team at MSD (Sebastian Escotet, Stephen Ma, and Annick Nusselder) for their contributions and valuable comments. This work was also supported by the Dutch Science Foundation, NWO-VIDI grant.

Endnotes

[1] Note that theoretical mass flow (kg/h) = feed factor (g/revolutions) × screw speed (revolutions/min) × 60 (min/h) × 1/1,000 (kg/g).

[2] The specific control algorithm used in our system is proprietary and not disclosed.

References

Balfour H (2021) Continuous biomanufacturing – Where are we now? European Pharmaceutical Review. Accessed November 23, 2023, https://www.europeanpharmaceuticalreview.com/article/164036/continuous-biomanufacturing-where-are-we-now/.

BioPhorum Operations Group (2022) Biomanufacturing technology roadmap. Accessed November 21, 2023, https://www.biophorum.com/wp-content/uploads/bp_downloads/Overview-1.pdf.

Center for Drug Evaluation and Research, The U.S. Food and Drug Administration (2023) Artificial intelligence in drug manufacturing. Accessed September 29, 2023, https://www.fda.gov/media/165743/download?attachment.

European Medicines Agency (2012) Guideline on real time release testing (formerly Guideline on parametric release). Accessed November 22, 2023, https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-real-time-release-testing-formerly-guideline-parametric-release-revision-1_en.pdf.

De Jonge B, Scarf PA (2020) A review on maintenance optimization. *European J. Oper. Res.* 285(3):805–824.

European Medicines Agency (2020) EMA regulatory science to 2025 strategic reflection. Accessed October 1, 2023, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf.

Food and Drug Administration (2019) Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML) based software as a medical device (SaMD) – Discussion paper and request for feedback. Accessed September 30, 2023, https://www.regulations.gov/document/FDA-2019-N-1185-0001.

Food and Drug Administration (2023) Artificial intelligence and machine learning (AI/ML) for drug development. Accessed October 1, 2023, https://www.fda.gov/science-research/science-and-research-special-topics/artificial-intelligence-and-machine-learning-aiml-drug-development.

Freeman Technology (2023) Powder flow testing with the FT4 powder rheometer. Accessed December 1, 2023, https://www.freemantech.co.uk/powder-testing/ft4-powder-rheometer-powder-flow-tester.

Garguilo L (2023) AI beats back bad batches. Outsource pharma. Accessed October 1, 2023, https://www.outsourcedpharma.com/doc/ai-beats-back-bad-batches-0001.

GEA Pharma and Healthcare (2021) ConsiGma® CDC-LB Line, continuous direct compression with linear blending. Accessed November 28, 2023, https://www.youtube.com/watch?v=nOLwUhQlAGc.

Giaquinto M, Auchincloss K (2022) US Congress passes bill supporting new era of biopharma advanced & continuous manufacturing. Bioprocess online. Accessed November 23, 2023, https://www.bioprocessonline.com/doc/us-congress-passes-bill-supporting-new-era-of-biopharma-advanced-continuous-manufacturing-0001.

Globerman S, ed. (2016) Intellectual property rights and the promotion of biologics, medical devices, and trade in pharmaceuticals. Fraser Institute. Accessed November 21, 2023, http://www.fraserinstitute.org.

Grand View Research (2022) Animal health market size, share & trends analysis report by animal type (companion), by product (biologics), by distribution channel, by end-use, by region, and segment forecasts, 2023–2030. Accessed November 22,2023, https://www.grandviewresearch.com/industry-analysis/animal-health-market.

Holman J, Tantuccio A, Palmer J, van Doninck T, Meyer R (2021) A very boring 120 h: 15 million tablets under a continuous state of control. *Powder Technol.* 382:208–231.

ICH Q8, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2017) ICH guideline Q8 (R2) on pharmaceutical development. Accessed November 23, 2023, https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human-use_en-11.pdf.

ICH Q10, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2015) ICH guideline Q10 on pharmaceutical quality system. Accessed November 23, 2023, https://www.ema.europa.eu/en/documents/scientific-guideline/international-conference-harmonisation-technical-requirements-registration-pharmaceuticals-human_en.pdf.

ICH Q13, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2023) ICH guideline Q13 on continuous manufacturing of drug substances and drug products. Accessed November 21, 2023, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-q13-continuous-manufacturing-drug-substances-drug-products-step-5_en.pdf.

Jiang M, Severson KA, Love JC, Madden H, Swann P, Zang L, Braatz RD (2017) Opportunities and challenges of real-time release testing in biopharmaceutical manufacturing. *Biotechnol. Bioeng.* 114(11):2445–2456.

Koca Y, Martagan T, Adan I, Maillart L, van Ravenstein B (2023) Increasing biomanufacturing yield with bleed–feed: Optimal policies and insights. *Manufacturing. Service Oper. Management* 25(1):108–125.

Lee S (2017) Modernizing the way drugs are made: A transition to continuous manufacturing. The U.S. Food and Drug Administration. Accessed November 23, 2023, https://www.fda.gov/drugs/news-events-human-drugs/modernizing-way-drugs-are-made-transition-continuous-manufacturing.

Markl D, Warman M, Dumarey M, Bergman EL, Folestad S, Shi Z, Manley LF, Goodwin DJ, Zeitler JA (2020) Review of real-time release testing of pharmaceutical tablets: State-of-the art, challenges and future perspective. *Internat. J. Pharm.* 582:119353.

Martagan T, Krishnamurthy A, Leland PA (2020) Managing trade-offs in protein manufacturing: How much to waste? *Manufacturing Service Oper. Management* 22(2):330–345.

Martagan T, Adan I, Baaijens M, Dirckx C, Repping O, van Ravenstein B, Yegneswaran PK (2023) Merck animal health uses operations research methods to transform biomanufacturing productivity for lifesaving medicines. *INFORMS J. Appl. Analytics* 53(1):85–95.

McNeil B, Harvey L, eds. (2008) *Practical Fermentation Technology* (John Wiley & Sons, West Sussex, UK).

Mišić VV, Perakis G (2020) Data analytics in operations management: A review. *Manufacturing Service Oper. Management* 22(1):158–169.

NIIMBL, The National Institute for Innovation in Manufacturing Biopharmaceuticals (2023) NIIMBL-led programs and projects. Accessed November 22, 2023, https://niimbl.my.site.com/s/niimbl-led-programs.

Putra MD, Abasaeed AE (2018) A more generalized kinetic model for binary substrate fermentations. *Process Biochem.* 75:31–38.

Repping O, Martagan T, van Ravenstein B, Dirckx C, Nusselder A, Baaijens M (2023) Applications of OR/MS methodologies in biomanufacturing: Our vision and future research directions. M&SOM Practice Platform Seminar Series. Accessed November 21, 2023, https://pubsonline.informs.org/page/msom/biomanufacturing-2-3-2023.

Skyquest (2023) Global pharmaceuticals market. Accessed November 22, 2023, https://www.skyquestt.com/report/pharmaceuticals-market#:~:text=Pharmaceuticals%20Market%20size%20was%20valued,is%20used% 20for%20medical%20purposes.

Wang Z, Escotet-Espinoza MS, Ierapetritou M (2017) Process analysis and optimization of continuous pharmaceutical manufacturing using flowsheet models. *Comput. Chem. Eng.* 107:77–91.

Wang B, Xie W, Martagan T, Akcay A, van Ravenstein B (2023) Optimizing biomanufacturing harvesting decisions under limited historical data. Preprint, submitted January 11, 2021, https://arxiv.org/abs/2101.03735.

Wuest T, Kusiak A, Dai T, Tayur SR (2020) Impact of COVID-19 on manufacturing and supply networks—The case for AI-inspired digital transformation. Preprint, submitted May 5, http://dx.doi.org/10.2139/ssrn.3593540.