IISE Transactions

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Mind the gap between research and practice in operations management

Xiaohong Chen, Tianhu Deng, Zuo-Jun Max Shen & Yi Yu

To cite this article: Xiaohong Chen, Tianhu Deng, Zuo-Jun Max Shen & Yi Yu (2023) Mind the gap between research and practice in operations management, IISE Transactions, 55:1, 32-42, DOI: 10.1080/24725854.2022.2046893

To link to this article: https://doi.org/10.1080/24725854.2022.2046893

Published online: 15 Apr 2022.

IISE TRANSACTIONS
2023, VOL. 55, NO. 1, 32–42
https://doi.org/10.1080/24725854.2022.2046893

Taylor & Francis
Taylor & Francis Group

Mind the gap between research and practice in operations management

Xiaohong Chen[a,b], Tianhu Deng[c], Zuo-Jun Max Shen[d,e], and Yi Yu[c]

[a]Hunan University of Technology and Business, Changsha, China; [b]Business School, Central South University, Changsha, China; [c]Department of Industrial Engineering, Tsinghua University, Beijing, China; [d]Faculty of Engineering and Faculty of Business and Economics, The University of Hong Kong, Hong Kong, China; [e]College of Engineering, University of California, Berkeley, CA, USA

ABSTRACT

The mission of the Institute of Industrial and Systems Engineers (IISE) is to serve those who solve complex and critical problems of the world. Notably, the research–practice gap in Operations Management (OM) marginalizes the value and relevance of the IISE. To maintain and enhance the impact of the IISE, we identify major bottlenecks that limit the industrial installation of OM research outcomes. Ranked by the relative importance, the three bottlenecks are verifying the performance improvement, building trust with practitioners, and balancing model accuracy and simplicity, respectively, in the stages of value verification, implementation and development. We propose potential research opportunities and illustrate the challenges and opportunities using real case studies from three Fortune Global 500 companies. In particular, we emphasize the role of data-driven decision methods in dealing with the three bottlenecks.

ARTICLE HISTORY
Received 10 November 2021
Accepted 23 February 2022

KEYWORDS
Operations management; research–practice gap; real-case studies; data-driven

1. Introduction

It is crucial for the Institute of Industrial and Systems Engineers (IISE) to guide Operations Management (OM) practitioners toward achieving OM practice excellence. The mission of the IISE is to serve OM practitioners who solve complex and critical problems encountered in the world, and often need to deal with complicated and integrated systems that suffer from customer dissatisfaction, excess inventory costs, and low product quality. Therefore, it is imperative for the IISE to capture the full potential of successful OM practice.

However, the academia that studies OM has generally been following, rather than leading, business practices (Simchi-Levi, 2014). The inability to lead OM practices is often due to numerous failed implementations of OM models. After many years of OM projects with industry collaborators, we cannot emphasize enough how important a successful implementation is for a research project. Further, we are astonished by the difficulties associated with ensuring an OM implementation with precise and unified standards across different departments. Companies have been appropriately cautious about new OM models because they cannot fully reap its benefits (Ibanez *et al.*, 2018; Sun *et al.*, 2021).

Why is it so challenging to ensure precise and unified implementations of OM research outcomes? A modern system integrates geographically dispersed firms whose decisions and performances are not directly observable and verifiable. In addition, intra-firm coordination is challenging because divergent responsibilities are assigned to various departments. Nonetheless, it is impossible to list all the reasons for failed OM implementations. Instead of answering "why", we focus on answering "how" to increase the likelihood of a successful OM implementation.

The primary goal of this article is to identify the major characteristics of OM research that can lead to more rapid and effective implementation. Three bottlenecks are discussed in the order of their importance to successful implementation. In particular, we start with how to verify the benefits in Section 2. Next, in Section 3, we proceed to discuss how to build trust between models and their users. Finally, we propose the trade-off between model simplicity and accuracy in Section 4. We emphasize that these three bottlenecks are interrelated. As illustrated in Figure 1, failure to verify performance improvement leads to lack of user trust. Using overcomplicated models and goals may increase the difficulty level of performance verification.

The impacts of data-driven decision methods on bridging the research–practice gap are discussed in Section 2 to Section 4. While traditional Industrial and Systems Engineering (ISE) research hits a plateau, the opportunities brought by big data are leading OM scholars to explore new sources of value. Meanwhile, there are more and more voices calling for strengthening the multidisciplinary team working between researchers from traditional ISE and other fields, e.g., statistics, computer science (Royston, 2013; Ranyard *et al.*, 2015; Johnson *et al.*, 2018), who have much to offer when it comes to leveraging data to improve business operations. As argued in Qi *et al.* (2020) and Wang *et al.* (2022), it is critical for future OM research to integrate data-driven decision methods, along with our own advantages in OM domain knowledge and optimization

CONTACT Zuo-Jun Max Shen maxshen@hku.hk

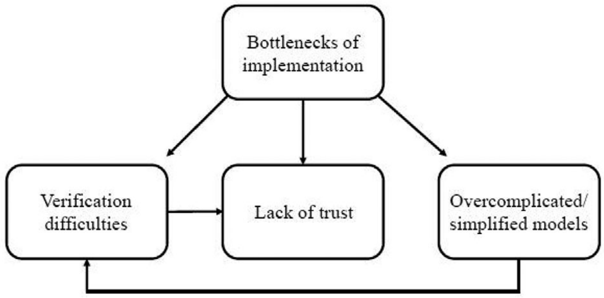


Figure 1. Bottlenecks in the successful implementation of OM research.

techniques, so that the methods can showcase higher levels of performance in domain-specific problems and the impact of ISE research in practice can be further enhanced.

2. Verifying the performance improvement

Implementing the results of OM research requires the approval of top-level managers who can take full responsibility for the associated costs and risks. Scientific and solid evidence must be presented to demonstrate the potential performance improvement. Unfortunately, verifying the outcomes of OM research (e.g., substantiating that the model has a satisfactory performance that is consistent with the expectations within its scope of application (Sargent, 2013)) is particularly challenging, which may lead to failed OM implementations.

2.1. Challenges

In a survey of supply chain simulation research, Oliveira *et al.* (2016) analyze 189 papers published over the last 25 years and find that only 29.5% of papers test and evaluate the performance of models using real case studies. Evidently, a large number of verification results cannot be easily accepted by managers. There are two main reasons for situation. First, numerous studies have tested the validity of models based on statistical methods in the verification stage. However, managers undoubtedly expect to see the performance of decisions from an operational perspective compared with statistical results (Besbes *et al.*, 2010). Second, the verification processes of many studies do not involve models' users, which means users' possible deviation behaviors are not incorporated. For these reasons, the percentage of research with valid performance verifications may be lower than anticipated.

In addition, verifying the revenue increment or cost savings is particularly challenging for brick-and-mortar stores and traditional manufacturers. Unlike online platforms, e.g., Alibaba, JD.com, and Airbnb that can carry out online experiments to test the performance of optimization models, it is considerably more expensive to conduct verification experiments for brick-and-mortar stores and traditional manufacturers. For instance, the prices of products sold in a brick-and-mortar store are relatively fixed and cannot be adjusted frequently. In Section 2.3, we further illustrate the challenges of verifying performance with a traditional manufacturer.

2.2. Research opportunities

A field experiment is a method for testing causality under real scenarios by randomly assigning subjects to treatment or control groups. By conducting field experiments, researchers can obtain previously implicit or unobservable information to make unbiased estimates. For instance, a field experiment conducted by Gaur and Fisher (2005) in toy stores yields an unexpected result that sales sometimes increase with a price increase. They find that two factors can explain this observation. One is that price is regarded as an indicator of quality in some cases, and the other is that consumers treat some products as gifts, so the sweet spots of pricing are more popular. Fisher *et al.* (2018) explore demand signals that cannot be observed by researchers based on a field experiment with randomized prices. We refer readers to the following studies that use field experiments to investigate the effects of inventory levels on demands (Craig *et al.*, 2016), information disclosure (Allcott and Sweeney, 2017), incentive design (Brahm and Poblete, 2018), fit information in online retail (Gallino and Morenob, 2018), etc.

There is growing literature that incorporates the effects of real operations and users' behaviors into performance verifications by conducting field experiments in governmental or industrial partners. Based on a field experiment in an online retailer, Ferreira *et al.* (2016) demonstrate that the price increase suggested by the proposed algorithm will increase revenue without causing a significant decrease in sales, thus alleviating managers concerns. Levi *et al.* (2020) design a new two-stage auction for online agri-platforms and conduct a field implementation in Karnataka, India, to examine the effects of bidders' behavioral factors on their bidding strategies.

We broadly classify the OM literature that uses field experiments in performance verifications into two categories: online platform and offline platform experiments.

2.2.1. Field experiments in offline platforms

Mukhopadhyay and Kekre (2002) conduct a field experiment in a large industrial supplier to examine the strategic and operational benefits of electronic integration in procurement processes. Caro and Gallien (2012) work on the design, implementation, and evaluation of a new system for clearance pricing in Zara. Through a field pilot experiment, they demonstrate that the decisions from the new system yield significant revenue growth and have a cultural impact on the operations of Zara. In collaboration with a beverage vending retailer, Kawaguchia (2021) finds that the performance of the proposed assortment algorithm in a field experiment is not as good as that in a simulation, due to workers noncompliance with the algorithm. Other interesting topics studied by offline field experiments include initial shipments (Gallien *et al.*, 2015), the effects of stockout-based substitution on demand estimation and inventory planning (Lee *et al.*, 2016), etc.

Another important question is how to determine the appropriate number of experiments so that the results are valid without causing excessive costs. This question is particularly relevant for offline field experiments that are relatively expensive to conduct. Li *et al.* (2015) demonstrate that the number of required experiments can be reduced when the problems being studied have favorable structures. It is worth noting that there is growing literature verifying the performance of models by statistically analyzing data from direct observations rather than designed experiments, e.g., Shin *et al.* (2018) develop a covariate matching method to quantify the performance improvement of wind turbine updates. Studying how to make better use of naturally observed data can be a promising research direction.

2.2.2. Field experiments in online platforms

Schwartz *et al.* (2017) conduct an online field experiment to verify the performance of an advertising policy in collaboration with a large retail bank, demonstrating that the policy yields an 8% increase in customer acquisition rate. We refer readers to the following latest studies that verify model performance by online field experiments, including partnering with Airbnb (Cui *et al.*, 2020), Alibaba (Zhang *et al.*, 2019; Feldman *et al.*, 2021; Sun *et al.*, 2021), Amazon (Cui *et al.*, 2019), and JD.com (Qi *et al.*, 2021).

In particular, recent studies have noticed the violations of the Stable Unit Treatment Value Assumption (SUTVA) that occur in common experimental designs for two-sided markets, such as demand-side randomization and supply-side randomization. For instance, Johari *et al.* (2022) show that single-sided randomization will cause estimation biases for two-sided markets such as Airbnb. To address this issue, Ha-Thuc *et al.* (2020) propose a counterfactual framework for seller-side A/B testing on Facebook Marketplace, demonstrating that the framework satisfies SUTVA. Johari *et al.* (2022) introduce a two-sided randomization that is unbiased when there is an extreme imbalance between supply and demand. In the experimental design for verifications, a two-sided field experiment is used by Ye *et al.* (2020) to ensure SUTVA when they verify the performance of a data-driven optimization algorithm for cold start in online advertising platforms. Future research can explore other experimental design methods that guarantee SUTVA or quantify the biases introduced by the violations of SUTVA.

2.2.3. Data-driven decision methods

The difficulty in verification is largely caused by the lack of information of the indirect benefits, e.g., reduced lost sales and diminished brand value. We illustrate this point using a case study in Section 2.3. There is growing literature leveraging data-driven decision methods to tackle this issue. For example, Derhami and Montreuil (2021) develop a data-driven algorithm to estimate the potential lost sales in a distribution network.

2.3. An industry case with a manufacturing company

Based on our joint project with a manufacturing company, we illustrate the challenges of verifying performance improvement. The collaboration is dedicated to promoting the digital transformation of OM in the company, including demand forecasting, production scheduling, and marketing resources allocation. We encountered the following major challenges.

First, some critical data at the stage of performance verification were not available and hard to obtain. When calculating the production scheduling algorithm, an important criterion is the number of production line changeovers reduced. Nonetheless, the energy consumption caused by each production line changeover can not be precisely determined. To quantify the energy consumption, we needed to compute the total energy consumption per day, the average working time per production line per day, and the average time of each production line changeover. The amount of effort to accurately calculate these data was exhausting. The lesson from this challenge is that we could have collected the energy data at the beginning, instead of acquiring these data at the end. Put differently, the criterion of the performance verification should better be clarified at the beginning of the project.

Second, industry practitioners may be hypercritical of model assumptions in the verification stage. In our joint research with the manufacturer on marketing resources allocation, we proposed that the base level and potential of a market are influenced by eight factors, which was acceptable to our collaborative researchers from the company. However, when it came to the performance verification stage, practitioners outside our joint research team began to challenge us by listing the various factors that should be considered. This is reasonable because the practitioners who carry the burden of risk from a failed implementation are cautious when it comes to implementing new approaches recommended by people who do not have the same depth of knowledge about the industry. Therefore, we recommend that researchers allow extra time in the verification phase to respond to potential new requirements.

Third, indirect and long-term benefits were difficult to compute. In the performance verification of the production scheduling algorithm, direct costs such as energy consumption decrements due to the reduction in working time and production line changeovers, along with the reduced labor cost, were easier to compute than the indirect economic factors, such as reduced out-of-stocks. Although the indirect benefits were difficult to verify, they were often more important.

The same challenge arose in the performance verification of our joint research with the manufacturer on demand forecasting. Figure 2 displays both the short-term and long-term effects of demand forecasting. The short-term impact of demand forecasting was relatively straightforward to calibrate. Nonetheless, the effects of the forecasting model for the medium- and long-term decisions, such as optimizing competitive strategies and strategic planning, were extremely hard to measure.

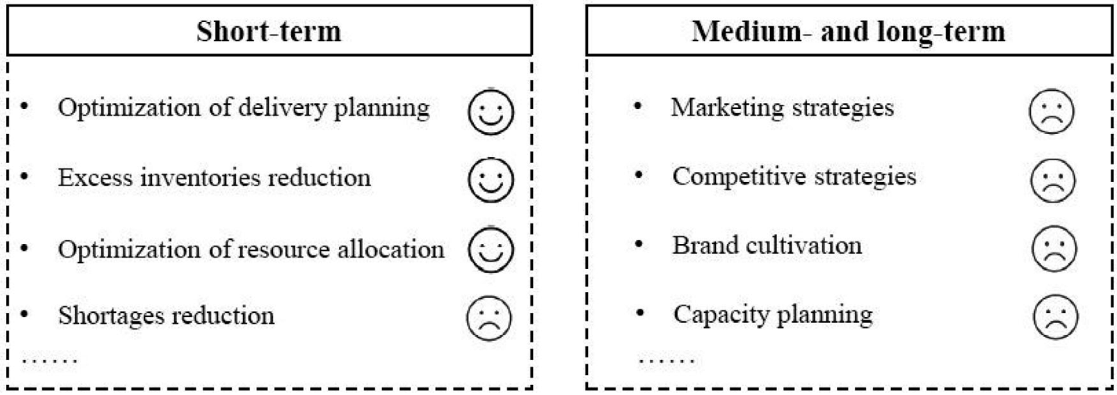


Figure 2. Benefits from the demand forecasting model.

Last but not the least, it is critical to establish direct communication with managers at the "right" level. One important criterion for judging whether it is "right" is that the managers should not only be familiar with the daily operations, but also have strategic and systematic thinking.

3. Building trust between algorithms and practitioners

If we expect OM practitioners to implement an idea recommended by academics, the idea must be properly grounded in the practitioners' minds. Unfortunately, some of the most obvious ideas in OM academia take years to work their way into practitioners' minds. Thus, it often happens that practitioners' decisions deviate from algorithms' recommendations, causing significant costs (Ibanez *et al.*, 2018; Sun *et al.*, 2021). On the other hand, we must acknowledge that the lack of user trust in OM models is sometimes due to model developers' lack of understanding of practical issues from business and user perspectives. Facing these trust issues, redesigning algorithms by analyzing the behavioral factors and improving the ease of interpretation of algorithms' results are two promising directions.

3.1. Challenges

Trust issues mainly stem from the information mismatch between algorithms and users, the complexity of algorithms' results which makes it difficult for practitioners to use, and the lack of explainability.

3.1.1. Information mismatch

This mismatch is mostly reflected in the user's knowledge of information unobserved by algorithms. For example, in bin packing problems, Sun *et al.* (2021) demonstrate that deviations partly stem from the fact that workers possess more information about the items to be packed, e.g., some items can be folded or compressed, and some items are fragile. This information mismatch is common in a variety of optimization problems, e.g., in inventory control (Van Donselaar *et al.*, 2010), task scheduling (Ibanez *et al.*, 2018), etc.

3.1.2. Complexity of results

"People tend to reject what they do not understand." (Little, 1970). According to Rossit *et al.* (2019), users may prefer a suboptimal policy over an optimal policy in vehicle routing problems if the optimal policy is too complicated and the suboptimal policy is simple to use. In warehouse management, packing workers sometimes choose larger boxes rather than following the instructions from the bin packing algorithms. This result is due to the complexity of the algorithm, due to workers being unable pack items in the manner prescribed by the algorithm (Sun *et al.*, 2021).

3.1.3. Lack of explainability

Practitioners seek more transparency in their algorithms. For example, there are more than 20 petroleum refining factories owned by the China National Petroleum Corporation. The company's headquarters issues monthly production plans to each factory based on the total crude oil resources. Nonetheless, due to the lack of explainability, central planners often need to spend a considerable amount of time explaining to each factory why the allocation is preferred. The lack of succinct and understandable explanations from a user perspective hinders the trust-building process and leads to the limited usage (Shin, 2021).

3.2. Research opportunities

According to Dietvorst *et al.* (2018), people are more willing to use algorithms that allow users to make appropriate modifications. To build trust and reduce the deviations between users' actual decisions and algorithm results, model developers should be user-centric and provide users with easy-to-execute and explainable instructions.

3.2.1. User-centric algorithms

Traditional OM models usually regard users' deviating responses as randomness or an error (Boudreau *et al.*, 2003). In fact, these deviations appear in many OM contexts, including the Newsvendor problem (Schweitzer and Cachon, 2000), forecasting (Lawrence *et al.*, 2006; Kremer *et al.*, 2011), inventory risk allocation (Davis *et al.*, 2014), and information sharing (Cui *et al.*, 2015).

There are some pioneering OM studies that focus on analyzing and capturing behavioral deviations to improve models and algorithms. Little (1970) emphasizes that managers prefer models that are concise, robust, controllable, and adaptive. Based on this preference, he proposes a set of model-based procedures called decision calculus. Van Donselaar *et al.* (2010) modify replenishment algorithms by incorporating previously ignored factors, which come from the learning of retailers' ordering deviations. Based on capturing drivers' routing behaviors and travel time predictions, Liu *et al.* (2020) improve order assignment algorithms in the last-mile delivery. Sun *et al.* (2021) predict users' deviations by using machine learning techniques and then redesign traditional bin packing algorithms to avoid complex and incomprehensible instructions, thereby improving the user experience and operating efficiency. It would be interesting to investigate how user-centric algorithm design may help improve performance in other problem settings.

3.2.2. Explainability

There are two streams of literature on explainability. The first stream is to explore explanation techniques. For instance, Ribeiro *et al.* (2016) propose an algorithm to explain the predictions of models and show that the explanations are important for increasing users' trust. Further, a framework for the explanations of predictions is presented by Lundberg and Lee (2017), where six existing explanation techniques are unified. The second stream of literature suggests using explainable models directly, such as decision trees (Bertsimas and Stellato, 2021). According to Bertsimas and Dunn (2017), decision trees are often preferred over other models that may have higher accuracy but are relatively unexplainable in practice. In fact, there is a trade-off between explainability, and accuracy (Arrieta *et al.*, 2020). Although there have been plenty of studies on explainability, few papers consider users' cognitive biases and social expectations in the explanation process, except for Miller (2019), Poursabzi-Sangdeh *et al.* (2021), etc.

3.2.3. Data-driven decision methods

Facing the challenge of information mismatch, some researchers try to use data-driven decision methods to acquire previously unknown information. Van Donselaar *et al.* (2010) find that store managers have more information than inventory models by learning retailers' behavior data, which explains why they often disagree with the recommended ordering policies. By analyzing the data from 2,400,000 radiological diagnoses, Ibanez *et al.* (2018) find that doctors tend to deviate from the prescriptions given by task scheduling algorithms because they prioritize similar tasks and those tasks they expect to complete faster. These previously hidden messages can be used to redesign algorithms, making them more user-centric and trustworthy.

Another highly related research stream is user trust studies on human factors. Hoff and Bashir (2015) systematically review empirical evidence on factors that influence trust in automation and propose a three-tier trust model. Recently, Lee and Kolodge (2020) study trust in self-driving vehicles by text analysis.

3.3. An industry case with an automobile manufacturer

In this subsection, we recommend an explainable modeling framework proposed by Wang *et al.* (2021). Based on three types of user needs, the explainable modeling is aimed at building trust between modelers and stakeholders and then supporting successful Digital Twin implementations. In our joint production scheduling project with an automobile manufacturer, we applied the framework to answer the following three questions that practitioners are most concerned about:

- Q1: What is the impact of different constraints on objective values?
- Q2: Does the model perform well in all production scenarios?
- Q3: If the company decides to focus more on a certain performance measure, how should the model be adjusted?

The framework of explainable modeling in Wang *et al.* (2021) is summarized in Figure 3. Each option represents a set of constraints that can be incorporated in the scheduling model, e.g., the total production quantity limit, the batch volume limit, etc. We considered two production scenarios: (i) low product variety and high demand and (ii) high product variety and low demand. Performance measures include the on-time delivery rate, the total number of changeovers, model complexity in terms of running time, data integration cost (a large part of data require manual input and maintenance), etc.

To answer Q1, we defined *model-based explanation* as the comparative analysis of optimal objective values with different constraint sets in a given production scenario. To answer Q2, we defined *scenario-based explanation* as the comparison of the optimal objective values of a model in different production scenarios. Finally, to answer Q3, we defined *goal-oriented explanation* as the improvement of performance measures of users' interests, e.g., users may be interested in increasing the on-time delivery rate by 10% and reducing the number of total changeovers by 5% in a production scenario. Automatically and systematically generated explanations can speed up the implementation process.

4. Balancing model accuracy and simplicity

The contradicting goals of OM practitioners and academic researchers call for the need to balance model accuracy and simplicity. Practitioners expect the model to be as close to the practical problem as possible, so that they can directly use the solutions without additional adjustments. In academia, a concise and general theory that is only approximately accurate is usually more useful than a theory that is completely accurate, but involves many details and exceptions (Bohanec and Bratko, 1994).

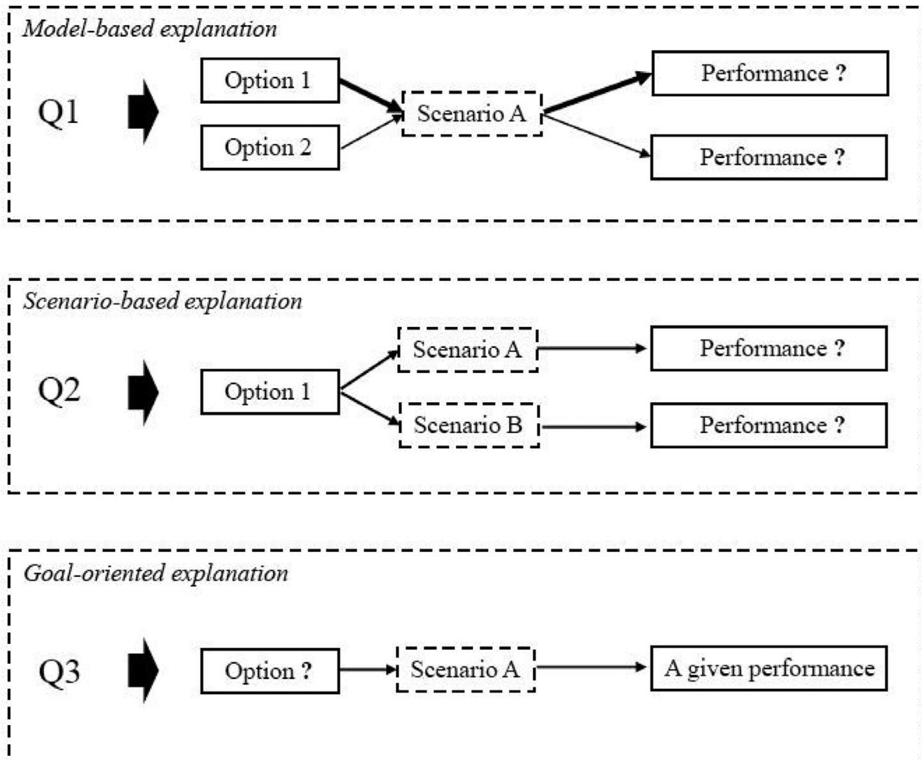


Figure 3. A framework of explainable modeling.

4.1. Challenges

In complex systems, a practically challenging and time-consuming task is to clarify each of the elements, their boundaries and importance to each other, how they change over time, and interact as a system to accomplish a specific purpose. Without simplifying real-world systems, the elements and the complex interactions may make it difficult to analyze the issue about which we are most concerned (Oliveira *et al.*, 2016). First, building a complex model that closely reflects real-world supply chains takes an extraordinary amount of effort and time. By the time the modeling is complete, the environment in which supply chains exist may have totally changed. Second, it is computationally challenging to solve a complex model.

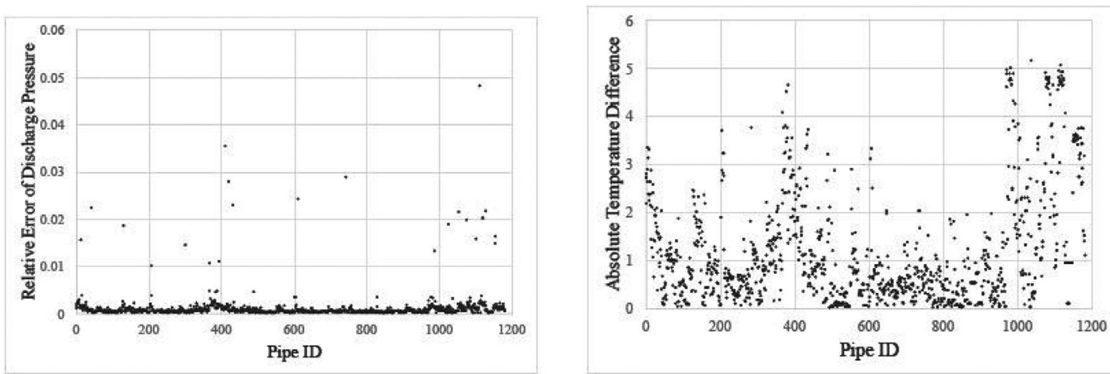
Meanwhile, oversimplified models are not suitable for real-life problems. There has always been an adherence to theoretical puzzles without ascertaining how close a theory is to the truth (Reisman and Kirschnick, 1994; Ormerod and Kiossis, 1997; Ormerod, 2002). Consequently, an oversimplified model may suggest a production or transportation plan that is impossible to execute. The inaccuracy in a model may also hurt the validity of the claimed performance improvement.

4.2. Research opportunities

Where balance can be found is mainly based on the cost of computing resource usage and computational time and the requirements of acceptable accuracy (Hunt, 1965). Bohanec and Bratko (1994) use decision trees to represent the definitions of concepts and translate the simplification of a concept into finding the smallest pruned tree that satisfies the accuracy requirements. According to Min and Zhou (2002), a model builder should define the scope of a supply chain model based on the key dimensions of practical problems while ensuring that the solution is not excessively complicated so as to compromise the trade-off between model simplicity and accuracy. They also summarize two guidelines for defining the scope of a supply chain model: one is based on three levels of decision hierarchy, and the other is determined by the structure of a supply chain network. Fotouhi *et al.* (2016) propose a framework where users are allowed to select an appropriate compromise between model accuracy and simplicity, here by using Pareto optimal sets.

4.2.1. Data-driven decision methods

In the field of OM, researchers usually simplify models by assuming distributions for uncertain parameters. Nonetheless, there may be significant differences between the assumed and actual distributions, thereby weakening the accuracy of models. To tackle this issue, Levi *et al.* (2007) study the sample average approximation approach for data-driven Newsvendor problems. Chen and Chao (2019) develop parametric data-driven learning algorithms for demand functions in a joint pricing and inventory control problem. Garcia *et al.* (2020) introduce a data-driven stochastic optimization framework to avoid the need to make distributional assumptions for uncertainty in risk estimates. Recently, machine learning methods have been applied to uncertainty modeling, e.g., KM estimator (Huh *et al.*, 2011) and neural networks (Cao and Shen, 2019; OroojlooyJadid *et al.*, 2020; Qi *et al.*, 2021). To sum up, a more practice-



(a) Relative error (b) Absolute error

Figure 4. Error of isothermal model.

oriented and accurate model can be built using data-driven uncertainty modeling without increasing excessive computational complexity.

Some future research directions are promising. First, most of the existing methods for the accuracy–simplicity trade-off aim at problems and models with relatively low computational complexity. For complex situations such as Markov Decision Process models with high-dimensional states and actions, how to find the simplest model within a target accuracy is a challenging question. Second, parsimonious predictive models from the machine learning literature are widely used to seek a solution that enforces some sparsity. For example, LASSO helps improve model simplicity and interpretation as, it intuitively enforces a smaller number of features. It would be interesting to investigate the accuracy–simplicity trade-off with the help of these approaches.

4.3. An industry case with China National Petroleum Corporation

To illustrate the trade-off between model accuracy and simplicity, we describe an industrial project with the China National Petroleum Corporation on natural gas pipeline transmission. This project was one finalist of the 2018 INFORMS Franz Edelman prize (Xue *et al.*, 2016; Han *et al.*, 2019). We show the accuracy and complexity of each model employed in the literature or recommended by our industry collaborators.

The main nonlinearity of the mathematical problem arises from the relationship between the gas flow and gas pressure in pipelines. For example, for a pipeline connecting locations i and j, let $f_{i,j}$ be the amount of gas flow. Let p_i and p_j be the locations i and j's gas pressures, respectively. One common modeling approach is to assume that gas pressures p_i and p_j determine the gas flow $f_{i,j}$ in a nonconvex way:

$$\beta_{i,j} f_{i,j}^2 = p_i^2 - p_j^2, \qquad \text{(F-P)}$$

where $\beta_{i,j}$ is a constant and positive parameter.

One key underlying assumption behind (F-P) is that the natural gas system is in a steady state. In transient systems, three important partial differential equations are used (Mahlke *et al.*, 2010). We note that (F-P) is not the only way to describe a steady system's gas flow and pressure. For instance, on the left-hand side of (F-P), the power of gas flow is 2.0, whereas Zhang and Zhu (1996) use a power of 1.85. This difference is minor and insensitive to the major solution approaches. Other work delves deeper into the parameter β_{ij}. Borraz-Sánchez and Haugland (2013) suggest that β_{ij} further depends on the suction pressure of the pipeline in a nonlinear way, making the relationship even messier.

At the beginning of the project, we found that the elevation difference in the pipeline network was not explicitly considered. In China's pipeline network, it is common to have pipelines connecting nodes with elevation differences of more than 200 meters. In Xue *et al.* (2016), we revise (F-P) to (F-P-1) in order to consider the effects of nonuniform network elevation. In (F-P-1), $\alpha_{i,j}$ represents the effects of the elevation differences between nodes i and j on the pipeline's flow-pressure relationship:

$$\beta_{i,j} f_{i,j}^2 = p_i^2 - \alpha_{i,j} p_j^2, \qquad \text{(F-P-1)}$$

Furthermore, the effects of temperature on the relationship between gas flow and gas pressure in pipelines and compressors have been rarely considered. One exception is Chaczykowski (2010), in which the author studies the effect of different nonisothermal gas flow models on the simulation results of fluid transients in natural gas transmission pipelines. In practice, the values of $\alpha_{i,j}$ and $\beta_{i,j}$ in (F-P-1) change with the gas flow rates, gas pressures, and gas temperature.

We collected real data for over 1000 Chinese pipelines and performed numerical tests to analyze the variations in the temperature. In the natural gas pipeline transmission problems in China, the accuracy of a pipeline model is determined by the relative difference between the discharge pressure obtained by the pipeline model and that obtained by a simulation method. In practice, simulation methods are considered to yield the most accurate results, so they are used as benchmarks.

Figures 4(a) and 4(b) show the relative error and absolute error for each pipeline, where the X-axes represent the pipeline ID and the Y-axes denote the average error with pipeline model. For each pipeline, we generated a set of 50 pairs for the suction pressure and gas flow rate. These pairs comprised most of the pipeline's working conditions. Next, we computed the average relative and absolute error for each pipeline in all 50 conditions. It is clear that the effects of temperature should be considered if the requirement on model accuracy is high.

Finally, we considered each node's temperature as a decision variable and took the model:

$$\beta_{i,j}(\pi_i, T_i) \cdot f_{i,j}^2 = \pi_j - \alpha_{i,j}(\pi_i, T_i) \cdot \pi_i. \qquad \text{(A-P-New)}$$

The new model can achieve high model accuracy but may be very difficult to solve. Figure 5 shows the solution accuracy of isothermal and nonisothermal models in the Chinese XN network. In Figure 5, we can see that for most of the pipelines (149 of 151), the nonisothermal model generates more accurate solutions than the isothermal model.

One thing with which researchers can help industry practitioners, is to scientifically measure each model's accuracy and complexity. When we signed the contract with China National Petroleum Corporation, we thought (F-P) would be enough. We spent 3 months designing and implementing the code only to find its accuracy was not sufficient to meet practical needs. Then, we further incorporated the effects of elevation variation and temperature to obtain new models (F-P-1) and (A-P-New). This change in the model required a few more months of research and testing. In addition, the availability of precise data is key to accurate models. Without precise data, it is impossible to verify and improve the model accuracy.

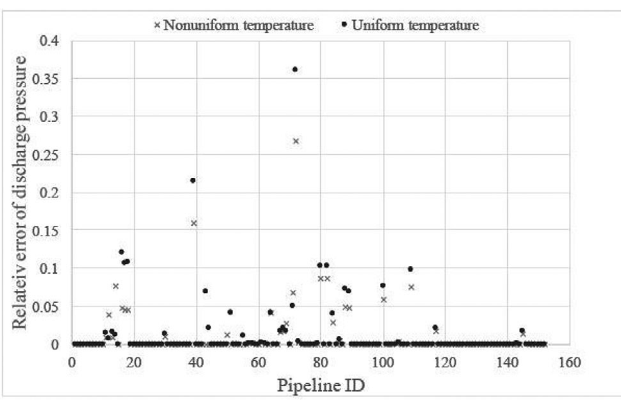


Figure 5. Solution accuracy of the isothermal and nonisothermal models.

5. Conclusions

The future is promising as more OM research tends to bring the theoretical studies close to practical issues. The authors have extracted the data of all the 300 OM-related projects funded by the National Natural Science Foundation of China (NSFC) from 2016 to 2021. Table 1 categorizes the subjects of all the projects by new business model, new technology & Supply Chain Management (SCM) finance, product & behavior, green & health-care & agriculture, and traditional SCM. It is surprising to see that almost all the subjects are driven by real practices, and few project proposals even involve preliminary plans for industrial implementations.

The academic community has always emphasized novel and mathematically provable methods and reproducible results, yet none of which are required for complex real-world issues (Sodhi and Tang, 2008). To thrive, the ISE research community should focus on overcoming the challenges that hinder successful practical applications and embrace cooperation with data-driven decision methods so that our research community will stay relevant and impactful. Combined with real case studies, we list and analyze the

Table 1. OM-related projects funded by NSFC.

New business model	*New technology & SCM finance*	*Product & behavior*	*Green & health-care & agriculture*	*Traditional SCM*
Crowdfunding	Data-driven	Intelligent product	Carbon emissions	Network design
Presale	IoT	Production line design	Carbon trading	Inventory optimization
C2M	Internet plus	User experience	Green innovation	Pricing
O2O	Robotics	Social learning	Green SC	Bullwhip effect
Live streaming	Block chain	Product upgrade	Green logistics	Time-limited service
Sharing economy	Drone	Return	Low carbon SC	Order picking
E-commerce	RFID	Information disclosure	Remanufacturing SC	Rebate
Omnichannel	3D printing	Overconfidence	Circuit breaker	Closed-loop SC
Emerging economy	Digital media	Dynamic fairness	Contactless delivery	Disruption risk
Fulfillment service	Electric vehicle	Repurchase	Post-pandemic era	Flexible procurement
Grey market	Equity investment	Social network	Medicine e-commerce	Emergency
Reverse liability	Green finance	Product externality	Medical platform	Route planning
Crowd innovation	Trade financing	Strategic consumer	Healthcare reform	Demand forecasting
Cloud service	Inventory financing	Individual learning	Vaccination	Negotiation
New retail	Financing platform	Product development	Farmer default	Intelligent warehouse
	Rental arrangement	Behavior recognition	Poverty alleviation	Risk control
	Financial innovation	Customization	Community agriculture	Port container
	Dynamic hedging	Cognitive bias	Fresh agriculture	Revenue management
		Online review	Rural logistics	Perishables
		Reference effect	Farmers' financing	Information sharing
		SC advertisement	Precision agriculture	
		Hunger marketing		
		Price subsidy		

Note. "IoT" stands for "Internet of Things". "C2M" stands for "Customer-to-Manufacturer". "O2O" stands for "Online-to-Offline". "RFID" stands for "Radio Frequency Identification". "SC" stands for "Supply Chain".

key challenges. Ultimately, these challenges can stimulate research opportunities such as user-centric algorithms and explainability. We hope that this article will encourage more ISE researchers to embark on research that can help close the gap when it comes to implementing new OM research ideas into practice.

Acknowledgments

The authors acknowledge Professor Hong Huo, Editor-in-Chief Yu Ding and two anonymous referees for their constructive suggestions that have significantly improved this paper.

Funding

This work was partially supported by the National Natural Science Foundation of China (Grant 71991462) and (Grant 71822105).

Notes on contributors

Dr. Xiaohong Chen obtained her bachelor's degree from Central South University of Technology in 1983 and PhD degree from Tokyo Institute of Technology in 1999. She joined the Hunan University of Technology and Business (HUTB) as the President in 2014, now serves as the party secretary of the HUTB Committee. She is also an academician of Chinese Academy of Engineering, a member of the National Natural Science Foundation of China, a convenor of the Management Science and Engineering Discipline Review Group of the Academic Degree Committee of the State Council, the deputy director of Management Commission of Science and Technology Committee of Ministry of Education, the director of national basic science center for Resource and Environmental Management Theory and Application in Digital Economy Era. Her research interests include decision theory and decision support systems, big data analysis and intelligent management, SME financing, resource-conserving and environment-friendly society and ecological civilization.

Dr. Tianhu Deng obtained his bachelor's degree from Tsinghua University in 2008 and PhD degree from University of California, Berkeley in 2013. He joined the department of Industrial Engineering, Tsinghua University as an assistant professor in 2013, and was promoted to associated professor in 2016. His research interests include supply chain management and production scheduling.

Dr. Zuo-Jun Max Shen is a professor in the Department of Industrial Engineering and Operations Research and Professor of the Department of Civil and Environmental Engineering at UC Berkeley. He joined The University of Hong Kong in 2021 as the Vice President of Research. Max is a Fellow of the Institute for Operations Research and the Management Sciences (INFORMS), the President of the Production and Operations Management Society (POMS), and a past President of the Society of Locational Analysis of INFORMS. His primary research interests are in the general area of integrated supply chain design and management, and data driven decision making.

Yi Yu obtained her bachelor's degree from Nankai University in 2019. Now she is a PhD candidate in the Department of Industrial Engineering at Tsinghua University. Her research interests include supply chain flexibility and mechanism design.

ORCID

Xiaohong Chen http://orcid.org/0000-0003-3919-5215
Tianhu Deng http://orcid.org/0000-0001-6012-6076
Zuo-Jun Max Shen http://orcid.org/0000-0003-4538-8312
Yi Yu http://orcid.org/0000-0002-4029-8150

References

Allcott, H. and Sweeney, R. (2017) The role of sales agents in information disclosure: Evidence from a field experiment. *Management Science*, **63**(1), 21–39.

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

Bertsimas, D. and Dunn, J. (2017) Optimal classification trees. *Machine Learning*, **106**(7), 1039–1082.

Bertsimas, D. and Stellato, B. (2021) The voice of optimization. *Machine Learning*, **110**(2), 249–277.

Besbes, O., Phillips, R. and Zeevi, A. (2010) Testing the validity of a demand model: An operations perspective. *Manufacturing & Service Operations Management*, **12**(1), 162–183.

Bohanec, M. and Bratko, I. (1994) Trading accuracy for simplicity in decision trees. *Machine Learning*, **15**(3), 223–250.

Borraz-Sánchez, C. and Haugland, D. (2013) Optimization methods for pipeline transportation of natural gas with variable specific gravity and compressibility. *TOP*, **21**(3), 524–541.

Boudreau, J., Hopp, W., McClain, J. and Thomas, J. (2003) On the interface between operations and human resources management. *Manufacturing & Service Operations Management*, **5**(3), 179–202.

Brahm, F. and Poblete, J. (2018) Incentives and ratcheting in a multi-product firm: A field experiment. *Management Science*, **64**(10), 4552–4571.

Cao, Y. and Shen, Z. (2019) Quantile forecasting and data-driven inventory management under nonstationary demand. *Operations Research Letters*, **47**(6), 465–472.

Caro, F. and Gallien, J. (2012) Clearance pricing optimization for a fast-fashion retailer. *Operations Research*, **60**(6), 1404–1422.

Chaczykowski, M. (2010) Transient flow in natural gas pipeline–the effect of pipeline thermal model. *Applied Mathematical Modelling*, **34**(4), 1051–1067.

Chen, B. and Chao, X. (2019) Parametric demand learning with limited price explorations in a backlog stochastic inventory system. *IISE Transactions*, **51**(6), 605–613.

Craig, N., DeHoratius, N. and Raman, A. (2016) The impact of supplier inventory service level on retailer demand. *Manufacturing & Service Operations Management*, **18**(4), 461–474.

Cui, R., Allon, G., Bassamboo, A. and Van Mieghem, J. (2015) Information sharing in supply chains: An empirical and theoretical valuation. *Management Science*, **61**(11), 2803–2824.

Cui, R., Li, J. and Zhang, D. (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Science*, **66**(3), 1071–1094.

Cui, R., Zhang, D. and Bassambooc, A. (2019) Learning from inventory availability information: Evidence from field experiments on Amazon. *Management Science*, **65**(3), 1216–1235.

Davis, A., Katok, E. and Santamaría, N. (2014) Push, pull, or both? A behavioral study of how the allocation of inventory risk affects channel efficiency. *Management Science*, **60**(11), 2666–2683.

Derhami, S. and Montreuil, B. (2021) Estimation of potential lost sales in retail networks of high-value substitutable products. *IISE Transactions*, **54**(6), 563–577.

Dietvorst, B., Simmons, J. and Massey, C. (2018) Overcoming algorithm aversion: People will use imperfect algorithms if they can even slightly modify them. *Management Science*, **64**(3), 1155–1170.

Feldman, J., Zhang, D., Liu, X. and Zhang, N. (2021) Customer choice models versus machine learning: Finding optimal product displays on Alibaba. *Operations Research*, **70**(1), 309–328.

Ferreira, K., Lee, B. and Simchi-Levi, D. (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, **18**(1), 69–88.

Fisher, M., Gallino, M. and Li, J. (2018) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Science*, **64**(6), 2496–2514.

Fotouhi, A., Auger, D., Propp, K. and Longo, S. (2016) Accuracy versus simplicity in online battery model identification. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **48**(2), 195–206.

Gallien, J., Mersereau, A., Garro, A., Mora, A. and Vidal, M. (2015) Initial shipment decisions for new products at Zara. *Operations Research*, **63**(2), 269–286.

Gallino, S. and Morenob, A. (2018) The value of fit information in online retail: Evidence from a randomized field experiment. *Manufacturing & Service Operations Management*, **20**(4), 767–787.

Garcia, G., Lavieri, M., Jiang, R., McCrea, M., McAllister, T. and Broglio, S. (2020) Data-driven stochastic optimization approaches to determine decision thresholds for risk estimation models. *IISE Transactions*, **52**(10), 1098–1121.

Gaur, V. and Fisher, M. (2005) In-store experiments to determine the impact of price on sales. *Production and Operations Management*, **14**(4), 377–387.

Ha-Thuc, V., Dutta, A., Mao, R., Wood, M. and Liu, Y. (2020, July) A counterfactual framework for seller-side A/B testing on marketplaces, in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, Virtual Event, China, pp. 2288–2296.

Han, J., Xu, Y., Liu, D., Zhao, Y., Zhao, Z., Zhou, S., Deng, T., Xue, M., Ye, J. and Shen, Z. (2019) Operations research enables better planning of natural gas pipelines. *INFORMS Journal on Applied Analytics*, **49**(1), 23–39.

Hoff, K. and Bashir, M. (2015) Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, **57**(3), 407–434.

Huh, W., Levi, R., Rusmevichientong, P. and Orlin, J. (2011) Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Operations Research*, **59**(4), 929–941.

Hunt, J. (1965) Balancing accuracy and simplicity in determining reorder points. *Management Science*, **12**(4), 94–102.

Ibanez, M., Clark, J., Huckman, R. and Staats, B. (2018) Discretionary task ordering: Queue management in radiological services. *Management Science*, **64**(9), 4389–4407.

Johari, R., Li, H. and Weintraub, G. (2022) Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 0(0), 1–19.

Johnson, M., Midgley, G. and Chichirau, G. (2018) Emerging trends and new frontiers in community operational research. *European Journal of Operational Research*, **268**(3), 1178–1191.

Kawaguchia, K. (2021) When will workers follow an algorithm? A field experiment with a retail business. *Management Science*, **67**(3), 1670–1695.

Kremer, M., Moritz, B. and Siemsen, E. (2011) Demand forecasting behavior: System neglect and change detection. *Management Science*, **57**(10), 1827–1843.

Lawrence, M., Goodwin, P., O'Connor, M. and Önkal, D. (2006) Judgemental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, **22**(3), 493–518.

Lee, J., Gaur, V., Muthulingam, S. and Swisher, G. (2016) Stockout-based substitution and inventory planning in textbook retailing. *Manufacturing & Service Operations Management*, **18**(1), 104–121.

Lee, J. and Kolodge, K. (2020) Exploring trust in self-driving vehicles through text analysis. *Human Factors*, **62**(2), 260–277.

Levi, R., Rajan, M., Singhvi, S. and Zheng, Y. (2020) Improving farmers' income on online agri-platforms: Theory and field implementation of a two-stage auction. Available at SSRN: https://ssrn.com/abstract=3486623.

Levi, R., Roundy, R. and Shmoys, D. (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, **32**(4), 821–838.

Li, J., Rusmevichientong, P., Simester, D., Tsitsiklis, J. and Zoumpoulis, S. (2015) The value of field experiments. *Management Science*, **61**(7), 1722–1740.

Little, J. (1970) Models and managers: The concept of a decision calculus. *Management Science*, **16**(8), 466–485.

Liu, S., He, L. and Shen, Z. (2020) On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science*, **67**(7), 4095–4119.

Lundberg, S. and Lee, S. (2017, December) A unified approach to interpreting model predictions, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Curran Associates Inc., Long Beach, CA, pp. 4768–4777.

Mahlke, D., Martin, A. and Moritz, S. (2010) A mixed integer approach for time-dependent gas network optimization. *Optimization Methods and Software*, **25**(4), 625–644.

Miller, T. (2019) Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, **267**, 1–38.

Min, H. and Zhou, G. (2002) Supply chain modeling: Past, present and future. *Computers & Industrial Engineering*, **43**(1–2), 231–249.

Mukhopadhyay, T. and Kekre, S. (2002) Strategic and operational benefits of electronic integration in B2B procurement processes. *Management Science*, **48**(10), 1301–1313.

Oliveira, J., Lima, R. and Montevechi, J. (2016) Perspectives and relationships in supply chain simulation: A systematic literature review. *Simulation Modelling Practice and Theory*, **62**, 166–191.

Ormerod, R. (2002) On the nature of OR: Taking stock. *Journal of the Operational Research Society*, **53**(5), 475–491.

Ormerod, R. and Kiossis, I. (1997) OR/MS publications: Extension of the analysis of U.S. flagship journals to the United Kingdom. *Operations Research*, **45**(2), 178–187.

OroojlooyJadid, A., Snyder, L. and Takác, M. (2020) Applying deep learning to the newsvendor problem. *IISE Transactions*, **52**(4), 444–463.

Poursabzi-Sangdeh, F., Goldstein, D., Hofman, J., Wortman Vaughan, J. and Wallach, H. (2021, May) Manipulating and measuring model interpretability, in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, Yokohama, Japan pp. 1–52.

Qi, M., Mak, H. and Shen, Z. (2020) Data-driven research in retail operationsa review. *Naval Research Logistics*, **67**(8), 595–616.

Qi, M., Shi, Y., Qi, Y., Ma, C., Yuan, R., Wu, D. and Shen, Z. (2021) A practical end-to-end inventory management model with deep learning. *Management Science* (in press).

Ranyard, J., Fildes, R. and Hu, T. (2015) Reassessing the scope of or practice: The influences of problem structuring methods and the analytics movement. *European Journal of Operational Research*, **245**(1), 1–13.

Reisman, A. and Kirschnick, F. (1994) The devolution of OR/MS: Implications from a statistical content analysis of papers in flagship journals. *Operations Research*, **42**(4), 577–588.

Ribeiro, M., Singh, S. and Guestrin, C. (2016, August) "Why should I trust you?" Explaining the predictions of any classifier, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, San Francisco, CA, pp. 1135–1144.

Rossit, D., Vigo, D., Tohmé, F. and Frutos, M. (2019) Visual attractiveness in routing problems: A review. *Computers & Operations Research*, **103**, 13–34.

Royston, G. (2013) Operational research for the real world: Big questions from a small island. *Journal of the Operational Research Society*, **64**(6), 793–804.

Sargent, R. (2013) Verification and validation of simulation models. *Journal of Simulation*, **7**(1), 12–14.

Schwartz, E., Bradlow, E. and Fader, P. (2017) Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, **36**(4), 500–522.

Schweitzer, E. and Cachon, G. (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, **46**(3), 404–420.

Shin, D. (2021) The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, **146**(102551), 1–10.

Shin, Y., Ding, Y. and Huang, J. (2018) Covariate matching methods for testing and quantifying wind turbine upgrades. *Annals of Applied Statistics*, **12**(2), 1271–1292.

Simchi-Levi, D. (2014) OM forum-OM research: From problem-driven to data-driven research. *Manufacturing & Service Operations Management*, **16**(1), 2–10.

Sodhi, M. and Tang, C. (2008) The OR/MS ecosystem: Strengths, weaknesses, opportunities, and threats. *Operations Research*, **56**(2), 267–277.

Sun, J., Zhang, D., Hu, H. and Van Mieghem, J. (2021) Predicting human discretion to adjust algorithmic prescription: A large-scale field experiment in warehouse operations. *Management Science*, **68**(2), 809–1589.

Van Donselaar, K., Gaur, V., Van Woensel, T., Broekmeulen, R. and Fransoo, J. (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Science*, **56**(5), 766–784.

Wang, L., Deng, T., Shen, Z., Hu, H. and Qi, Y. (2022) Digital twin-driven smart supply chain. *Frontiers of Engineering Management*, **9**(1), 56–70.

Wang, L., Deng, T., Zheng, Z. and Shen, Z. (2021, December) Explainable modeling in digital twin, in *Proceedings of the 2021 Winter Simulation Conference* (in press).

Xue, M., Deng, T. and Liu, D. (2016) CNPC uses an iterative two-stage convex relaxation approach to operate natural gas pipelines. *INFORMS Journal on Applied Analytics*, **46**(6), 533–546.

Ye, Z., Zhang, D., Zhang, H., Zhang, R., Chen, X. and Xu, Z. (2020) Cold start to improve market thickness on online advertising platforms: Data-driven algorithms and field experiments. Available at SSRN: https://ssrn.com/abstract=3702786.

Zhang, D., Dai, H., Dong, L., Wu, Q., Guo, L. and Liu, X. (2019) The value of pop-up stores on retailing platforms: Evidence from a field experiment with Alibaba. *Management Science*, **65**(11), 5142–5151.

Zhang, J. and Zhu, D. (1996) A bilevel programming method for pipe network optimization. *SIAM Journal on Optimization*, **6**(3), 838–857.