Invited Review

Demand management for attended home delivery—A literature review

Katrin Waßmutha,*, Charlotte Köhlerb, Niels Agatc, Moritz Fleischmanna

a Chair of Supply Chain Management, University of Mannheim, Schloss, Mannheim 68131, Germany
b Chair of Business Analytics, European University Viadrina, Große Scharrnstraße 59, Frankfurt (Oder) 15230, Germany
c Department of Technology and Operations Management, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Rotterdam 3062 PA, the Netherlands

A R T I C L E  I N  F O

Article history:
Received 11 March 2022
Accepted 25 January 2023
Available online xxx

Keywords:
Logistics
Last-mile operations
Attended home delivery
Demand management
Planning framework

A B S T R A C T

The continuing e-commerce boom, the design of efficient and effective home delivery services is increasingly relevant. From a logistics perspective, attended home delivery, which requires the customer to be present when the purchased goods are delivered, is particularly challenging. To facilitate the delivery, the service provider and the customer typically agree on a specific time window for service. In designing the service offering, service providers face complex trade-offs between customer preferences and profitable service execution. In this paper, we map these trade-offs to different planning levels and demand management levers, and structure and synthesize corresponding literature according to different demand management decisions. Finally, we highlight research gaps and future research directions and discuss the linkage of the different planning levels.

1. Introduction

The COVID-19 pandemic has boosted the demand for online shopping and home delivery across the globe, and it is likely that some shifts in demand will also have long-lasting effects (OECD, 2020). For example, the global online share of grocery annual sales increased from 7% before the pandemic to 10% at its peak and remains at a high level of 9%, even after the peak.1 Filling this growing demand requires effective and cost-efficient last-mile delivery operations. While the last mile is generally recognized as the most challenging part of the fulfillment process, this is especially true for attended home delivery (AHD), where the customer must be present to receive the goods.

AHD is common for home services and products that require special handling, such as groceries, large appliances, or furniture. To reduce missed deliveries and waiting times, service providers typically let customers choose a delivery time from a menu of time windows or deadlines (referred to as service options). This step involves the customer directly in the service creation process, a characteristic that is typical of the field of service operations management (see, e.g., Coltman & Devinney, 2013).

The concept of AHD is especially well established in the context of online grocery retailing, which is a particularly challenging sector, as profit margins are low, and the delivery of fresh or even frozen goods requires special care in planning and execution. Consequently, many online supermarkets are struggling to create a profitable business.2,3 To manage profitability, service providers can manage both supply and demand. The supply-side levers involve traditional supply chain planning tasks, such as network design, inventory management, and vehicle routing. In general, these levers seek the most cost-efficient fulfillment of a given demand (see, e.g., Han et al., 2017). Demand management focuses on managing customer demand to maximize profitability of a given supply. Typical levers include the specific service options and prices offered to customers. Through these levers, demand management can enhance profits in two ways. First, by increasing revenues by prioritizing high-value customers or by serving more customers due to better capacity utilization. Second, demand management may reduce costs by facilitating more efficient order delivery. In addition to profit maximization, demand management can also contribute to other goals, such as prioritizing specific customer groups when demand exceeds capacity (Schwamberger et al., 2022) or steering customers toward more sustainable delivery times (Agatz et al., 2021).

1 Corresponding author.
E-mail addresses: katrin.wassmuth@uni-mannheim.de (K. Waßmuth).

https://doi.org/10.1016/j.ejor.2023.01.056
0377-2217/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite this article as: K. Waßmuth, C. Köhler, N. Agatz et al., Demand management for attended home delivery—A literature review, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2023.01.056
While traditional supply-oriented approaches have been studied for decades, demand management has only started to attract substantial attention in the research community more recently. Technological advances have been driving this development by allowing for a better understanding of customer behavior and by providing the flexibility to change offered services and prices in real time. When considering current practice, we observe that different e-grocers make different choices regarding their service offerings. In the Netherlands, for example, Albert Heijn offers up to 15 different time windows per day with various lengths (one to six hours) and different delivery fees, whereas Picnic offers any customer a single, free, one-hour time window for each day of the week. We also observe a dynamic development in terms of business models, including on-demand grocery delivery, as offered by Gorillas and Flink. Given the recent progress in the field, the time appears right for a review of demand management for AHD to synthesize the current knowledge and identify relevant open questions.

Demand management generalizes the concept of revenue management, which aims to maximize revenues (Strauss et al., 2018). Costs are generally sunk or proportional to demand in traditional revenue management settings (Klein et al., 2020). In contrast, delivery costs in AHD cannot simply be attributed to individual orders but depend on the specific set of accepted orders (Snoeck et al., 2020). Demand management in AHD involves deciding on the assortment of the delivery service options. This links the topic to the field of assortment planning of physical products across different retail channels (see, e.g., Bernstein et al., 2019).

This paper contributes to the existing literature in the following ways. First, we refine and extend the framework by Agatz et al. (2013) and classify different demand management decisions along strategic, tactical, and operational planning levels. Thereby, our work is the first to explicate the different interrelated planning levels in demand management for AHD. Second, we structure and synthesize the current literature according to the different demand management decisions and planning levels. This provides an up-to-date overview of the literature and identifies research gaps and directions for future research. Third, we introduce a consistent terminology to help bring together different strands of research within the fields of revenue management and vehicle routing. In this way, our work complements previous review papers on online order fulfillment and customer behavior (Nguyen et al., 2018) and integrated demand and revenue management in vehicle routing (Fleckenstein et al., 2023; Snoeck et al., 2020).

The remainder of this paper is organized as follows. In Section 2, we define and structure the field of demand management and develop our classification framework to structure the academic research field systematically. Based on this framework, we review the demand management literature in detail and cluster them into different research streams (Sections 3–5). In Section 6, we highlight our observations and identify gaps and future research opportunities for each planning level. We also discuss the connection between planning levels in that section. Finally, we conclude this literature review in Section 7 by summarizing our main findings and pointing out general avenues for future research.

2. Order fulfillment process

Demand management for AHD aims to generate customer demand and, at the same time, shape it in a way that benefits the fulfillment process. To identify the potential of demand management in this context, we thus need to understand the fulfillment process. At a broad level, it involves activities in sourcing, warehousing, delivery, and sales (Agatz et al., 2008). However, in our context, the most relevant part of the fulfillment process is the one that follows the interaction with the customer, i.e., the customer order decoupling point. This downstream part comprises three main steps, namely, order capture, order assembly, and order delivery (Campbell & Savelsbergh, 2005). In what follows, we briefly discuss each of these steps (Section 2.1.1) and how to coordinate them for multiple orders (Section 2.1.2).

2.1. Fulfillment steps

During order capture, the customer and the service provider mutually agree on when and where the order is to be delivered. To reach such an agreement, the service provider commonly presents an assortment of service options from which the customer can choose. The offered service options may differ in their timing within and across days, their lengths, and their associated delivery prices. Some providers offer the same set of options to all customers, while others tailor them to the customer’s shopping history, delivery location, or basket composition. To ensure a smooth booking process, the service provider must decide on the offered service assortment very quickly, within, at most, a few seconds. Customers choose from the offered options according to their preferences – not placing an order if none of the options meets their expectations. Once the customer chooses a service option, the service provider confirm the order, and the delivery agreement is fixed. It is illustrative to position this process relative to adjacent research fields: In the terminology of the production planning literature, the described process is denoted as real-time single-order capture (Meyr, 2009), while service operations management classifies it as nonsequential offering (Liu et al., 2019).

Order assembly denotes all warehousing operations that are required to prepare an order for delivery, including order picking, sorting, and packaging. Handling the items may be demanding depending on the product category. For example, grocery orders may contain dry, fresh, refrigerated, and even frozen food. This makes order picking quite time consuming. Many service providers therefore seek economies of scale by consolidating the order assembly in larger fulfillment centers that allow for (semi-)automated picking processes. This, however, usually moves the order assembly location further away from the delivery areas, thereby increasing the overall fulfillment lead time. Constraints on innercity space further exacerbate this effect. Service providers that compete on short click-to-door times may therefore opt for a different approach, relying on smaller fulfillment centers situated near customer locations. In particular, on-demand service providers often use a dense network of small innercity depots or even assemble orders in physical stores.

Order delivery refers to the physical delivery of purchased products to customers’ homes within a certain time frame. As this step typically involves assigning customer orders to vehicles and determining the delivery sequence, it can be modeled as a vehicle routing problem (VRP). Service providers often run a proprietary delivery fleet; only a few use external carriers. The fleet can be composed of trucks, vans, cars, or bicycles that visit one or more customers along a specified route. The service includes delivery to the customer’s doorstep, and thus, delivery includes a service time for handover, parking, unloading and – for apartment buildings – carrying the order upstairs. For online supermarkets, the service time is approximately 10 minutes (Klein et al., 2019).
2.1.2. Fulfillment process design

For a single customer order, the three steps of the fulfillment process naturally follow the sequence outlined above. However, the service provider has multiple options to coordinate these steps across multiple orders. For example, the order assembly literature discusses wave and waveless release times, where the former means that incoming orders are held back to be later released in larger batches, whereas in the latter, arriving orders are released immediately and individually (see, e.g., Çeven & Gue, 2017). Similar options apply to order delivery, as discussed in the literature on dynamic consolidation by means of dispatch waves (see, e.g., Klapp et al., 2018). For AHD, we distinguish between a periodic and order-based design of the fulfillment process.

In a periodic fulfillment process, the service provider defines periodic cut-off times, after which all captured orders are assembled and delivered. In other words, there is a fixed period for assembly and delivery that does not overlap with the respective order capture period. This approach exploits economies of scale by consolidating orders in the assembly and delivery steps. The resulting efficiency benefit comes at the expense of a longer click-to-door time since captured orders have to wait until the cut-off time before being further processed. The service provider can choose the cut-off frequency to manage the speed/efficiency trade-off. For online groceries, daily or semi-diurnal cut-offs are common.

In an order-based fulfillment process, the service provider decides dynamically on each customer request whether to initiate the assembly and delivery of orders captured up to that time. In particular, this includes the option to assemble and deliver each order individually immediately after capture. Intuitively, this process design is common for businesses that compete aggressively on speed. It is worth pointing out that a ‘same-day delivery’ service does not necessarily imply an order-based fulfillment process. In fact, under periodic fulfillment, a cut-off time early in the day may also allow for deliveries later on that same day. Thus, from a planning perspective, there is a greater distinction between periodic and order-based processes than between ‘same-day’ and ‘next-day’ delivery. We illustrate this point with specific examples below and visualize it in Fig. 1.

The Dutch grocery retailer Albert Heijn follows a periodic fulfillment process with cut-off times at noon for deliveries the next morning, and at midnight for deliveries the next afternoon. After each cut-off, delivery routes are planned, and order assembly takes place in one of five online fulfillment centers. Similar to Albert Heijn, the German e-grocer REWE also operates a periodic fulfillment process. REWE uses a cut-off time of 1 pm, which allows orders to be delivered in the late afternoon on the same day. To enable fast delivery and handling of more than 20,000 products, the company invests in semi-automated fulfillment centers close to delivery areas. In contrast, the German beverage delivery service Flaschenpost does not communicate periodic cut-off times but guarantees delivery within 120 minutes for every incoming order – a service proposition that requires a particularly fast fulfillment process. To meet this requirement, Flaschenpost operates 23 fulfillment centers to distribute an assortment of approximately 2000 products to more than 150 German cities. Each of these facilities is equipped with approximately 70 vans that deliver up to ten orders per trip. We denote this fulfillment approach as order-based with dynamic order consolidation.

Further speeding up the fulfillment process, German start-up Gorillas offers on-demand grocery delivery within 10 minutes. To meet the extremely short delivery times, the company sets up micro fulfillment centers in each delivery area and limits the offered product assortment to 2500 products. In addition, they hand-pick each captured order immediately and deliver it by bicycle. Such a fulfillment process is order-based without consolidation.

2.2. Demand management decisions

In the previous subsection, we highlighted the main steps of the fulfillment process in AHD services. How efficiently a company can execute these steps depends on the properties of individual orders, such as their click-to-door time (e.g., Ulmer, 2017) and delivery time specificity (e.g., Lin & Mahmassani, 2002), as well as on the temporal and geographical distribution of the overall set of captured orders (e.g., Ehmke & Campbell, 2014). At the same time, these factors are intimately linked to customer preferences and thus to the popularity of delivery service options. Demand management aims to manage the resulting trade-offs between captured demand (revenue) and assembly and delivery efficiency (costs). In this sense, Fig. 2 illustrates the interdependence between demand management and the steps of the fulfillment process and the implied impact on revenue and costs.

Demand management encompasses a diverse set of different decisions. We propose mapping these out along two dimensions, distinguishing three planning levels (strategic, tactical, and operational) and two levers (offering and pricing). This approach gives rise to six different sets of demand management decisions, as shown in Table 1. In what follows, we briefly discuss both dimensions of this framework.

---

2.2. Planning levels

As is common in many areas of supply chain planning and operations management (Fleischmann et al., 2015), we distinguish between different hierarchically-linked planning levels, i.e., strategic, tactical, and operational. We define these levels based on their aim, their time horizon, and their relation to the fulfillment process timeline. Strategic decisions are design choices specified over a long horizon, while tactical and operational decisions consider the management of service options over a shorter time span. Strategic and tactical decisions take place before order capture while operational decisions are based on real-time information on actual demand. In what follows we elaborate on each of these levels in some more detail.

Strategic demand management defines the boundaries within which tactical and operational demand management are embedded. It constitutes a special case of the service design stage in service operations management (see, e.g., Roth & Menor, 2003) and also bears resemblance with structural decisions in revenue management (Talluri & van Ryzin, 2004). Strategic demand management reflects the overall business strategy and, to gain a competitive advantage, must be carefully aligned with the competitive environment, customer preferences and willingness to pay, and operational implications. Respective decisions determine the target markets and design the general service assortment, based on a market’s demand potential. This includes selecting the service region and pricing model, designing the service options, and defining appropriate service segments for subsequent tactical planning. The term service segment refers to a customer group that should receive the same service assortment (e.g., a geographical area).

The subsequent planning levels address the management of the designed service assortment within the established boundaries. We classify any such decisions taken before order capture as tactical demand management. Tactical decision-making is based on (aggregated) demand forecasts and exploits the heterogeneity of customers in the delivery market. Corresponding decisions include differentiation of service options and prices for different service segments. Moreover, tactical planning can be applied to simplify short-term operational planning, for which only limited computational time is available.

We denote any decisions made during order capture as operational demand management, i.e., decisions that are made in real time, based on detailed information on actual customer orders. Thus, operational decisions are highly time-critical and directly affect the interaction with the customer. They include accepting customer orders and adjusting the availability of service options and attached prices in the short term. For order-based fulfillment processes, these decisions are additionally combined with simultaneous fulfillment planning, as the order capture step overlaps with order assembly and delivery. This differs from periodic designs, where fulfillment planning can be postponed until after the cut-off. Both tactical and operational demand management share analogies with traditional revenue management (Agatz et al., 2013; Snoeck et al., 2020).

In this subsection, we introduced the planning levels top-down from strategic to operational, thereby reflecting the natural sequence of decision-making. However, we observe that the corresponding literature is evolving in the opposite direction, with many demand management approaches starting at the operational level and gradually providing insights to the strategic level. We follow this development in Sections 3 to 5 and review the demand management literature bottom-up, from operational to strategic planning.

2.2.2. Levers

The demand management levers, offering and pricing, capture the main characteristics of the delivery service. Offering refers to both the design of service options and the management of their availability. The latter are binary decisions (an option is either offered or not offered) that can (i) ensure feasibility and (ii) steer customer choice by intentionally withholding some feasible options. Service providers can also manage demand through pricing decisions. We use ‘pricing’ to denote a variety of monetary and non-monetary incentives to steer customer choice and generate additional revenue by exploiting differences in willingness to pay. The pricing lever allows a more fine-grained demand management since prices can be chosen from a continuous interval, rather than from a binary set. Previous research in the context of e-grocery suggests that small incentives may suffice to change customer behavior (Campbell & Savelbergh, 2006).

Offering and pricing can be used as substitutes to steer demand. However, it should be noted that customers might perceive them very differently, as the willingness to pay is generally low (Goethals et al., 2012). Furthermore, the two levers also have complementary features and constitute building blocks that can be combined into an overarching demand management approach. For example, in the case of operational demand management, pricing usually builds on the feasibility decision, i.e., the service provider first determines which options could be offered, and then sets prices for the feasible set of options. Therefore, and in line with the dichotomy of quantity- and price-based revenue management (Talluri & van Ryzin, 2004), we present and discuss offering and pricing separately in what follows.

3. Operational demand management

In this section, we review the literature on operational demand management, distinguishing offering and pricing decisions. We provide an overview of the corresponding literature in Table 2. We characterize published work with respect to the considered problem setting, the decision-making process, and the computational study. We further elaborate on these characteristics below. They then lead us to identifying clusters of closely related papers that we discuss in Sections 3.1 and 3.2.

We distinguish different problem settings for operational demand management by the design of the fulfillment process (periodic or order-based; see Section 2.1.2) and by the type of service options offered to the customer, i.e., time window or deadline.

To characterize the decision-making process, we highlight the service provider’s assortment decision approach, that is making decisions either independently for individual service options or
<table>
<thead>
<tr>
<th>Offering</th>
<th>Problem setting</th>
<th>Decision-making process</th>
<th>Computational study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assortment decision</td>
<td>Order assessment</td>
<td>Order assessment components</td>
</tr>
<tr>
<td></td>
<td>Fulfillment approach</td>
<td>Choice behavior</td>
<td>Fulfillment feasibility</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Casazza et al. (2016)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – ADV – Service</td>
</tr>
<tr>
<td>Hungerländer et al. (2017)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – ADV –</td>
</tr>
<tr>
<td>Köhler &amp; Haferkamp (2019)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – APR –</td>
</tr>
<tr>
<td>Visser et al. (2019)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – ADV –</td>
</tr>
<tr>
<td>Köhler et al. (2020)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – INS Cost</td>
</tr>
<tr>
<td>Trudcn et al. (2022)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – ADV –</td>
</tr>
<tr>
<td>van der Hagen et al. (2022)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – APR –</td>
</tr>
<tr>
<td>Campbell &amp; Sueslesberger (2005)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Indep. – INS Profit</td>
</tr>
<tr>
<td>Mackert (2019)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Avraham &amp; Raviv (2021)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Lang et al. (2021a)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Lang et al. (2021b)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT APR Profit</td>
</tr>
<tr>
<td>Azi et al. (2012)</td>
<td>Order-based</td>
<td>Time window</td>
<td>Indep. – ADV Profit</td>
</tr>
<tr>
<td>Klapp et al. (2020)</td>
<td>Order-based</td>
<td>Deadline</td>
<td>Indep. – ADV Cost</td>
</tr>
<tr>
<td>Pricing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campbell &amp; Sueslesberger (2006)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint EXO INS Profit</td>
</tr>
<tr>
<td>Yang et al. (2016)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Klein et al. (2018)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Koch &amp; Klein (2020)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Asdemir et al. (2009)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT APR Revenue</td>
</tr>
<tr>
<td>Yang &amp; Strauss (2017)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT APR Profit</td>
</tr>
<tr>
<td>Vinsensius et al. (2020)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint EXO – Profit</td>
</tr>
<tr>
<td>Lebedev et al. (2021)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT APR Profit</td>
</tr>
<tr>
<td>Strauss et al. (2021)</td>
<td>Periodic</td>
<td>Time window</td>
<td>Joint RUT APR Profit</td>
</tr>
<tr>
<td>Prokhorchuk et al. (2019)</td>
<td>Order-based</td>
<td>Deadline</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Ulmeer (2020)</td>
<td>Order-based</td>
<td>Deadline</td>
<td>Joint RUT INS Profit</td>
</tr>
<tr>
<td>Klein &amp; Steinhardt (2023)</td>
<td>Order-based</td>
<td>Deadline</td>
<td>Joint RUT ADV Profit</td>
</tr>
</tbody>
</table>
jointly for a set of options. Related to this aspect, some papers explicitly model customer choice behavior, either through exogenous substitution rates (EXO) or based on random utility theory (RUT). The remaining papers do not model customer choice but assume demand to be independent of the service offering. We also consider two attributes concerning the assessment of an incoming order. First, the service provider must verify the fulfillment feasibility of each service option, given the available fulfillment capacity and the previously committed orders. The feasibility check can be based on either a functional approximation (APR) or a tentative route plan, using simple insertion heuristics (INS) or more advanced routing methods (ADV). Note that for each paper only one of possibly several methods applied is given in the table. In addition to checking feasibility, the service provider may assess the present order value according to different metrics, including cost, service, revenue, and profit. If no order value is considered, they make decisions based on feasibility only. Papers also differ in the components of the fulfillment process that they consider in the assessment of the current order. These may include subsequent order assembly and order delivery. In addition, papers may or may not consider the impact on the fulfillment of future orders, reflected in opportunity costs.

For the computational study, we list the type of demand data (synthetic or empirical) and the business sector of the motivating application.

3.1. Operational offering

Operational offering decisions determine the service options to offer to a customer during order capture. To structure our discussion, we cluster papers with similar characteristics as shown in the upper part of Table 2. In particular, we identify three clusters based on the design of the fulfillment process and the consideration of opportunity costs in the order assessment.

**Periodic fulfillment with focus on order delivery assessment.** Most papers that focus on operational offering decisions consider periodic fulfillment. We can further classify these papers based on whether or not they take into account opportunity costs and thus future orders. Table 2 shows that the papers that ignore the opportunity costs consider single time windows independently and do not explicitly model customer choice behavior. Most of these papers focus on assessing fulfillment feasibility.

One of the challenges of integrating routing aspects into operational demand management is to quickly obtain good solutions to allow for real-time feasibility checks. Hungerländer et al. (2017) develop an adaptive neighborhood search heuristic (ANS) to determine feasible time windows during order capture. They also tailor their ANS to the specific time window problem structure to find better solutions in less time. Truden et al. (2022) study a number of different solution methods for the AHD setting. In line with Hungerländer et al. (2017), they show that it is beneficial to adapt time window heuristics to the specific problem settings. Köhler & Haferkamp (2019) compare various vehicle routing methods to facilitate fast high-quality assessments of the available fulfillment capacity. The authors also introduce an acceptance mechanism based on Daganzo (1987) to approximate expected travel times. Using real-world booking data of an online supermarket, they show that the delivery area and expected demand impacts the performance of different approaches. Visser et al. (2019) study a setting in which multiple customers interact with the booking system simultaneously. It is therefore not only important to do a fast initial time slot feasibility check but also a second check when the customer commits to a certain time slot. Their detailed computational study shows that combining a fast insertion heuristic with a sophisticated background procedure ultimately leads to more accepted orders. van der Hagen et al. (2022) study the use of machine learning (ML) methods to predict the fulfillment feasibility by framing the problem as a binary classification problem. Their results suggest that ML methods can generate accurate feasibility assessments in a fraction of the time needed for common heuristic-based methods.

Another challenge of delivery-oriented order assessment is to account for uncertainty at the time of decision-making. Ehmke & Campbell (2014) seek a reliable feasibility assessment in a setting with uncertain travel times. They compare assessment methods, including a novel insertion-based heuristic that accounts for time-dependent and stochastic travel times. Based on a computational study using real travel data, they find that considering time-dependent travel times is especially valuable in suburban areas, whereas buffers against travel time uncertainty are effective in downtown areas. In addition to feasibility checks, some papers also estimate the present order value using cost and service metrics to maximize the number of orders accepted. In contrast to the cost metric, the service metric explicitly measures customer satisfaction with respect to the service options. Casazza et al. (2016) try to insert a new customer into the current route plan. If this is infeasible, the service provider does not reject the order, but shifts or enlarges the delivery time window. The authors use a dynamic programming algorithm to assess feasibility in real-time and evaluate several decision policies based on different service measures. The results highlight the trade-off between customer service and increasing the number of accepted orders. Köhler et al. (2020) introduce flexibility mechanisms that incorporate myopic information about routing efficiency and delivery locations to dynamically decide whether to offer a long or short time window to a given customer. Their results confirm that the more customers book long time windows, the more flexibility can be maintained for the fulfillment, which increases the availability of time windows for later customers.

**Periodic fulfillment with opportunity cost assessment.** Within the second cluster, we find literature that considers opportunity costs in the assessment of a given order so as to better steer customers to more profitable or cost-efficient options. Contrary to the first cluster, most of the papers simultaneously consider multiple time windows and explicitly model customer choice behavior. However, the techniques applied to test fulfillment feasibility are simpler than in the previous cluster.

In contrast to other papers in this cluster, Campbell & Savelsbergh (2005) decide on individual time window offers independently but are the first to provide a rough estimate of future profits. In particular, for each new request, they solve a routing instance including already accepted customers, the current customer under consideration, and a number of expected future customers.

The remaining papers explicitly model customer choice behavior based on random utility theory. Incorporating customer choice behavior is crucial for joint assortment decisions. However, it is challenging to incorporate a detailed customer choice model taking into account choices and substitution across multiple days, time windows, and delivery prices. Therefore, these models try to balance modeling detail and computational effort. To this end, Mackert (2019) apply a generalized attraction model (GAM) which ranks each time window offer based on the customer’s perceived attractiveness. The authors use the choice probabilities in combination with a mixed-integer programming (MIP) based profit estimation to determine the subset of most profitable time windows for a given customer. They conclude that applying the GAM can lead to a more accurate estimation of customer choice than applying the most frequently used multinomial logit (MNL) model (e.g., Avraham & Raviv, 2021; Lang et al., 2021a; 2021b). Lang et al. (2021a) propose several methods for anticipatory profit estimation using, inter alia, extensive offline training based on samples of expected demand and value function approximation (VFA; see, e.g.,...
Powell, 2016). They highlight the modular composition of the associated routing and revenue management techniques. Lang et al. (2021b) additionally account for multiple short- and long-term revenue metrics, including basket value, the visibility of branded trucks, and popularity among influential customers. In contrast, Avraham & Raviv (2021) focus on efficient multi-day assortment decisions. Different from the previous work, the authors anticipate future demand to maximize the number of expected accepted customers. They use tentative route information both for feasibility checks and as features of a VFA jointly predicting route efficiency over multiple consecutive days. The presented results show that taking into account inter-day dependencies create more efficient fulfillment routes that allows for more accepted orders.

**Order-based fulfillment.** The third cluster addresses offering decisions in order-based fulfillment systems. To date, only a few publications pertain to this stream of demand management literature. The work in this cluster presents sophisticated order delivery methods for order assessment and also takes rough proxies of order assembly into account. We conjecture that the importance of considering all fulfillment steps in the offering decision stems from the order-based fulfillment setting itself, and is due to the high time pressure in this setting.

Azi et al. (2012) consider a setting in which new customer requests arrive during the execution of the routes of previously accepted customers. There are no predetermined cut-off times. However, new customers can only be inserted into time windows of routes that have not yet started. To the best of our knowledge, this is the first paper to integrate vehicle dispatching and order capture. By assuming a load-dependent setup time, this paper also models the interaction between order capture and order assembly. The authors formulate a dynamic decision model in which the acceptance of a customer request depends on a scenario-based opportunity costs. The embedded routing problem is solved with an ANS heuristic. Instead of time windows, Klapp et al. (2020) consider the acceptance of requests that must be delivered no later than the end of the operating day, which constitutes a common delivery deadline. The objective is to minimize the sum of expected travel costs and penalties for rejecting a request. The authors approach this problem as an extension to the dynamic dispatch waves problem (Klapp et al., 2018), adding efficient request acceptance as a demand management decision. They evaluate fulfillment feasibility based on dispatch plans that include a constant parameter representing assembly time, and construct and upgrade the plans using neighborhood search heuristics.

**3.2. Operational pricing**

Operational pricing involves dynamically adjusting the prices of the service options offered during the order capture step. This means setting (customer-specific) delivery prices or other incentives associated with the service options that are displayed when customers arrive over time. Such incentives can stimulate efficient fulfillment operations and maximize revenue in the short term.

We present the literature for operational pricing in three clusters, based on the attributes displayed in the lower part of Table 2. Even across clusters, the available operational pricing models have many aspects in common. Intuitively, each of them accounts for joint assortment decisions and some form of customer choice behavior. We especially highlight the work of Yang et al. (2016) who calibrate an MNL choice model based on a large amount of real booking data from an e-grocer. Many subsequent publications refer to this model and its data to capture customer choice behavior. Other common features among operational pricing approaches are the use of revenue-based metrics (revenue or profit) for order value assessment and accounting for order delivery as well as opportunity costs in the order assessment. These characteristics largely correspond to those of the second operational offering cluster, which also focuses on the anticipatory steering of customer choice. Within this overall picture, we identify three clusters of publications that differ in terms of the fulfillment process design and the method for the fulfillment feasibility assessment.

**Periodic fulfillment with tentative route plans.** Similar to offering, the vast majority of the operational pricing literature assumes a periodic fulfillment process. Within this relatively homogeneous group, the approaches differ mainly in the way they determine fulfillment feasibility. The papers in the first cluster perform a tentative route planning, using insertion heuristics. The tentative route information is also used to estimate profits for assessing the present order value – with or without considering opportunity costs.

Campbell & Savelsbergh (2006) do not consider opportunity costs but estimate the profit contribution of a given order as the sales margin minus the insertion cost, taking into account already accepted customers. An incentive optimization model then trades off price discounts against the increased likelihood that customers will choose time windows with higher profit expectations. More recent approaches seek to also capture opportunity costs, i.e., the impact of demand management decisions on future demand (management). To this end, they typically model the decision problem as a stochastic dynamic program. Yang et al. (2016) are the first to present such a formulation, taking into account the fulfillment costs incurred in the order delivery step. Since this problem is computationally intractable, the authors propose an approximation to compute optimal prices for feasible options in real time.

Similar to Campbell & Savelsbergh (2006), the approximation relies on insertion cost estimates, which are offset against the immediate profit before fulfillment. However, the authors incorporate estimates of future demand as they draw on pools of route plans that involve already existing orders and samples of expected future order locations. Koch & Klein (2020) replace the anticipatory insertion cost by a linear VFA that uses the information retrieved from tentative route planning as features. While the former method can only account for cost-related effects in the opportunity cost estimation, this one accounts for both cost- and revenue-related displacement effects. Instead of applying statistical learning, Klein et al. (2018) choose a model-based approach to capture these effects. Their MIP formulation combines myopic insertion costs derived from tentative route plans with anticipatory seed-based routing that draws its information from a choice-based demand prediction model.

A major challenge in using tentative route information is computational complexity: The insertion cost calculation is a primary bottleneck (Yang et al., 2016), and it may be necessary to periodically recalculate opportunity costs to decrease online computation times (Klein et al., 2018).

**Periodic fulfillment with capacity approximation.** The second operational pricing cluster relies on static capacity controls to assess feasibility instead of using tentative route plans. Alternatively, they skip the feasibility checks altogether and incur penalty costs on capacity shortage. The papers use different approaches to capture the routing aspects of the order delivery step. In addition, they differ in how they link the approximation method used for feasibility assessment to the method used to assess the present order value – in terms of profit or revenue.

Asdemir et al. (2009); Lebedev et al. (2021) study the structure of an optimal pricing policy under MNL customer choice, assuming static capacity controls. Asdemir et al. (2009) assess the present order value using a revenue metric assuming sunk fulfillment costs. They introduce a balanced capacity utilization constraint to implicitly model the order delivery step. Lebedev et al. (2021) account for delivery costs in the terminal state of their dynamic programming formulation and refer to route approximation methods.
(Daganzo, 1987) to determine the assumed capacity controls. The studies show that optimal delivery prices increase dynamically as fulfillment capacities are depleted during order capture (Asdemir et al., 2009), and are monotonic in the number of accepted customers (Lebedev et al., 2021).

The other work in this cluster presents solution methods to the operational pricing problem that involve capacity approximation. Yang & Strauss (2017) build their solution method around Daganzo (1987). Specifically, they use this approximation method not only to determine static capacity controls for feasibility assessment, but also to train an affine VFA to anticipate profit based on the current number of accepted customers and the time remaining for order capture. Strauss et al. (2021) incorporate a similar feasibility assessment but tailor it to a setting with flexible time windows. In particular, they consider a setting in which customers select multiple delivery time windows that are acceptable to them. The customer receives a discount for providing the service provider with more flexibility in order fulfillment. The authors estimate profit through an anticipatory linear program that uses the capacity information from the approximate feasibility assessment. In contrast, Vinsensius et al. (2020) completely ignore feasibility checks at the order capture phase. Instead, they account for infeasibilities in order delivery by means of penalty costs. Yet, the authors incorporate routing properties faced during order delivery: Similar to Yang & Strauss (2017), they estimate profits using VFA. However, rather than relying on approximations, they train their VFA with solutions to a VRP variant with service choice. In particular, they perform the training on simulated historical data and solve the VRP instances using a minimum regret construction heuristic. Thus, although the authors apply explicit route planning within the offline training, they do not perform tentative route planning during the decision-making process, as for example Koch & Klein (2020) do.

Order-based fulfillment. Analogous to operational offering, operational pricing literature addressing order-based fulfillment is scant. In contrast to periodic order fulfillment, delivery decisions are dynamic and stochastic. In what follows, we point out how papers in this cluster deal with this aspect. We also explain how they use tentative route information for assessing opportunity costs. Interestingly, different from the cluster of order-based operational offering literature, none of the considered papers takes order assembly into account.

Ulmers (2020) dynamically set prices for one-hour and four-hour delivery deadlines. Their model optimizes both the pricing strategy and dynamic route dispatch times, where the former aims to maintain fleet flexibility while charging customers according to their expected willingness to pay. The solution method uses tentative route information obtained from an insertion heuristic that is based on already existing orders only. Besides facilitating feasibility checks, the myopic route information is used to derive fleet flexibility measures as features for a linear VFA that assists profit anticipation. Prokhchuk et al. (2019) extend this work and aim to make pricing decisions for reliable service assortments to reduce the number of missed deadlines and increase long-term customer loyalty. To this end, they integrate penalties for late deliveries and account for stochastic travel times that materialize while delivery routes are executed. Similar to the above study, the authors build on myopic route information and apply a linear VFA using flexibility- and reliability-based features for anticipatory profit estimation. In contrast, Klein & Steinhardt (2023) apply a more advanced tentative routing procedure and consider future orders in both profit estimation and route planning. Compared to previously applied insertion heuristics in combination with route-based VFA, the authors perform a sample-scenario state value approximation that involves heuristically solving a profitable multi-trip VRP with release and due times for every sampled scenario.

4. Tactical demand management

Table 3 lists the literature on tactical offering (upper part) and tactical pricing (lower part). Similar to the previous section, we categorize the publications based on their problem setting, the decision-making process, and the computational study. However, the attributes considered within each of these categories differ from those used to structure the operational literature. Again, the table entries allow us to identify clusters of closely related publications, which we discuss in Sections 4.1 and 4.2.

First, we distinguish different problem settings underlying tactical demand management in terms of the number of service options from which an individual customer can choose (single or multiple) and the service segments for which different offering and pricing decisions are made (individual customers or aggregated customer groups).

Second, we consider the forecast-based, tactical decision-making process. Corresponding demand management methods apply different optimization approaches and demand forecasting methods. Optimization approaches differ in terms of the linkage between planned shifts, i.e., they determine the decisions either independently for single shifts or jointly for multiple shifts. Further, we distinguish different model decisions, including assortment decisions, price decisions, and availability controls. While assortment decisions assign sets of service options to the given service segments, availability controls (e.g., booking limits) are set for given assortments with the aim of simplifying subsequent operational decisions. Finally, we list the model objective (cost, revenue, or profit) and the type of service and capacity constraints, if any. In the case of a cost objective, service constraints ensure an exogenously imposed service level with respect to the number of service options (frequency), the distribution of service times (balance), or subsets of service options that can be either continuous (interval) or discrete (candidates). Capacity constraints capture the necessary fulfillment operations and are represented by continuous approximation models (CA), simulation (SIM), or routing models that can be either explicit (ROUTE) or seed-based (SEED). Note that for each paper only one of the possibly several methods applied is given in the table. Concerning the demand forecast, we distinguish between a deterministic and stochastic demand model and indicate whether papers explicitly model customer choice behavior based on random utility theory (RUT). Other papers do not model customer choice but assume demand to be independent of the service offering.

Third, analogous to the operational planning models, information on the computational study includes the type of demand data (synthetic or empirical) and the business sector of the motivating application.

4.1. Tactical offering

Tactical offering decisions determine the availability of service options before the order capture step. In other words, they allocate the corresponding fulfillment capacity to different service segments, based on demand forecasts. In the upper part of Table 3, we observe three clusters of publications that share similarities with respect to the considered service segments and model decisions. As discussed below, each of the clusters represents a specific planning task within the domain of tactical offering – from the simplification of short-term operational planning to service differentiation and long-term customer agreements.

Availability controls. The first cluster focuses on establishing availability controls for a given assortment of service options, i.e., thresholds that guide the decision on the availability of service options for different service segments. This simplifies operational decision-making and resembles the concept of allocation planning in supply-constrained production planning (Meyr, 2009).
In this vein, Cleophas & Ehmke (2014) propose an iterative algorithm to allocate the fulfillment capacities of a geographically differentiated service assortment to value-based customer groups. They first simulate the order capture phase based on historical booking data and by applying customer acceptance rules from the literature (Ehmke & Campbell, 2014). From the simulation results, they derive booking thresholds for each time window and delivery area. The authors then refine the thresholds for discrete order value buckets using the expected marginal seat revenue (EMSR) heuristic, a classical revenue management tool (Belobaba, 1987). The computational results show that the proposed method can generate significant revenue gains in the case of heterogeneous order values. In contrast, Visser & Savelsbergh (2019) focus on foresighted delivery routes to maximize the generated revenue. Inspired by Dutch e-grocer Picnic, which offers a single time window per day for each delivery area, they present an approach to (i) determine the specific time window to offer in each area and (ii) establish an operational control mechanism to determine when time windows should be closed. Both decisions are guided by a priori routes that are constructed over a set of delivery points with known order volumes and revenues. Order placement and order sequence are uncertain. The authors develop a two-stage stochastic program, where routes are determined in the first stage and generated revenue is simulated in the second stage. To reduce complexity, the study assumes a single vehicle, thereby turning the routing problem into a traveling salesperson problem (TSP). The study presents insight into the structure of optimal a priori routes.

**Assortment decisions for aggregated customer groups.** Papers in the second cluster determine an assortment of service options for each geographical area within the service region. In particular, by differentiating the assortment over different areas, the service provider can spatially cluster demand but also temporally sequence the clusters to facilitate efficient delivery routes.

In this light, Agatz et al. (2011) determine the service assortment per shift across days for different geographic areas. They assign a fixed number of service options out of a given pool of options to each service area with the objective of minimizing the expected fulfillment cost. To decompose the problem per shift, the authors assume weekly demand to be evenly distributed over the service assortment. Additionally, expected demand is known and independent of the service assortment. The paper proposes two solution approaches, one based on continuous approximation (Daganzo, 1987) and the other based on integer programming. The authors evaluate the resulting assortments by simulation on the operational level and based on real demand data. The results show a reduction in delivery costs compared to uniform assortments, which is most significant if delivery capacity allows a vehicle tour to span several time windows. Mackert et al. (2019) extend the integer programming-based method with a finite-mixture customer choice model that accounts for heterogeneous revenues and preferences. Furthermore, they eliminate the specification of exogenous service requirements by moving from cost minimization to profit maximization. The authors linearize the choice-based MIP to apply a standard solver and propose a decomposition heuristic for large instances. The computational results confirm that incorporating customer choice behavior can increase profits. The effect is amplified when preferences are more heterogeneous. The authors also investigate the impact of predefined service requirements on profits and find that an inaccurate specification can reduce profits. Hernandez et al. (2017) consider independent demand but account for interdependencies between service assortments over consecutive days. Thus, the assortment decision does not decompose by shift, and the authors use a periodic vehicle routing approach to assign weekly assortments to geographic areas. Routes are modeled at the aggregated area level rather than at individual customer locations. The computational study focuses on the performances of

![Table 3](image-url)
two tabu search-based solution methods, which are also compared to an exact solution method.

In another subset of papers, uncertainties in demand forecasts are explicitly considered. Bruck et al. (2018) discuss the business case of an Italian gas provider that cannot apply operational demand management but must ensure service to all customers at regulated prices. The authors make assortment decisions by assigning capacities (i.e., technicians) to a given pool of time windows and ensure service quality by balancing the assortment over all the days of an operating week. The customers’ time window choice is uncertain yet independent of the assortment offered. The authors incorporate the stochastic choice in a simulation stage that is part of a two-stage stochastic program. Combined with a multi-depot multiple TSP, this stage enables the evaluation of first-stage assortment decisions. Using real-life booking instances of the industry partner, the authors demonstrate that their method reduces delivery and penalty costs compared to the company’s manual process. Coté et al. (2019) extend the degree of uncertainty to customer locations, basket sizes, and service times. They evaluate an assortment’s delivery and penalty costs in the second stage of a two-stage stochastic program using a vehicle routing approach that accounts for multiple interrelated periods. The authors perform a computational study on real instances of a Canadian retail company, the results of which show the effectiveness of their method, which outperforms the manual solution obtained by the company.

Assortment decisions for individual customers. The third cluster is concerned with the assignment of single service options to individual customers, which can be interpreted as long-term customer agreements – a special case of service differentiation. The set of customers is fixed and known in advance, and all customers have to be served.

Spliet & Desaulniers (2015); Spliet & Gabor (2015) consider a business-to-business (B2B) case inspired by a Dutch retailer. In this context, ‘customers’ refer to retail stores that are replenished periodically. The supplier assigns to each store a time window in which it will receive deliveries. This assignment decision is driven by stochastic demand volumes. The authors present a two-stage stochastic linear program that evaluates assignment decisions based on a vehicle routing model. The objective is to minimize delivery costs subject to the stores’ preferred delivery time intervals (Spliet & Gabor, 2015) or candidate options (Spliet & Desaulniers, 2015). Both formulations are solved to optimality using a branch-and-price-and-cut algorithm with route relaxations. In subsequent work, Spliet et al. (2018) add time-dependent travel times and seek arrival time consistency. The authors propose an exact solution method and evaluate its performance.

4.2. Tactical pricing

We define tactical pricing as the planned differentiation of prices across both customer groups (e.g., by geographic location or order value) and service options (e.g., premiums for evening delivery). While tactical offering limits an assortment’s breadth, tactical pricing steers customers to favorable options within a (potentially broader) assortment. As seen in the lower part of Table 3, we are aware of one single publication focused on tactical pricing.

Klein et al. (2019) consider price differentiation between time windows offered in given geographic areas, with the objective of maximizing total profit. Assortments are fixed, but prices can be selected from a finite price list. Akin to the majority of operational pricing studies, the authors explicitly model customer choice behavior based on random utility theory. Specifically, they apply a non-parametric rank-based model that captures a customer segment’s choice behavior through preference lists over all possible service options, including non-purchase. The authors formulate the pricing problem as an MIP that either features aggregate vehicle routes or cost approximations with respect to the geographic areas. The computational results confirm the benefits of differentiated pricing over uniform pricing. For industry-sized instances, the authors recommend their approximation-based approach since it is able to find good solutions in a limited amount of time.

5. Strategic demand management

The studies on operational and tactical demand management discussed in the preceding sections make assumptions regarding the setting defined by strategic-level decisions. These include decisions on the service region, appropriate service segments, the service design, and the pricing model. Interestingly, publications that address these decisions in their own right are few and far between. Therefore, rather than creating a literature table similar to those in Sections 3 and 4, we present the problem- and methodology-related focus of the current state-of-the-art literature on strategic demand management at a glance in Table 4. We discuss the relevant aspects of key strategic planning tasks and contextualize current perspectives in the literature. As in the preceding sections, we distinguish between offering and pricing levers.

5.1. Strategic offering

Strategic offering refers to identifying target markets and designing an appropriate service proposition, which translates to three major planning tasks that guide our discussion: The selection of the service region, service design, and the definition of service segments (cf. Roth & Menor, 2003).

We start with the literature that sheds light on the choice of service region. Here, a decision has to be made whether to offer service in a densely or sparsely populated area. The former includes mostly metropolitan areas and inner cities with dense road networks and high demand potential but also more fierce competition. The latter is characterized by sparser road networks and lower customer density but may allow the retailer to achieve a monopoly. In this vein, several studies have examined the operational implications of urban and rural service regions (Belavina et al., 2017; Boyer et al., 2009; Lin & Mahmassani, 2002; Ramaekers et al., 2018) and conclude that customer density has a significant positive effect on route efficiency. Beyond strategic demand management literature, Jiang et al. (2019) discuss general challenges of last-mile delivery in rural, more sparsely populated areas. In the operational demand management literature, Ehmke & Campbell (2014); Köhler & Haferkamp (2019) show that the characteristics of the service region also influence which real-time order evaluation method is most appropriate.

Second, we consider the literature addressing service design. This planning problem refers to a broad spectrum of design elements that characterize a delivery service offer and its service level. This includes decisions on delivery speed (e.g., click-to-door time), precision (e.g., time window length), and service frequency. Further design decisions concern possible interactions between service assortment and physical assortment, customer flexibility in terms of changes in the time window and shopping basket, and value-added services such as returns management. To gain a competitive advantage, it is important to understand both the sales impact and operational implications of different service designs (Amorim et al., 2020). Thus, on the one hand, many empirical studies have investigated customer preferences and expectations regarding particular delivery service attributes (Amorim et al., 2020; de Magalhães, 2021; Milioti et al., 2020; Wilson-Jeanselme & Reynolds, 2006). Most recently, Rodríguez García et al. (2022) present a framework on how to map value proposition to logistics strategy, thereby qualitatively assessing operational impli-
cations of a service design. All of these studies shed light on how service design attributes affect the generated demand volume.

On the other hand, there is a wide field of exploratory research that examines the operational implications of a service design. Starting in the early 2000s, Lin & Mahmassani (2002) show by simulation that increasing the time window length can reduce vehicle idle time, lower total miles traveled, and allow for more customers to be served. Boyer et al. (2009) support their results, and Ramaekers et al. (2018) report similar effects for both delivery and assembly operations. Ulmer (2017) focus on the impact of offering delivery deadlines, and Manerba et al. (2018) investigate both click-to-door time and time window length from an environmental perspective. Agatz et al. (2011) perform a sensitivity analysis on the choice of service frequencies, and Mackert et al. (2019) show that an inadequate specification can reduce profits. Very recently, Phillipson & van Kempen (2021) have assessed the cost implications of allowing customers to change their chosen time window before the delivery day, and Fikar et al. (2021) have examined the integration of product shelf-life options into demand management decisions. Some of these findings have already been picked up in operational demand management: Casazza et al. (2016) perform dynamic service design adjustments, and Campbell & Savelbers (2006); Köhler et al. (2020) offer and price time windows depending on their length.

Lastly, we present literature that concerns defining appropriate service segments which form the basis for tactical service differentiation (see Section 2.2.1). It should be noted that these segments do not necessarily coincide with the customer segments used to capture different preference structures within customer choice models. Tactical demand management commonly assumes given service segments based on geographic characteristics such as a customer’s zip code affiliation; only Cleophas & Ehmke (2014) additionally group customers based on their basket value (see Table 3). We are aware of just a single contribution that determines optimal service segments in this context. Bruck et al. (2020) extend the tactical approach of Bruck et al. (2018) and integrate strategic offering. They determine optimal service segments by solving a P-median facility location problem to group municipalities within the considered service region. A service constraint handles potential imbalances among segments’ total expected demand. The authors evaluate their approach using real industry data and emphasize its value for assessing entry into new service regions and analyzing past service segment configurations.

5.2. Strategic pricing

Strategic pricing refers to the overall pricing model and depends on the competitive environment, customer preferences, and price sensitivities within the target market. Determining a pricing model includes decisions about free or paid delivery, whether to use a delivery charge per order or a subscription fee per service period, and other incentive schemes. Tactical and operational demand management commonly assume a per-order pricing model within a given price range to steer customer choice. However, we are aware of several studies that shed light on the impact of specific pricing models.

Belavina et al. (2017) consider grocery delivery and build a stylized model to examine per-order and subscription-based pricing models with respect to equilibrium customer behavior and resulting profit and environmental performance. Their results show that subscription-based models lead to more frequent delivery requests, which in turn impact the provider’s revenue, route efficiency, and food waste. The authors conclude that the subscription model tends to be more environmentally friendly because the reduction in food waste emissions outweighs the increase in delivery emissions, but they still recommend the per-order model for high-margin providers that operate in sparsely populated areas. Wagner et al. (2021) show that on average, the increased order frequency entails a profit loss as the increase in assembly and delivery costs outweighs the increase in revenue. The authors explain this effect as a result of higher expectations of subscription customers; i.e., they choose narrower and more popular time windows. In addition, the authors develop a data-driven algorithm that predicts the expected post-subscription profitability to determine whether a particular customer should be offered a subscription plan. The algorithm is trained and evaluated based on real order data from a large omnichannel grocery retailer. The authors report that observed product assortment size and basket value are the strongest predictors of post-subscription profitability. In contrast, Gümü̈ş et al. (2013) investigate the joint design of a pricing model for product and delivery service. They analyze the competitive dynamics of price partitioning, where delivery and product prices are displayed separately in a partitioned setting, and free shipping is advertised in a non-partitioned setting because the delivery cost is already included in the product price. The authors determine the equilibrium market structure and validate their theoretical results through empirical analyses. In addition to traditional pricing models, Agatz et al. (2021) focus on non-monetary incen-
tives and study the impact of displaying green labels for environmentally friendly service options on customer behavior and operational performance. From their empirical experiments and simulation study, the authors verify that green labels effectively steer customer choice, also in combination with price incentives and for less attractive time windows.

6. Discussion

In this section, we synthesize our findings from reviewing the literature, highlight key challenges and potential future research for each planning level, and elaborate on the connection between the planning levels.

There is a growing number of academic contributions on operational demand management, predominantly directed at e-grocery. The computational challenges make it an active field of research in operations research. Most work in this area focuses on sophisticated solution methods for specific parts of the real-time decision problem, e.g., feasibility assessment, value anticipation, or customer choice behavior. In general, vehicle routing heuristics and dynamic programming can be identified as methodological cornerstones.

Building on the current body of research, we see several avenues for future research. First, given the modular structure of operational decision-making, there is a need for comprehensive benchmarks that guide the selection of suitable building blocks of solution methods. Lang & Cleophas (2020); Ulmer (2019) offer valuable starting points for this purpose. Second, in light of very limited computation time, there is still a need for fast solution methods. One potential research avenue is the application of machine and reinforcement learning in this context. Such methods have already been adapted for feasibility assessment (van der Haegen et al., 2022) and value anticipation (e.g., Koch & Klein, 2020) but have not yet been applied to predict customer choice. Alternatively, it may be beneficial to change the fulfillment process design to simplify operational planning. We see valuable starting points in the recent literature: Schwamberger et al. (2022) define an inverted order capture process in which the service provider proactively approaches customers with the opportunity to place an order, and Yildiz & Savelbergh (2020) explore the possibility of incentivizing accepted customers to change their chosen time window after the order capture cut-off time.

We see fewer contributions to tactical demand management that, however, cover a variety of planning problems from long-term customer agreements to short-term availability control. From a methodological perspective, MILP, two-stage stochastic programming, and simulation are prevalent and customer choice behavior is rarely modeled explicitly. Besides, we observe that tactical approaches are mainly tailored to specific business sectors and that the research is often conducted in collaboration with an industry partner, which indicates the practical relevance of the topic.

We see a need for future research, especially for innovative AHD concepts. Service providers that perform order-based fulfillment within a deadline benefit from tactical offering and pricing decisions: Different delivery deadlines can be offered in different geographic areas at different prices (e.g., longer and/or more expensive deadlines in peripheral areas). Stroh et al. (2021)’s tactical vehicle dispatch policies may serve as a starting point. Moreover, there is great potential for tactical offering under a subscription-based pricing model. Spliet et al. (2018); Spliet & Desaulniers (2015); Spliet & Gabor (2015) provide relevant insights from the business-to-business context that can be transferred to customers who are allowed to reserve a time window as part of their subscription plan.

Contributions to strategic demand management provide insight into many different aspects of strategic planning. The set of applied methodologies is much more diverse which we explain by the strong interdependencies with other domains. For example, selecting a service region interacts with location planning, determining service segments is influenced by delivery districting (e.g., Banerjee et al., 2022; Haugland et al., 2007), and service design and pricing models strongly depend on marketing and competitive considerations. As a consequence, we see that comprehensive decision support is still missing. Other reasons that might promote this gap are that (i) strategic demand management decisions are considered to have less leverage compared to strategic decisions in other research fields (e.g., network design) since they are less long-term and more easily reversible. (ii) Competitive constraints may leave only limited room for optimization. (iii) From a practitioner’s perspective, decision-making responsibilities are more dispersed and located at a higher managerial level than they are for tactical and operational demand management.

We see the opportunity for strategic demand management to provide comprehensive decision support to capture the greatest possible demand potential and to do so profitably. Thereby, important issues of competitive pressure and market share should also be addressed. Looking to adjacent research fields confirms this potential. Metters & Walton (2007) provide strategic decision support by proposing a service sector typology for multi-channel e-tailing. They develop a matrix of competitive positions along the dimensions of inventory pooling and shipping consolidation, and identify four types of strategies that can be adopted by multi-channel e-tailers. The authors also emphasize that e-tailers should align their supply chain configuration with their strategic objectives. For the express delivery business sector, Li et al. (2021) propose a two-dimensional decision matrix to select the most suitable delivery service mode among direct and indirect options. They measure the expected customer utility and calculate the expected cost of delivery service to map different service modes to the decision matrix.

We conclude our discussion with a few observations concerning the interaction between the different planning levels reviewed separately in Sections 3–5. Conceptually, longer-term decisions set the boundaries for decisions on the shorter term. One challenge is that actual performance can only be observed once orders materialize. Appropriately anticipating this performance impact is a core issue for long-term decisions. Given the scarcity of strategic demand management research highlighted in Section 5, the impact of corresponding long-term decisions on tactical and operational demand management is largely an open issue to date. Most contributions to the tactical and operational literature make assumptions on the strategic system design, based on choices observed in practice. However, the appropriateness of these choices, including the service region, service design, and service segments has received limited attention thus far.

As a potential starting point for future research in this direction, some studies consider the sensitivity of tactical or operational decisions and their performance to changes in selected strategic choices. Examples are strategic choices between suburban and downtown service regions (e.g., Ehmke & Campbell, 2014) and between different time window lengths (e.g., Campbell & Savelbergh, 2005; Côté et al., 2019). Conceptually, these studies follow a what-if approach to strategic-level decisions. A next step would be to turn the analysis into a systematic optimization approach that selects strategic options based on their impact on day-to-day operations and performance. For example, Agatz et al. (2021) conducted operational-level simulations to assess the potential of new ways for steering customer behavior. Their strategic concept of green labels can be incorporated in tactical and operational pricing, complementing the current monetary incentives.

Interactions between the tactical and operational planning levels have received more attention in the literature. This is primarily driven by the fact that operational demand management decisions must be made in real time to facilitate a smooth order capture pro-
cess. This limits the available time for computations on the operational level. There is, however, more time to support tactical decisions. We observe two approaches in the literature that exploit this relation.

First, tactical decisions can pre-structure and thereby simplify operational decisions by limiting the decision space on the operational level. In the reviewed literature, this holds true for service and price differentiation. To be effective, such approaches must capture the link with the operational level. The extent to which this is the case depends on the decision-making flexibility assumed at this level. Long-term service agreements (e.g., Spliet & Gabor, 2015), legal regulations (Bruck et al., 2018), or business policies (Côté et al., 2019) may severely limit operational levers. In these cases, we observe more accurate routing formulations and the use of two-stage stochastic programs to hedge against forecast errors. If, on the other hand, operational demand management opportunities are more extensive, the demand model and operational impacts are more coarsely estimated (Agatz et al., 2011; Hernandez et al., 2017; Klein et al., 2019; Mackert, 2019). However, operational performance may be tested outside of the decision model, through simulation studies (e.g., Agatz et al., 2011).

Second, it may be beneficial, or even necessary, to shift some decisions from the operational to the less time-constrained tactical planning level altogether. Essentially, this implies a choice between an elaborate ex-ante planning model and a simpler heuristic using real-time information. Given the discussed computational limits, it makes sense to reserve real-time planning to those decisions for which the available real-time information really makes a difference. One example of shifting decisions to the tactical level is the ex-ante calculation of availability controls such as booking limits for specific time slots (Cleophas & Ehmke, 2014). Corresponding literature uses simulation and two-stage stochastic programming to capture the effects on the operational level (Cleophas & Ehmke, 2014; Visser & Savelsbergh, 2019).

7. Conclusion

This review paper introduced a framework for classifying demand management decisions for AHD with respect to different planning levels and demand management levers. For each planning level, we presented and classified prescriptive analytics methods in the literature and identified research gaps. The following are our main observations. We have seen a rich set of studies on operational demand management, aimed at extracting the greatest potential from real-time decisions. Because manifold opportunities for real-time decision-making differentiate AHD from traditional brick-and-mortar retail, the appeal of this line of research is intuitive. The ensuing computational challenges have triggered sophisticated algorithmic contributions. However, all decisions clearly do not benefit equally from real-time information. In this light, we see yet unlocked opportunities for tactical demand management to simplify and prestructure operational decisions. Finally, there is a striking lack of research on underlying long-term, design-level decisions. Hence, we see great potential for future contributions to strategic demand management for AHD.

Taking a more general perspective, we highlight four topical themes that we believe hold opportunities for innovative and relevant future research on demand management for AHD. These themes give rise to novel analytics issues at all planning levels.

First, a natural direction concerns innovative business models and services in AHD. While research on standard ‘next-day’ grocery delivery is maturing, researchers have only started to study new delivery trends. On the one hand, on-demand e-grocery startups (e.g., Gorillas and Flink) promise ‘instant’ grocery delivery within a few minutes. This fundamentally different service offering challenges many assumptions of the current fulfillment strategies and corresponding demand management. On the other hand, established businesses are exploring novel customer interaction processes that deviate from the current standard process reflected in Section 2.1. Examples include long-term subscription agreements and proactive customer contacting. These developments give rise to novel decisions and call for corresponding analytics models and approaches.

Second, more research that addresses new objectives in demand management for AHD is needed. To date, the majority of publications focus on profit maximization as the primary goal of service providers. Given the expansion race between emerging on-demand e-grocery businesses, research should recognize market share as a relevant alternative objective. Furthermore, considering environmental objectives has become a standard in many research fields, and delivery services are subject to particular public scrutiny with regard to sustainability (Siragusa & Tumino, 2022).

Belavina et al. (2017); Manerba et al. (2018) are the first to investigate the leverage of demand management in light of environmental objectives. Future research should expand this development and explore the impact of multiple conflicting objectives, for example, related to social responsibility toward internal stakeholders (e.g., delivery workers) and external stakeholders (e.g., customers, residents, and administrators). Recent literature has underlined the relevance of this perspective: Belanche et al. (2021) show that customers’ purchase intentions depend on their perception of the working conditions for delivery workers, Chen et al. (2022); Soeffker et al. (2017) investigate demand management regarding fairness to customers, and Bjørgen et al. (2021) discuss the integration of e-grocery logistics into urban spaces. The rapid expansion of micro depots to support instant grocery deliveries, so-called ‘dark stores’, have already sparked public and political debate: The Dutch cities of Amsterdam and Rotterdam recently restricted the opening of new facilities because of noise and the blocking of pedestrian walkways.

Third, we see potential for demand management addressing the interaction between the delivery service and the product assortment. Fikar et al. (2021); Gümüş et al. (2013) provide initial work in this direction. Future research may strengthen the integration of product assortment-related aspects into demand management and extend demand management levers accordingly. For example, while existing levers have been shown to effectively reserve fulfillment capacity for more valuable customers, the inventory rationing literature demonstrates a similar effect with respect to product availability by reserving inventory for high-margin customers (e.g., Jimenez G et al., 2020). In addition, integrating the product assortment naturally draws attention to the order assembly process. We have seen few contributions that explicitly account for order assembly in demand management methods. Among those is research exploring the impact of time windows on both assembly and delivery (Ramaekers et al., 2018) and research presenting operational ordering for order-based fulfillment (Azi et al., 2012; Klapp et al., 2020). Product-related demand management requires new analytical models and approaches that enable integrated decision-making at all planning levels.

Fourth, we call for more empirical validation of demand management for AHD. On the one hand, we recognize that results based on empirical instances alone are difficult to generalize and should therefore be supported by carefully generated synthetic data. The classification of demand data presented in Tables 2 and 3 is intended to shed light on this crucial aspect, even though the observed situation is more nuanced than a strict dichotomy. While research on supply-oriented levers can more easily base the computational results on synthetic instances, empirical data are partic-

---

ularly important for demand management because of the strong role of customer interaction in this context. Many of the assumptions required for demand management relate to customer behavior, which is difficult to model realistically without empirical data. In addition, customer behavior changes over time, so empirical validation should be reviewed regularly.

To conclude, we expect demand management for AHD to continue to gain importance and to witness significant innovations to emerge. We hope that this review contributes to stimulating future research into this dynamic field.

References