Optimizing Visitor Load at National Parks: Balancing Accessibility and Overcrowding through Online Reservation Systems

Yamit Leon, Galit B. Yom-Tov

Technion—Israel Institute of Technology, yamitleon07@gmail.com, gality@technion.ac.il

Problem definition: The management of national parks (NPs) involves striking a delicate balance between conserving nature and its inhabitants by limiting human access versus promoting awareness of the park's wonders by allowing access to designated trails and public areas. Many park authorities (PA) restrict access by establishing visiting hours and limiting the number of visitors allowed to enter per day. Such limitations are managed through a reservation system, requiring visitors to reserve a visiting permit before arrival. We analyze reservation system data provided by the PA of Israel, showing cancellation and no-show behaviors that change dynamically over time and depend on the number of days a reservation is made before the visiting day.

Methodology/results: We develop a dynamic policy for the number of reservations the system should be allowed to make every day for a specific focal date. The solution depends on the ratio between the cost of not allowing visitors to make a reservation to the park ("blocking cost") and the park's overloading cost normalized by the probability of a reservation being realized (due to cancellations or no-shows). The policy takes the form of a two-threshold policy, where the first is a time threshold determining periods where reservations are unrestricted, and the second is a capacity threshold limiting the total number of reservations made over the entire reservation horizon. We simulate the performance of our fluid policy and compare it to several benchmarks, showing that our proposed policy is the only one inducing minimal costs in all scenarios.

Managerial implications:

We show how our policies can inform NP authorities regarding the number of days in which the reservation system should be opened for reservations before the focal day. Our model can help NP to better conserve natural resources and provide access to public spaces in a balanced way.

---

1. Introduction

Managing national parks (NPs) involves a delicate balance between protecting nature and its inhabitants and promoting awareness through public access (Cole 2012, 2019, Yung et al. 2010). Traditionally, this balance has been achieved by restricting human access both to designated trails and to specific visiting hours. However, increasing population growth and rising tourism have led to a significant surge in visitors to NPs. As a result, it is now common to encounter traffic jams at park entrances and overcrowding on popular trails during the summer and holiday periods. These challenges pose risks to the parks' ecosystems (Hammitt et al. 2015, Monz et al. 2013, Pickering 2010, Cole 1990). Therefore, the issue of limiting daily visitor numbers is raised annually in many popular parks worldwide. However, denying entry to visitors at park gates is often viewed as overly restrictive of individual rights. Consequently, such measures are implemented only in extreme situations where public health is at risk, such as during wildfires. Instead, many NPs around the world balance nature conservation with visitor management using reservation systems through which visitors can ensure their accessibility to the park on a certain day while the park authority (PA) manages the visitor load.

For example, in the US, the federal government manages a travel planning and reservation platform called recreation.gov through which visitors can reserve park permits or participate in a lottery for popular trails. In New Zealand, a permit is required to hike the Great Walks. In Israel, the trigger to implement such a reservation system was the COVID-19 pandemic, which transformed the issue of overcrowding to a public health concern. In 2020, the World Health Organization recommended limiting gatherings in both enclosed and open spaces. In response, the Israel Nature and Parks Authority (INPA) launched a preregistration system in May 2020 to regulate park access. Under this system, visitors were required to reserve a specific date and, in some cases, a specific entry time to visit a NP[1]. No fee was charged for obtaining an entry permit. As in the US, this system limits the maximum number of visitors based on the estimated capacity of each park. The INPA continued to use these preregistration procedures even after the pandemic stopped, primarily in smaller parks (e.g., Ein Yehuda) where overcrowding remains an issue. To this day, the system remains active but is not mandatory anymore in most NPs. We claim that a lot can be learned from the period when the park registration system was mandatory, ranging from how to manage NP capacity, to the public's reaction to reservation restrictions.

Managing reservation systems for NPs presents several operational challenges. Some of these challenges are similar to those faced by other reservation-based service systems. For instance, data from the INPA for 2019–2020 (see Section 3) shows that approximately 19.5% of individuals who schedule a visit cancel their reservation, and another 29.3% neither arrive at the park nor cancel their reservation (no-shows). The latter statistic is akin to the no-show rates observed in healthcare appointments (Liu et al. 2010). There are, however, differences between the healthcare and park settings. For one, while patient cancellations and no-shows affect a physician's work schedule and one another, they have no such effect in a park setting, where there is no specific "server" attending to visitors.

Determining NP capacity is a complex problem in and of itself (Whittaker et al. 2011). It is influenced by a variety of social and ecological factors including current environmental conditions;

[1] For brevity, we include Israel's nature reserves and national parks in the umbrella term "national park (NP)".

ecosystem type; the levels, timing, and type of visitor use; and visitor behavior (Hammitt et al. 2015, Monz et al. 2013, Pickering 2010). The PA can influence some of these factors through development or actions. For example, the PA can create elevated paths to protect the natural environment against impact, direct visitors to specific areas of a park, limit park use, or encourage minimum-impact visitor behavior (Cole 1990). Furthermore, visitor load itself affects the way visitors utilize recreation areas (D'Antonio and Monz 2016, Cole and Hall 2010).

In this paper, we will address the following research questions:

1. *Discover similarities and differences between NP visitor behavior and known customer behavior in other reservation systems.* Previous literature has analyzed customer behavior in reservation systems for various sectors, including hotels (Dole 2023), healthcare (Feldman et al. 2014), and flights (Lawrence et al. 2003). These studies identify phenomena such as reservation cancellations and no-shows. In this paper, we utilize data from the INPA and observe similar patterns of time-varying dynamics of cancellation and no-show behaviors across parks (Section 3). Moreover, in NPs that applied a daily quota of reservations with no hourly limitations, we observe a time-varying arrival-rate pattern of visitors to the park that is similar to other service systems (e.g., healthcare (Yom-Tov and Mandelbaum 2014) and call center (Gans et al. 2003)) without reservations.
2. *How do we determine optimal park capacity?*

   We propose to analyze visitor load at NPs as an $M_t/G/\infty$ queue with time-varying arrivals and general LOS (Section 4.1). Based on fluid approximation, we determine the optimal number of visitors that balances overcrowding and accessibility. By "accessibility" we mean to allow as many as possible visitors to enter the park, which holds both economic and educational significance. By "overcrowding," we refer to situations when the number of visitors exceeds the maximal capacity that was predefined to a specific park.
3. *How do we optimize a park reservation system?*

   The main contribution of this paper is in optimizing the NP reservation system while considering behavioral factors, such as no-shows and cancellations. Here, we strive to balance not only accessibility and overcrowding at the park but also the problems resulting from requiring people to reserve a permit before their visit. This affect is captured via dynamic blocking costs, where the system is penalized for preventing potential visitors from making a reservation, and that penalty depends on the time left till the requested visiting day. In Section 5, we analyze the reservation system optimization problem at the fluid level, proving that the system's reservation capacity is determined by the ratio between the blocking and overcrowding costs and is influenced by the probability of a reservation being realized (due to cancellations or no-shows). The fluid policy takes the form of a two-threshold policy, where the first is a time threshold determining periods where reservations are unrestricted, and the second is a capacity threshold limiting the total number of reservations made over the entire reservation horizon. This analysis has implications for determining the time when the PA should begin accepting reservations for a specific focal visiting day (Section 5.3).

To the best of our knowledge, this is the first paper to analyze the effectiveness of a national park reservation system as a service system, applying operations research methods.

The rest of the paper is organized as follows. In Section 2, we review relevant literature on managing visitor load in NPs and optimizing reservation systems. In Section 3, we present descriptive data analytics of Israeli NPs based on data provided by the INPA. In Section 4, we discuss how to determine park capacity and connect this to determining the reservation system's acceptance policy. In Section 5, we develop a mathematical model for a reservation management system, propose a policy that is optimal for the fluid approximation version of this model, and discuss managerial implications. In Section 6, we simulate the performance of this fluid policy and compare it to several benchmarks, showing that our proposed policy is the only one achieving minimal costs in all scenarios. In Section 7, we discuss avenues of future research and conclude.

2. Literature Review

Here we review relevant literature. Most of it was developed for other types of service systems, such as healthcare and flight companies, or for tourism and leisure systems, such as hotels and theme parks. We explain the differences when relevant.

2.1. Managing Visitor Load in Parks and Tourism Areas

National parks are characterized by a time-varying arrival rate to the park (see Section 3). Time-varying arrival rates are typical for many service systems, creating a time-varying customer load that increases and then decreases during a working day, unless server capacity compensates for the variation. NPs are somewhat different from typical service systems because NP capacity is constant and cannot be changed throughout the day; a park's capacity is determined by its individual characteristics (e.g., the park's sitting areas and trails (Hammitt et al. 2015, Monz et al. 2013, Pickering 2010)) that do not change during the day. These facts not only reinforce the time-varying dynamics of visitor load in NPs, but also limit the PA's ability to manage this load. As explained, what the PA can manage is the total number of visitors entering the park, by limiting the number of permits available in the reservation system. Hence, while we take the arrival-rate time-varying pattern as given, the total number of arrivals we wish to allow during a day is viewed as a decision variable. A given park capacity and the time-varying patterns of arrivals and LOS are important factors in determining a target total number of visitors we wish for the reservation system to allow to visit the park during the day. This could be viewed as the inverse of the mathematical problem studied by Zychlinski et al. (2020), who determined the optimal fixed capacity that minimizes under-utilization and overload costs in healthcare systems for given time-varying arrival rate and service times. Yet, our problem does not have a close form solutions.

A prerequisite to solve such problems is to be able to forecast demand—how many people would wish to visit the NP, as done for campground or tourism areas using historical data (Rice et al. 2019) or web search (Peng et al. 2017, Law et al. 2019). Dependencies between parks can be an important aspect (see Section 7 and Appendix EC.1). As mentioned, the park's individual characteristics of trails and leisure areas (Hammitt et al. 2015, Monz et al. 2013, Pickering 2010) determine its capacity. Load within the NP depends on how people use these areas—are they sitting for a picnic or moving through a trail? Hence, load depends on the park layout and design. Meijles et al. (2014) analyzed visitor spatial flow and overcrowding patterns in a NP using GPS data. They suggested that visitor tracking can be used to steer visitors to less overcrowded areas. Similar suggestions were made by Ahmadi (1997) who analyzed visitor movements in a different leisure industry, theme parks. They suggested that such analysis can shed light on the efficient spatial design of the theme park's rides and attractions as well as its impact on load, wait times, and visitor experience (Ahmadi 1997). By contrast, NPs have less control over the spatial design itself, and rather try to leave the park environment with minimal disruption. PAs focus on developing safe walking trails and providing rest areas. Their location is very much determined by the park terrain.

2.2. Reservation Systems

The above challenges have led many PAs to install reservation systems to control load. Reservation systems are widely used not just in NPs and other tourism sites but also in healthcare facilities (where they are usually referred to as appointment systems). A review of the literature on appointment systems is given in Mondschein and Weintraub (2003) and Pinedo et al. (2015); Cayirli and Veral (2003) and Gupta and Denton (2008) provide excellent surveys on appointment systems in healthcare specifically. As Pinedo et al. (2015) showed in their review paper, a common theme across varying industries is to either minimize the costs or maximize the gains of the appointment system. However, the authors highlight that the efficiency of appointment systems is highly dependent on each service industry's context-specific constraints and objectives.

The literature on appointment systems in healthcare usually considers intra-day scheduling aiming to minimize patients' waiting time and staff costs and idle time. The literature also takes into account customer no-shows and cancellations, as these reduce system efficiency and revenue (Hassin and Mendel 2008, Zacharias and Pinedo 2013, Hassin and Mendel 2008). One way to deal with no-shows and cancellations is to book more customers than the system can handle. For example, Zacharias and Pinedo (2013) model an appointment reservation system that overbooks appointments using an index policy that is based on patients' no-show characteristics. A different approach is to optimize the intra-day sequence of appointments in order to minimize the impact of no-shows on staff utilization. For example, Hassin and Mendel (2008) suggest sequencing appointments in varying intervals throughout the day (e.g., have lower number of appointments at the beginning and the end of a day compared to the middle of the day). Mandelbaum et al. (2019) consider the joint problem of determining the appointment date, appointment capacity, and appointment sequencing throughout the day, taking into account patient punctuality and service duration variability. Wang and Fung (2015) suggest a dynamic programming model to optimize appointment scheduling that takes into account patients' preferences for a particular physician and time slot. This approach is reinforced by research showing a correspondence between honoring patients' preferences and their no-show and cancellation behaviors (Liu et al. 2018). Our problem is different from the above research because a NP is not a service system with strict staffing capacity, where one customer waits for a previous customer to finish service before starting service themselves. Instead, all visitors enjoy the park at the same time. Hence, there are no delays caused by the exact arrival time of visitors. This lack of a strict capacity limit led us to model the NP as an infinite-server system (Section 4.1), an approach suggested for large healthcare appointment systems by Mandelbaum et al. (2019) and Huang et al. (2022).

A key factor in estimating no-show and cancellation probabilities in a medical context is that such behavior increases as a function of the time difference (in days) between the day the appointment is made and the appointment day (Green and Savin 2007, Liu et al. 2010, Norris et al. 2014, Feldman et al. 2014, Leeftink et al. 2022). We observe the same dynamics, where no-show and cancellation probabilities increase with that time difference (Section 3). Relatedly, in the context of restaurant reservations, Alexandrov and Lariviere (2020) show that not having a reservation system is generally the worst policy, since reservations may increase demand on slow nights when demand is naturally low. We also consider the reservation horizon length problem, as Leeftink et al. (2022) did for healthcare clinics, first analytically in Section 5.3, where we determine the conditions under which one should allow the reservation horizon to be as long as possible vs. as short as possible, and then numerically, in Section 6, where we use simulation to compare our reservation policy to a no-reservation policy.

In our model, the system may block customers from making a reservation, and this blocking cost plays an important role in our model and policy. Most of the above-mentioned healthcare papers do not consider blocking costs as part of their model, since healthcare systems usually do not prevent patients from making an appointment. An exception is Schütz and Kolisch (2013) who maximize revenue. They take into account blocking costs, overtime costs due to overbooking, and refunds for customers who cancelled their reservation or did not show up for their appointment. Refunds can be relevant in NPs in which fees are collected during the reservation process (e.g., in US), and are closely connected to lost revenues due to blocking demand. Overtime is not relevant in our model, because NPs do not have strict capacity that can cause queues and delays during the visiting day.

Our reservation system has also some parallels with reservation systems in the hospitality and tourism industry. For example, hotels and airlines also overbook to offset customer cancellations and no-shows (Lawrence et al. 2003). Like the NP case, hotels and airlines optimize the total number of arrivals during the focal day and care less about their specific arrival time. Unlike NPs, hotels and airlines have a strict capacity, and when the number of reservations that are realized on the focal day exceeds that strict capacity, high overbooking penalties are paid due to lost potential revenue.

An interesting question is how to balance long-term and short-term demand. For example, Bitran and Gilbert (1994) ask whether last-minute walk-in hotel reservations should be accepted by creating overbooking, that is, taking into account the probability of denying entry of a customer arriving to a full hotel with a reservation. They employ intra-day information to predict cancellations and no-shows and the resulting probability of reaching a full hotel. Grant et al. (2022) study appointment systems in healthcare, where the trade-off is between same-day appointments with extra office hour costs for the clinic vs. late appointments that might result in a deterioration of the patient's health condition. The question of balancing long- and short-term demand also arises in our context in two ways: (a) should we keep capacity for customers planning their visit just 1–2 days before the visiting day (Section 5.3), and (b) how should we use real-time predictions of no-show and cancellation probabilities (Section 6.1).

3. The Reservation System in National Parks

In this section, we first describe the general processes of managing visits to a NP and then the way these processes are reflected in data.

This research was done in collaboration with the INPA. INPA manages 400 nature reserves (NRs) and 81 national parks (NPs) in Israel, covering over 20 percent of Israel's land mass, including sites with historical significance or unique natural attributes. INPA provided us with data that includes park visits in 2019–2020 and park reservations in 2020. In 2020, the COVID-19 pandemic triggered a change in how the INPA managed visits to its parks. A reservation system was installed. Reserving an entry permit became the first step in planning a visit to one of the INPA sites. These reservations could be changed up to the day of the visit to the park.

Making a Reservation

Reservation are done through the PA website, where people can observe the park availability within the reservation horizon (the horizon length depends on the park). Each PA sets its own horizon for reservations. For example, one can reserve a permit to visit popular NPs in the US six months in advance (see www.recreation.gov/pass). For some NPs, entry permits are distributed through a lottery (see www.recreation.gov/lottery/available). INPA allowed entry permits to be obtained up to 15 days in advance during 2020; currently (as of October 2024) permits can be obtained up to 9 months in advance.

When INPA established the reservation system during the COVID-19 pandemic, the aim was to control visitor load at the NPs. Different policies where implemented in different parks. In most parks, visitors were allowed to enter the park at the hour of their choosing without any limit on visiting duration while the park was open for that day, but in some parks, visitors were required to specify an entry time (from 2–4 time-slot options), and in others, LOS was also limited to that time slot, particularly during weekends (see examples in Table 1). The time-slot partition resulted in queues at the park entrance gate around the beginning time of each time slot and, therefore, the time-slot limitations were eventually cancelled. Table 1 provides examples of visiting time slots and maximum capacity per slot at several parks in the north of Israel. The visiting time limitations could change between weekdays and weekends.

Table 1 Sample of visiting time slots of nature parks in the north of Israel [May–August 2020].

Site	Weekday		Friday		Saturday	
	Visiting time slot	Max. capacity	Visiting time slot	Max. capacity	Visiting time slot	Max. capacity
Hermon Stream (Banias) NR — Springs Area	8AM– 4PM	1200	8AM–10AM	400	8AM–10AM	300
			10AM–12PM	400	10AM–12PM	300
			12PM– 3PM	400	12PM– 2PM	300
					2PM– 4PM	300
Snir Stream NR	8AM–11AM	400	8AM–10AM	400	8AM–11AM	400
	11AM– 2PM	400	10AM–12PM	400	11AM– 2PM	400
	2PM– 4PM	400	12PM– 3PM	400	2PM– 4PM	400
Iyon Stream (Tanur) NR	8AM– 4PM	1200	8AM–12PM	600	8AM–11AM	400
			12PM– 3PM	600	11AM– 2PM	400
					2PM– 4PM	400
Yahudiya NR	8AM– 4PM	800	7AM– 3PM	800	7AM– 4PM	800

Analyzing the reservation data reveals interesting insights into visitor behavior. For example, Figure 1 shows the proportion of reservations made at day t before the focal day out of the total number of reservations made for that day. We observe that about 55% of visitors make reservations several days ahead of the focal day of the visit, while about 45% make reservations on the focal day of the visit or the day before. Naturally, the NP's specific features and accessibility affect the number of reservations; for example, 25% of all reservations during May–December 2020 were made to the five most popular NPs. Interestingly, 11.6% of the people reserved more than one NP for the same focal day, though it is unclear whether this phenomenon indicates tentative plans or an intention to visit more than one NP on the same day. See Appendix EC.1 for popular site combinations.

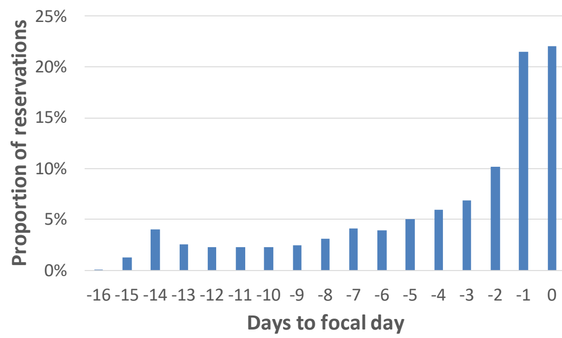


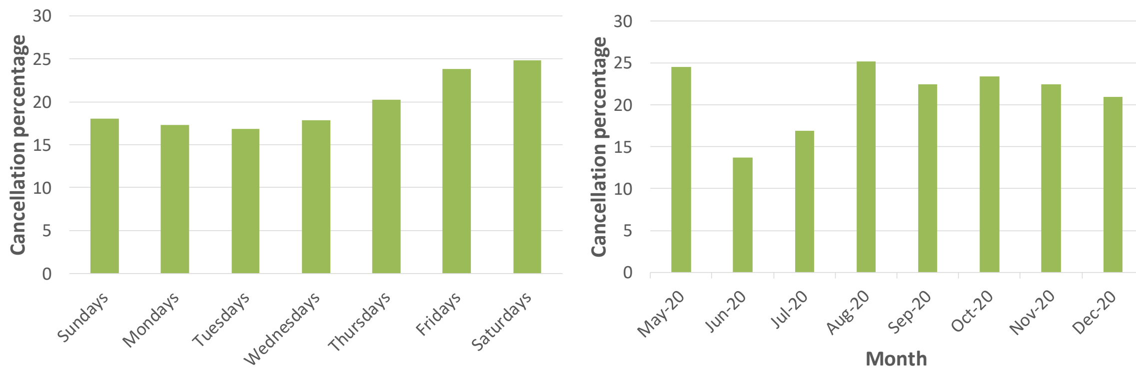
Figure 1 **Proportion of reservations made as a function of the distance from the focal day [All NPs, May–December 2020, All days]. Day 0 represents the focal day for which the reservations were made.**

Cancellations

Potential visitors may cancel or change their reservation (i.e., change the number of people in the group or the focal date of visit). Indeed, 19.5% (SD 15.8%) of the reservations in our sample were cancelled, and this percentage varied both by day of the week and by month (see Figure 2). Figure 2(a) shows that cancellations for weekends are higher than for weekdays: 23.8% (SD 13.7) and 17.6% (SD 16.1), respectively. Some of the variation in cancellations by month (see Figure 2(b)) may be attributed to variation in the reservation reminder system. Starting in June 2020, e-mail reminders were sent to the reservation holder 1–2 days before the focal day, and starting in November 2020, an SMS reminder was added. Reminders help to increase certainty over the focal day arrival rate, free reservation capacity for alternative visitors in case of cancellation, and reduce the no-show rate (since some people who would no-show without a reminder cancel instead).

Figure 3 shows the proportion of cancelled reservation at day t before the focal day from all cancelled reservations to that focal day. Most of the cancellations ($> 65\%$) were done on the focal day and the day before. Similar late cancellation behavior has been observed in healthcare systems. For example, Leeftink et al. (2022) find that about two-thirds of all appointment cancellations are cancelled less than five working days before the actual appointment date.

Figure 4 shows an estimation for the probability of a reservation being cancelled on each day before the focal day. This probability is estimated by the cancellation percentage: the number of cancellations from the number of reservations made at that day (the cancellations could have happened at any time from the reservation day until the focal day). We observe that the probability



(a) Day of week (b) Month

Figure 2 **Percentage of cancellations, by day of the week and month [All NPs, May–December 2020, All Days].**

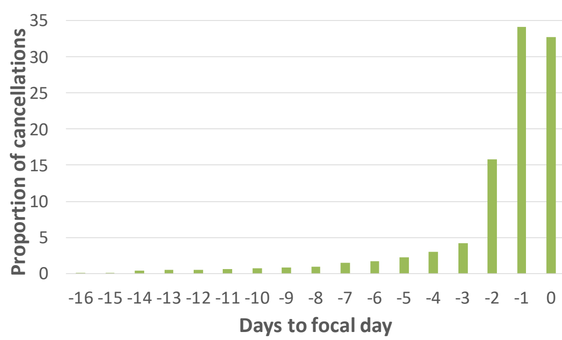


Figure 3 **Proportion of cancellations as a function of the time before the focal day [All NPs, May–December 2020, All days]. Day 0 represents the focal day for which the cancellations were made.**

of a cancellation increases with the number of days between the reservation day and the focal day. In particular, reservations that were made only 0–2 days before the visit day are significantly less likely to be cancelled. Similar behavior has been observed in healthcare, where the probability of canceling a physician appointment decreases with time leading up the appointment day (Gallucci et al. 2005, Liu et al. 2010).

Arrivals to the Park and No-shows

Analyzing the actual arrival rate to a park, we identify a clear time-varying pattern (that may differ between NPs). For example, Figure 5 shows the arrival rate (i.e., the proportion of arrivals at a specific hour out of the total arrivals during a day) in En Gedi Nature Reserve. The arrival rates peaks at the beginning of the day and at noon, then decrease toward the end of the day.

By comparing the number of active reservations (that were not cancelled before the focal day) to actual visits, we can estimate no-show probabilities. As in many other reservation systems, such as for doctor appointments, the no-show percentage is significant and has a direct impact on park accessibility. The no-shows are unrealized demand that may have prevented other potential visitors

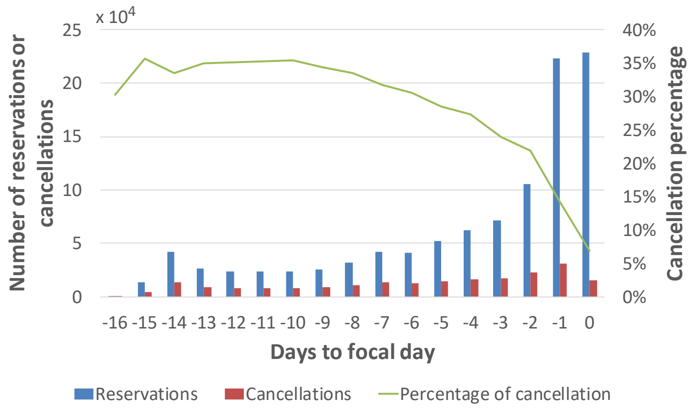


Figure 4 **Number of reservations (blue) and cancellations (red) as a function of the time before the focal day, and the percentage of cancellation from the reservations of that day (green) [All NPs, May–December 2020, All days]. Day 0 represents the focal day for which the reservations were made.**

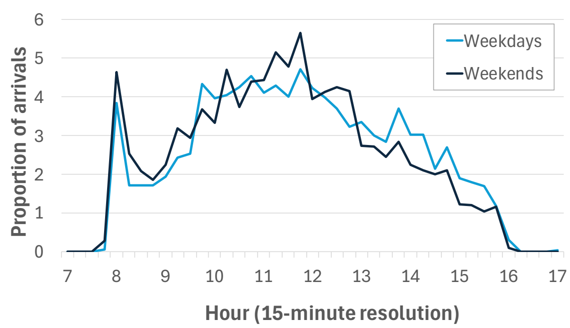


Figure 5 **Arrival rate to En Gedi NR by day of the week: Weekdays vs. weekends.**

from visiting the park by creating unused time slots. Figure 6, based on data from Gan HaShlosha (Sahne) NP, shows that no-shows are indeed a concern. The figure depicts the number of active reservations (solid orange line) and the number of visitors (solid blue line). The difference between them are no-shows. The number of active reservations for weekends reaches the maximum capacity of 2500 tickets (dashed black line), but the actual number of visitors is much lower. (Note that in August 2020, the park started to increase the maximum number of reservations that could be booked; accordingly, the number of actual visitors increased but was still below the maximum capacity of 2500 visitors on most days.)

Overall, the percentage of no-shows is 29.3% (see Figure 7(b)). The no-show probabilities seem to be higher for reservations for weekend focal days (see Figure 7(a)). Due to data limitations, we cannot observe the relation between the number of days before the focal date that the reservation was made and the no-show percentage. However, data from healthcare appointment systems shows a strong correlation between the time a reservation was made and the no-show percentage, where reservations that are made earlier are more likely to become no-shows (Feldman et al. 2014).

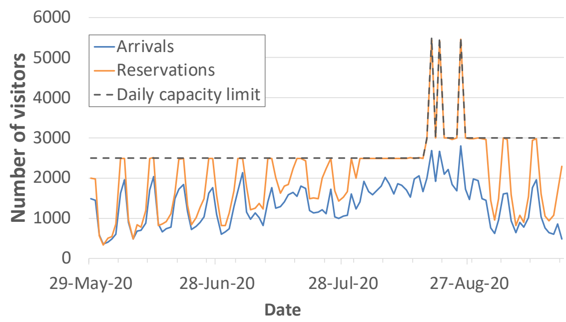
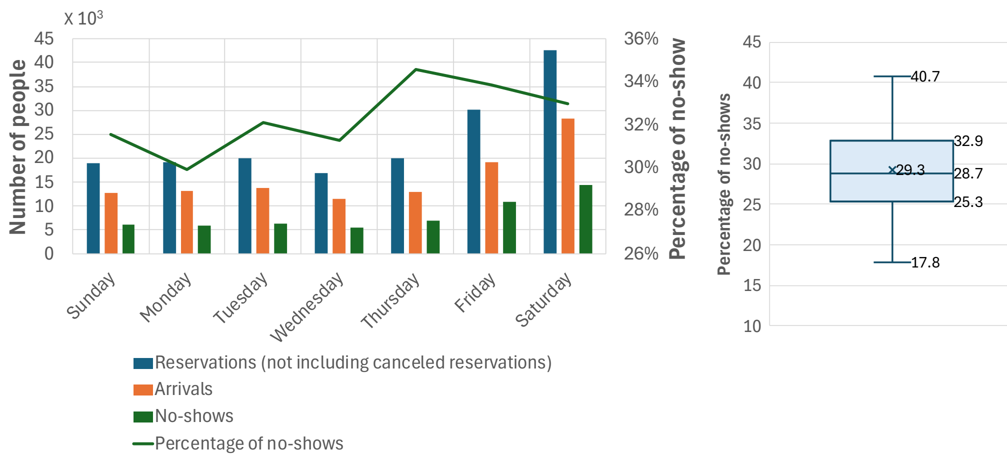


Figure 6 Number of active reservations vs. actual arrivals [Gan HaShlosha (Sahne) NP, May–September 2020, All days].



(a) As a function of day of the week (b) Average per park

Figure 7 Percentage of no-shows [May–December 2020, All NPs].

4. Optimizing NP Visitor Load

Our goal is to control the visitor load at NPs during visiting hours by optimizing the reservation system. We propose to separate the optimization problem into two parts and solve them sequentially. The first problem (Problem I) is to determine the optimal total number of visitors to the park on a focal day, given the park's individual features (arrival rate patterns, capacity, visit duration, etc.). This problem aims to balance overcrowding with accessibility. Based on fluid approximation we propose a numerical solution to this problem in Section 4.1.

The second problem (Problem II) is to design the reservation system to achieve the optimal number of visitors to the park on a focal day, as defined in Problem I. We analyze this problem in Section 5, where we minimize blocking costs as well as costs that result from a mismatch between the target number of visitors and the realized number of visitors. While the blocking costs accrue over time, the mismatch costs are incurred only on the focal day. The optimization of the reservation system is achieved by controlling the number of visitors who are allowed to make a reservation in the system. Because a reservation might be cancelled, rescheduled, or just not utilized (see Section 3), we need to allow overbooking of reservations (similar to airline and healthcare reservation systems), but not so much so as to cause overcrowding and a mismatch between the target number of visitors and the actual number of visitors. Because these dynamics change over time, we formulate a rolling-horizon model. The analysis in Section 5 will be on the fluid level, assuming that demand, no-show, and cancellation probabilities are known in expectation, and in Section 6 we confirm that the fluid policy preforms well also in its underlying stochastic environment. Previous research shows that fluid models have been successfully implemented in modeling service systems in various contexts, such as healthcare (Zychlinski et al. 2020) and contact centers (Yom-Tov et al. 2021). As noted by Zychlinski et al. (2020), "fluid frameworks are well adapted to large time-varying overloaded systems (Mandelbaum et al. 1998, 1999) [which is the case here]. Moreover, fluid models yield analytical insights, which typically cannot be obtained using their alternatives (e.g., simulation, time-varying stochastic queueing networks)."

4.1. Problem I: Determining the Target Total Arrival Volume

In this section, we want to determine the total number of visitors the park should allow to enter the park per day. The considerations the PAs need to take into account are wide and varied. These include (a) the park's features: how large is the physical space, how many interest points are there where visitors may spend time and what is the distance between these points, and whether the space is vulnerable, that is, does it include historical or natural areas that could be harmed by a large number of visitors; (b) the visitors' features: what is the popularity or demand for visiting the park, the visit duration distribution (usually this is affected by the park's features and its accessibility), and the arrival rate dynamics during the day; and (c) the costs: visit costs, maintenance costs of the park, and overcrowding cost, which represents visitors' experience and their wish to avoid crowds when venturing to nature. Both demand and visit duration are influenced also by external random features such as the weather.

Due to the above features, we assume that there is some *maximal number of visitors* (or "maximal capacity" for short), denoted by L, that the PA does not want to exceed *at any specific time of the day* and that exceeding this number will result in an over-cost of c^o per visitor per unit of time. On the other hand, it is assumed that the PA does not want to allow less than the maximal number of visitors and that falling below this number results in an under-cost of c^u per visitor per unit of time. The intra-day optimization problem should try to balance these two types of costs. Let $R(t)$ be the average number of visitors at the park at time t (also called "visitor load" for short). Then, we wish to minimize the total cost:

$$\int_0^T c^o(R(t)-L)^+ + c^u(L-R(t))^+ dt, \tag{1}$$

where T is the daily opening hours of the park.

We propose to model the number of visitors in the park using an $M_t/G/\infty$ queue, where arrival rate is according to a Poisson process with time-varying arrival rate λ_t, and general distribution of service times, denoted by S, with mean service rate μ. Because capacity in NP is a soft constraint, and visitors do not directly affect one another, we assume that the number of servers in infinite. Eick et al. (1993) finds that the visitor load, $R(t)$, of an $M_t/G/\infty$ system is $R(t) = E[\lambda(t - S_e)]E[S]$, where S is the service time random variable and S_e is the corresponding excess service time.

Denote by Λ the target number of visitors throughout the day (which we name "target arrivals" for short) at the park. Define $\lambda_t^{\%}$ as the percentage of the arrival rate at time t out of the daily target arrivals Λ, therefore, by definition $\int_0^T \lambda_t^{\%} = 1$. Then, the arrivals at time t are $\lambda_t = \Lambda\lambda_t^{\%}$. The optimization problem I can be written as

$$\begin{aligned} \min_{\Lambda} C(\Lambda) &= \int_0^T c^o(R(t) - L)^+ + c^u(L - R(t))^+ dt \\ s.t.\ R(t) &= \Lambda E[\lambda^{\%}(t - S_e)]E[S]. \end{aligned} \tag{2}$$

The above representation of $R(t)$ emphasizes that $R(t)$ is linear in Λ and independent of L. Denote $R^{\%}(t)$ as $R^{\%}(t) = E[\lambda^{\%}(t - S_e)]E[S]$. Define $R_i^{\%}(t)$ to be the increasing sorted version of $R^{\%}(t)$ (i stands for *increasing*). Let $\Phi(x) = \int_0^T R_i^{\%}(t)\,\mathbb{1}_{\left\{R_i^{\%}(t) \le x\right\}} dt$. Then, the optimal lambda is found by solving an integral equation as depicted by the following theorem:

Theorem 1. *There exists a solution to problem (2), and the optimal target total arrivals, Λ^*, that minimize it is given by the fixed point solution of the integral equation:*

$$\Phi\left(\frac{L}{\Lambda^*}\right) = \frac{c^o}{c^o + c^u} E[S]. \tag{3}$$

Providing $\Phi(x)$ is invertible on the relevant range, then, $\Lambda^* = \frac{L}{\Phi^{-1}\left(\frac{c^o}{c^o+c^u} E[S]\right)}$.

Remark 1. When $R_i^{\%}(t)$ is not strictly increasing, Φ may be constant on some interval of lambda values, so Φ^{-1} might not be single-valued. In that case, any x in the constant interval that satisfies $\Phi(x) = \frac{c^o}{c^o+c^u} E[S]$ can lead to an optimal Λ^*. This does not affect the existence of solutions or optimality but may yield a range of possible Λ^* values.

5. Developing a Real-time Reservation Management Policy (Problem II)

5.1. Model Definition

Next, we want to design an optimal reservation policy for which the actual number of visitors arriving at the park is as close as possible to the target number of visitors, Λ, with minimal costs. We assume that the demand for reservations is larger than Λ; therefore, the PA cannot satisfy all of the demand and must reject the reservation requests of some visitors.

The reservation process is a finite rolling-horizon process, meaning that the decision taken by the PA on day t regarding whether to accept a reservation depends on the decisions that were made up to day t as well as on the data on future demand, cancellations, no-shows, and costs. Let T be the *reservation horizon*—the number of days in which the reservation system is open for reservations to a focal day. Hence, this is also the maximal number of days between the day a reservation is made and the focal day. Let $t=1$ be the time when the reservation system opens for reservations, thus T is also the focal day.

We consider three types of costs: under- and over-costs (terms will be explained momentarily) and blocking costs. Under- and over-costs are paid according to the gap between the actual arrivals on the focal day T, denoted by I_T^e, and the target number of visitors on the focal day T, denoted by Λ. Let c^u be the cost incurred by the PA when the arrivals are less than Λ and c^o be the cost incurred by the PA when the arrivals are greater than Λ. Henceforth, we will refer to c^u as the *under-cost* and to c^o as the *over-cost*. The third type of cost is the blocking cost. Let c_t^b be the blocking cost at day t, which is the cost incurred by the PA for every customer it blocks from making a reservation during the reservation horizon T. Every demand for a ticket that is denied at day t increases the total blocking cost by c_t^b units. We will prove that the optimal reservation policy depends on the ratio between the blocking cost and the over-cost. Specifically, when we cannot avoid some costs, we need to decide which of these two types of costs we prefer to incur.

Denote by D_t the *demand* for customers entering at the reservation system on day t ($t \in \{1, ..., T\}$) wishing to reserve an entry permit to the NP and $\mathbf{D} = (D_1, ..., D_T)$ the vector of demands throughout the time horizon. Let Q_t be the number of customers whose request for reservation we accept at day t and $\mathbf{Q} = (Q_1, ..., Q_T)$ the vector of accepted requests throughout the reservation time horizon. Hence, $Q_t \leq D_t$. Let p_t be the probability of a ticket reserved on day t, for visiting of the focal day T, being cancelled before the focal day, that is, during the period $(t, T]$. Similarly, let φ_t be the no-show probability during the focal day T of a customer that made a reservation on day t. (Recall that a no-show customer is one who does not cancel their reservation but does not arrive at the park.) Assuming that cancellation and no-show behaviors are independent of each other, the mean number of customers that made a reservation on day t and arrive at the NP on the focal day T is $Q_t(1-p_t)(1-\varphi_t)$, and the total of that expression across all reservation days results in the mean number of arrivals at the NP on day T, given by

$$I_T^e = \sum_{t=1}^{T} Q_t(1-p_t)(1-\varphi_t). \tag{4}$$

The superscript e indicates that I_T^e is the number of *effective* reservations that are expected to actually arrive on the focal day out of the I_T reservations, given the cancellation and no-show probabilities.

Remark 2. NPs do not charge a fee for making a reservation; therefore, we assume that cancelling a reservation is costless. However, one could extend the proposed reservation system to account for reservation fees. In case there are reservation fees, they may apply differently for no-shows and cancellations. For example, in the hospitality industry, it is customary to charge a partial fee for cancellations that are made close to the focal day and charge a full fee for no-shows. These charges may also depend on the day the reservation was made.

Let I_t be the total number of reservations made up to day t: $I_t = \sum_{i=1}^{t} Q_i$. From those reservations, I_t^e visitors are expected to arrive. I_t^e can be also written recursively as $I_t^e = I_{t-1}^e + Q_t(1-p_t)(1-\varphi_t)$, which captures the rolling-horizon nature of the reservation process. Without loss of generality, we assume that $I_0 = I_0^e = 0$. These are the number of reserved permits when we open the reservation systems to the public.

We aim to find a policy for accepting reservations that minimizes the total costs function:

$$\sum_{t=1}^{T} c_t^b (D_t - Q_t)^+ + c^u \left(\Lambda - I_T^e\right)^+ + c^o \left(I_T^e - \Lambda\right)^+,$$

where Q_t is the decision variable (the number of accepted reservations at time t), D_t is the demand on each day t, and I_T^e is the number of arrivals at the focal day T, as given in Eq. (4). The function $(\cdot)^+$ stands for $\max\{0, \cdot\}$.

5.2. Solving the Fluid Optimization Problem

We solve the following fluid optimization problem:

$$\begin{aligned}
&\min_{\mathbf{Q}} \; \sum_{t=1}^{T} c_t^b (D_t - Q_t) + c^u (\Lambda - I_T^e)^+ + c^o (I_T^e - \Lambda)^+ && (5)\\
&s.t.\\
&\quad I_t^e = I_{t-1}^e + Q_t(1-p_t)(1-\varphi_t), && \forall t \in 1,...,T;\\
&\quad I_0^e = 0;\\
&\quad Q_t \in [0, D_t], && \forall t \in 1,...,T.
\end{aligned}$$

Intuitivly, the solution to problem (5) depends on the cost ratio $\frac{c^o(1-p_t)(1-\varphi_t)}{c_t^b}$. When this ratio is greater than 1, then the over-cost is higher than the blocking cost, and we should not allow more visitors than Λ. By contrast, when the ratio is less than 1, then the over-cost is lower than the blocking cost, and we should accept all the demand. Define an index function f such that $f_t = \frac{c_t^b}{(1-p_t)(1-\varphi_t)}$ for all t. Let $\hat{f}$ be the sorted vector of f from the highest to the lowest value. Denote by s the index of the sorted vector. Each index s maps to a real period t. We use $\hat{\cdot}$ to describe variables in the sorted system.

Theorem 2. *Let $\hat{f}$ be the sorted vector of f from the highest to the lowest value. Let index S_1 be maximal s such that $\hat{f}_s > c^o$ for all $s \le S_1$. Then, the optimal sorted solution to Eq. (EC.4) is $\hat{Q}^*_s = \hat{D}_s$ for all $s \le S_1$. If $\hat{I}^e_{S_1} = \sum_{s=1}^{S_1} \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) \ge \Lambda$, then $\hat{Q}^*_s = 0$ for all $S_1 < s \le T$. Otherwise, $\hat{\mathbf{Q}}^*_{(S_1+1,\ldots,T)} = \{\hat{D}_{S_1+1},\ldots,\hat{D}_{S_2-1},\frac{\Lambda-\hat{I}^e_{S_2-1}}{(1-\hat{p}_{S_2})(1-\hat{\varphi}_{S_2})},0,\ldots,0\}$, where S_2 is the index in which $\hat{I}^e_{S_1} + \sum_{s=S_1+1}^{S_2-1} \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) < \Lambda$ and $\hat{I}^e_{S_1} + \sum_{s=S_1+1}^{S_2} \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) \ge \Lambda$.*

The proof is in Appendix EC.2.

Theorem 2 identifies a two-threshold policy. The first threshold is a *time threshold* (at $s = S_1$)—before which all reservations are accepted regardless of the NP capacity—and the second is a *capacity threshold* (that we reach at $s = S_2$)—after which we stop accepting reservations in the case where $S_2 > S_1$ and therefore exists. The capacity threshold ensures the balance between over- and under-costs on the focal day. Theorem 2 can be implemented by the following algorithm:

Algorithm 1: The Two-Threshold Reservation Algorithm

1. Calculate all f_t values for every t, such that $f_t = \frac{c^b_t}{(1-p_t)(1-\varphi_t)}$.
2. Sort f_t from the highest to the lowest value. Let $\hat{f}$ (with indexes $s = 1,\ldots,T$) be the sorted function of f, and the *time-threshold* index $s = S_1$ be the maximal s in which $\hat{f} > c^o$, i.e., $S_1 = \max\{s|\hat{f}_s > c^o\}$.
3. For all days with index $s \in \{0,\ldots,S_1\}$, where $\hat{f}_s > c^o$, accept all demand for reservations.
4. For days with index $s \in \{S_1+1,\ldots,T\}$, where $\hat{f}_s \le c^o$, accept effective reservations sequentially (i.e., one at a time with increasing index-s days) as long as $\hat{I}^e_s$ doesn't exceed the *capacity threshold* Λ.

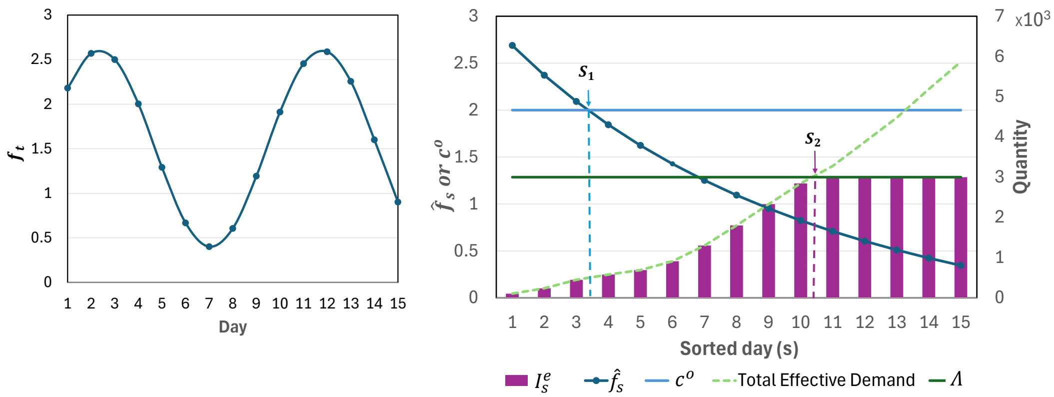
Algorithm 1 is illustrated in Figure 8. Figure 8(a) shows an arbitrary function f_t that may increase and decrease over time. The sorted function, $\hat{f}_s$ is illustrated in Figure 8(b) (marked solid blue line). This sorted function crosses $c^o = 2$ (solid light blue line) at time S_1. Therefore, all the effective demand for reservations (dashed light green line) in indexes 1–3 are accepted (purple columns). The total effective demand at index S_1 is below $\Lambda = 3000$ (solid dark green line), so all demand is accepted up to index S_2, where the total number of effective accepted reservations equals Λ. At this point, no further demand is accepted, that is, $I^e_T = \Lambda$.

The optimal solution according to this two-threshold policy is such that

$$I^e_T = \max\left\{\sum_{s:\hat{f}_s > c^o} D_s(1-p_s)(1-\varphi_s), \Lambda\right\}. \tag{6}$$

The first component in (6) occurs when (more than) Λ effective reservations are accepted before the time-threshold index S_1, that is, by accepting all the demand when $\hat{f}_s > c^o$. In this scenario, the sorted acceptance vector will be

$$\hat{Q}^* = (D_1,\ldots,D_{S_1},0,\ldots,0). \tag{7}$$



(a) f_t (b) Sorted solution

Figure 8 **Illustration of an optimal two-threshold policy.**

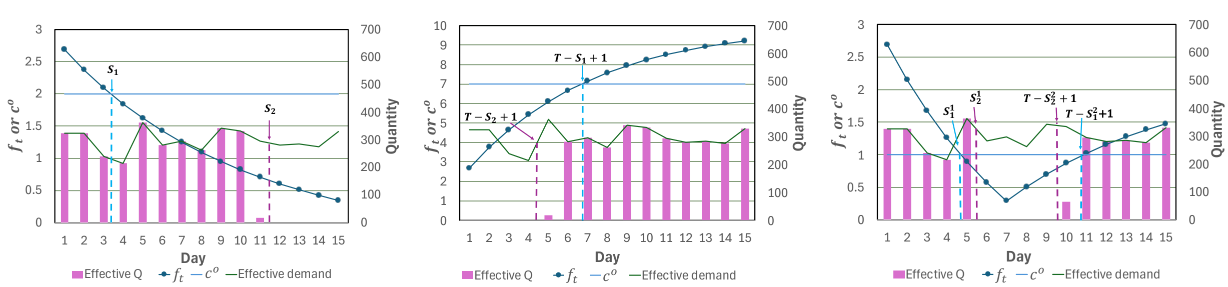
The second component in (6) occurs when the total effective demand up to the time-threshold index S_1 (i.e., in all the periods in which $\hat{f}_s > c^o$) is less than Λ. In this scenario, the sorted acceptance vector will be

$$\hat{Q}^* = \left(D_1, ..., D_{S_1}, D_{S_1+1}, ..., \frac{\Lambda - \hat{I}^e_{S_2-1}}{(1-\hat{p}_{S_2})(1-\hat{\varphi}_{S_2})}, 0, ..., 0 \right) \tag{8}$$

reservations. This policy for the sorted reservation function $\hat{Q}^*$ needs to be translated back to the actual day t to form Q^*.

5.3. Managerial Implications

Next, we explore how time-varying dynamics of the blocking cost and the cancellation and no-show probabilities impact (a) the time horizon in which the PA should open the reservation system for bookings and (b) whether the PA needs to actively limit the number of entry permits available for reservation in the reservation system between periods. We will discuss three scenarios as depicted in Figure 9: (a) decreasing f_t, (b) increasing f_t, and (c) decreasing-increasing f_t.



(a) Decreasing f_t (b) Increasing f_t (c) Decreasing-increasing f_t

Figure 9 **Reservation system optimal solution for different f functions' time-varying dynamics.**

5.3.1. Decreasing f_t As defined, the f function's time-varying dynamics may be attributed to one of the following parameter dynamics: the blocking cost and the cancellation and no-show probabilities, since $f_t = c_t^b/((1-p_t)(1-\varphi_t)$. In Section 3, we showed that the probability of a reservation being cancelled decreases as the time between the reservation day and the focal day decreases (see Figure 4). In other words, reservations that are made closer to the focal day are less likely to be cancelled. This means that, in practice, the cancellation probability, p_t, decreases over time. While we do not have information on the dynamics of the no-show probability, research on healthcare appointment systems shows decreasing dynamics for no-shows too (Feldman et al. 2014). Understanding how the blocking cost, c_t^b, behaves as a function of time is less clear. A reasonable assumption is that c_t^b is constant over time, taking the price of an entry permit lost due to the reservation blocking.

PROPOSITION 1. *Assume that the no-show probability, φ_t, and the cancellation probability, p_t, decrease over time and that the blocking cost, c_b, is constant over time. Then, the optimal time to open the reservation system is as early as possible. The PA should accept reservations continuously from that time onwards until Λ effective reservations have been accepted.*

This scenario also has a very easy implementation to a real-time algorithm as described graphically in Figure 9(a). As can be seen in the figure, when f is decreasing, the PA should open the reservation system as early as possible and fill up all the time slots continuously up to day S_2. If we add the reasonable assumption that the blocking cost is not too high, meaning that $f_1 \le c^o$, the PA should accept all reservations until the capacity threshold is reached. Therefore, the number of available entry permits for reservations on Day 1 should be set to $\Lambda/((1-p_1)(1-\varphi_1))$, and this cap should be updated daily according to accepted demand, realized cancellations, and the probability to cancel reservations (or not show up to the focal day) of active reservations. Hence, on day t the number of available entry permits for reservations will be $(\Lambda - I_t^e)/((1-p_t)(1-\varphi_t))$. (A more elaborate version will be described in Section 6, where we use a hazard rate function of the cancellation probability to get a prediction of the stochastic equivalent of I_t^e, denoted as $\mathbb{I}_t^e$, using real-time information on cancellations. See Algorithm 2.)

Some PAs implement policies in this spirit as if their f_t function is decreasing. In New Zealand, for example, park registration starts six months before parks open for the season, and within 10–20 minutes all tickets for the popular trails are taken for that season. Later reservations can be made only if someone cancels their reservation.

Proposition 1 also holds for blocking costs that decrease over time. Next, we explore the opposite scenario.

5.3.2. Increasing f_t Here, we assume that f_t increases over time. This can happen if c_t^b is increasing at a large enough rate such that f_t is increasing in spite of the decreasing no-show and cancellation probabilities.

PROPOSITION 2. *Assume that f_t increases over time. Then, the optimal time to open the reservation system is as late as possible at day* $\min\{T-S_1+1, T-S_2+1\}$*. The PA will accept reservations continuously from that day along the whole time horizon. In the case that index $s=S_2$ exists, at day $T-S_2+1$, partial demand may be accepted such that*

$$Q_{T-S_2+1} = \min\left\{\frac{\Lambda - \sum_{t=T-S_2+2}^{T} D_t(1-p_t)(1-\varphi_t)}{(1-p_{T-S_2+1})(1-\varphi_{T-S_2+1})}, D_{T-S_2+1}\right\}.$$

This case is illustrated in Figure 9(b), where the reservation system is open from Day 5 onwards. All demand before that day is blocked, and all demand after that day is accepted, while demand on Day 5 is partially accepted.

This scenario also has a very easy implementation to a real-time algorithm. A more elaborate version will be described in Appendix EC.3.

5.3.3. Decreasing-increasing f_t As noted above, the application of an early start or late start for the reservation period depends on what we assume regarding visitors' behavior. Do we assume that people plan their visit to the park well in advance, coming especially for this park and traveling long distances to reach it, or do we assume that people are more spontaneous, last-minute planners. Our data analysis in Section 3 suggests that both types of visitors exist. Hence, it may also be realistic to assume that c_t^b is convex, first decreasing and then increasing (see Figure 9(c)). In such a case, the PA should divide the number of entry permits it proposes on the reservation system between the two visitor populations, offering some entry permits to be reserved on the first few days of the horizon, then blocking all demand for the next period of days, and then offering some entry permits for reservations on the last few days of the horizon. Specifically, the index periods 1 to S_1 are divided between the beginning of the horizon (days 1 to S_1^1) and the end of the horizons (days $T-S_1^2+1$ to T). The index periods S_1+1 to S_2 are also divided between the beginning of the horizon (days S_1^1+1 to S_2^1) and the end of the horizon (days $T-S_2^2+1$ to $T-S_1^2$). We accept reservations up to day $\max\{S_1^1, S_2^1\}$ and from day $\min\{T-S_1^2+1, T-S_2^2+1\}$.

Figure 9(c) illustrates such a policy: reservations are accepted to and through Day 5 (S_2^1), from Days 6 to 9 ($T-S_2^2$) customers are informed that no reservations are available, from Day 10 ($T-S_2^2+1$) reservations are accepted again. On one of the days S_2^1 or $T-S_2^2+1$, reservations will be partially accepted up to a capacity threshold (the threshold can be calculated using the same logic as Q_{T-S_2+1} in proposition 2). On all other days, all reservations are accepted.

Remark 3. The policy suggested above enables a potential visitor to the park to observe openings in the reservation system for either short-term or long-term focal days. The PA saves some reservations for last-minute planners but also takes into account openings due to expected cancellations and no-shows (where same-day slots become available randomly). Saving reservations for last-minute customers is a common policy both in reservation systems, such as in hotels (Bitran and Gilbert 1994) or healthcare (Schacht 2018), and in non-reservation systems, as in hospitals (Kim et al. 2020).

Remark 4. Note that if the no-show and cancellation probabilities and the blocking costs are constant over time when $f < c^o$, then our two-threshold policy behaves like an overbooking policy with a capacity threshold of $\Lambda(1-p)(1-\varphi)$. We will use that as one of the benchmark policies tested in the next section.

6. Implementing the Fluid Policy in Practice

In Section 6.1, we propose an algorithm for the implementation of the fluid policy in real time, which we call the *adaptive two-threshold* (ATT) policy. In Section 6.2, we suggest to compare this policy to four benchmark policies. In Section 6.3, we conduct a simulation study to show that the ATT policy provides better performance than does the benchmark policies under stochastic behavior and real-time decision-making. We calibrate the simulation study using real data that was collected from the INPA for the period May–December 2020 (see Section 3). We present our simulation study results in Section 6.4, comparing performance under various load scenarios.

6.1. The Adaptive Two-Threshold (ATT) Policy

In this section, we propose ways to translate Algorithm 1 into a real-time decision-making policy using real-time information. In a real-time scenario, information on cancellations is obtained over time. By contrast, information on no-shows is obtained only on the focal day. Hence, our consideration of these two behaviors differs.

The real-time reservation process: In this section, we regard time as a continuous variable $t \in [0,T]$, instead of in days, where T is the time horizon in which the reservation system is open for reservations. It is assumed that at time t either a reservation or a cancellation arrives to the system. If, at time t, demand for a permit arrives to the system, a decision is made regarding acceptance. The number of customers who make a reservation at time t is 1 if the reservation is accepted ($Q_t = 1$), and 0 otherwise (in which case a blocking cost c_t^b is incurred). Alternatively, at time t a cancellation may arrive to the system. Let C_t denote a reserved ticket cancelled at time t. The total number of *active* reservations (i.e., reservations that were not cancelled so far) at time t, $\bar{I}_t$, is given by $\bar{I}_t = I_0 + \int_0^t (Q_i - C_i)di$ for $t \in [0, T^-]$. Since we know the number of cancellations, we also know the number of reservations made at time i that were not cancelled before time t, namely, $\bar{Q}_i^t$ (by definition, $\bar{Q}_t^t = Q_t$). Therefore, an alternative expression for the active number of reservations at time t, $\bar{I}_t$, is $I_0 + \int_{i=0}^{t} \bar{Q}_i^t di$, for all $t \in [0, T^-]$. Denote by N_T the number of customers who reserved permits but do not show up at the NP on time T. Then, $\bar{I}_T = \bar{I}_{T^-} - N_T = \int_{i=0}^{T^-} \bar{Q}_i^T di - N_T$. [$I_0 = 0$.] As before, we compare $\bar{I}_T$ to Λ and pay an over- or under-cost if they do not match.

Predicting cancellations: At time t, the ATT algorithm relies on real-time information regarding cancellations done until that time, and therefore needs to predict the probability of an active reservation to be cancelled from time t to T. We estimate this probability using the hazard rate function (see example from our data in Figure 10, in daily resolution). Let $r_i(t)$ be the hazard rate function for the cancellation probability on time t of a reservation made on time i, that is, $t \in [i, T]$. Thus, $r_i(t)$ is the risk that a customer will cancel her reservation *on* time t *given that her reservation was not cancelled before that time* (for a reservation made on time i). Therefore, the probability that a customer will cancel her reservation *on* time t (for a reservation made on time i) is $p_i(t) = r_i(t)S_i(t) = r_i(t)e^{-\int_{u=i}^{t} r_i(u)du}$. The sum of all $p_i(t)$ for all t gives us the probability of a reservation being cancelled at any time between the reservation day i and time T, that is, $p_i = \int_{t=i}^{T} p_i(t)dt$ (which is the parameter we used in the fluid analysis of Section 5). Finally, let $\mathbb{H}_i^t$ denote the probability of a reservation made on time i, and still active on time t, being cancelled before the end of the horizon T. That is,

$$\mathbb{H}_i^t = \frac{\int_{j=t}^{T} p_i(j)dj}{S_i(t)} = \int_{j=t}^{T} r_i(j)\frac{S_i(j)}{S_i(t)}dj = \int_{j=t}^{T} r_i(j)e^{-\int_{u=t}^{j} r_i(u)du}dj. \tag{9}$$

Eq. (9) is used to derive the number of active reservations at time t that are predicted to arrive to the NP on time T, namely, $\bar{I}_t^e$ by $\bar{I}_t^e = \int_{j=0}^{t} \bar{Q}_j^t \left(1 - \mathbb{H}_j^t\right)(1 - \varphi_j)dj$.

The ATT algorithm: Assuming that f is a decreasing function, Algorithm 2 for optimizing the reservation system is defined as follows:[2]

Algorithm 2: The Adaptive Two-Threshold (ATT) Algorithm for Decreasing f Function

1. Calculate all f_t values for every time t, such that $f_t = \frac{c_t^b}{(1-p_t)(1-\varphi_t)}$.
2. Identify $S_1 = \max\{t | f_t > c^o\}$.
3. Set $t = 0$.
4. If $t \leq S_1$, accept all the demand for reservations during that time, i.e., $Q_t = D_t$.
5. If $t > S_1$, calculate the number of reservations that are active just before time t and are predicted to arrive the NP by $\bar{I}_{t^-}^e = \int_{i<t} \bar{Q}_i^t (1 - \mathbb{H}_i^t)(1 - \varphi_i)di$. Then, if $\bar{I}_{t^-}^e \geq \Lambda$, accept no reservations (i.e., $Q_t = 0$). Otherwise, accept $Q_t = \min\{D_t, (\Lambda - \bar{I}_{t^-}^e)/((1 - p_t)(1 - \varphi_t))\}$ reservations.
6. Set $t = t + dt$. If $t \leq T$, go to Step 4; otherwise, stop.

[2] In Appendix EC.3 we provide Algorithm 3 for the less realistic scenario where f is an increasing function.

Here we assumed that no-show and cancellation probabilities decrease over time (see Section 6.3 for more details) while the blocking probability remains constant. Hence, by Proposition 1 the PA should open the reservation system as early as possible and accept all reservations up to the time threshold, S_1, and then continue to accept additional reservations up to Λ effective reservations. The decision at time $t \in (S_1, T]$, depends on the estimation that an active reservation at time t will be realized. (Note that the algorithm allows acceptance of more than one entry permit in each reservation, relaxing the assumption we made at the beginning of this section regarding Q_t, for predicting cancellations, since all reservations made at the same time are statistically identical.)

6.2. Benchmark Policies

We will compare the above-defined adaptive two-threshold (ATT) policy to four commonly used policies:

1. The lambda-level (LL) policy: Accept effective reservations until the capacity threshold is reached, where each reservation's effectiveness is predicted based on its no-show and cancellation probabilities. Hence, this policy ignores the time threshold S_1; that is, f_t is assumed to be smaller than c^o for all t. This policy should work well in under-loaded or medium-loaded systems. It is similar to the overbooking strategy suggested by Lawrence et al. (2003).
2. The no-show overbooking (NOB) policy: Let $\bar{\varphi}$ be the average no-show probability. Open $\Lambda/(1-\bar{\varphi})$ permits to be reserved in the system, and accept reservations until those permits are booked. This policy ignores the time-varying dynamics of cancellation and no-show probabilities. It sets an average no-show probability for all reservations and ignores cancellation probabilities, due to the fact that cancellations are realized before the focal day and can be replaced by demand arriving after the cancellation time. This policy should work in our NP setting since most of the demand arrives in the last few days. One problem of this policy is lost sales due to cancellations on the last day.
3. The cancellation and no-show average overbooking (OB) policy: Let $\bar{p}$ be the average cancellation probability and $\bar{\varphi}$ be the average no-show probability. Open $\Lambda/((1-\bar{p})(1-\bar{\varphi}))$ permits to be reserved in the system, and accept reservations until those permits are booked. This policy ignores the time-varying dynamics of cancellation and no-show probabilities and sets average no-show and cancellation probabilities instead. One problem of this policy may be accepting too many reservations.
4. No-reservation (NR) policy: In reality, in a no-reservation system, there would be no blocking and also no cancellations or no-shows of reservations. For a fair comparison, we regard the "demand for reservations" as people's intention to visit the NP, and the no-shows and cancellations as an event where people changed their plans to visit the NP, therefore not realizing those intentions. Hence, this policy is similar to having a reservation system with no limit on the number of permits opened in the system and accepting all the demand for reservations.

6.3. The Experiment Design of our Simulation Study

In this section, we describe a discrete-time simulation study designed to compare the performance of the above policies in a realistic setting. The simulation will examine costs of a single focal day in a typical NP. We rely on data analyzed in Section 3 to estimate model parameters. Specifically, the simulation inputs on the cancellation, no-show, and demand dynamics during the reservation time horizon are based on average customer behavior across all NPs and the total demand values are based on the average number of reservations made for a single NP.

Note that the demand in the simulation is stochastic. To distinguish between expected demand, D_t, and realized demand, we will denote the latter demand $\mathbb{D}_t$, and the resulted number of active tickets at day T after random no-show and cancelations as $\bar{\mathbb{I}}_T$. Denote V_π as the total cost under policy π, where $\pi \in \{ATT, LL, NOB, OB, NR\}$ for the above-defined policies. $V_\pi = \sum_{t=1}^{T} c_t^b(\mathbb{D}_t - Q_t) + c^u(\Lambda - \bar{\mathbb{I}}_T)^+ + c^o(\bar{\mathbb{I}}_T - \Lambda)^+$, where Q_t is determined by policy π. We simulate the decision made by each policy π and compare the cost V_π.

6.3.1. Cancellation and No-show Probabilities We use the cancellation probabilities presented in Section 3. Specifically, Figure 4 showed the probability of a reservation made on day t being cancelled before the focal day (see also the gray line in Figure 11(a)). For the simulation, we need an estimation of the respective hazard rate function, $r_i(t)$, presented in Figure 10. We observe an interesting behavior, where on the reservation day there is a 5% risk of same-day cancellation, which drops to 1%–2% until the last two days before the focal day. The cancellation risk then increases again to 5%–10% on Day -2, to 10%–15% on Day -1, and to 9%–10% on the focal day. This pattern is consistent regardless of when the reservation was made, except for reservations made on Day -1, which show different (and higher) risk for same-day cancellations.

Using this data, we calculate $\mathbb{H}_i^t$: the accumulated probability that a reservation made on day i and not cancelled before day t will be cancelled before the focal day T. This calculation is used for (a) randomizing cancellations made on day i and (b) making acceptance/blocking decisions

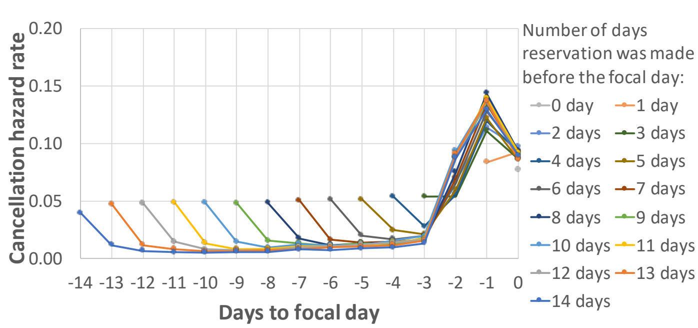
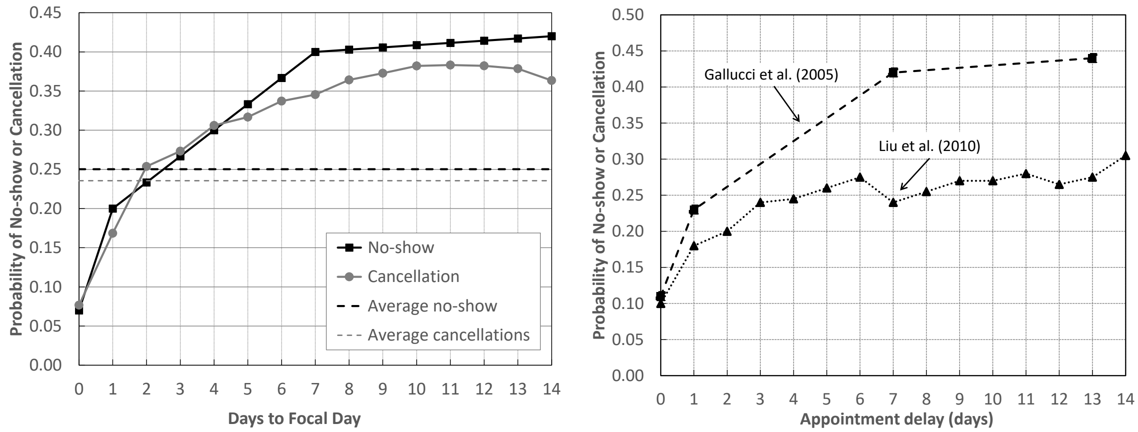


Figure 10 Hazard rate function for cancellation probabilities, by reservation day [May–December 2020, All NPs].



(a) NPs (b) Healthcare (Liu et al. 2010)

Figure 11 **No-show and cancellation probabilities as a function of the number of days until the focal day**

for the ATT and LL policies presented in Sections 6.1 and 6.2. We apply the average cancellation probabilities over time to the OB and NOB policies (see Section 6.2).

As discussed in Section 3, due to data limitations we cannot measure individual-level no-shows (i.e., we cannot connect a specific no-show to a specific reservation nor to the day it was made). Instead, we only know the total no-show percentage, by comparing the total active reservations made for a focal day to the total number of arrivals on that focal day. Previous studies on healthcare appointment no-shows have shown that the no-show probabilities decrease over time (e.g., Gallucci et al. 2005, Liu et al. 2010). Therefore, in our simulation, we use the no-show time-varying dynamics from Gallucci et al. (2005) (see the upper line in Figure 11(b)) and proportionally decrease the values by a few percentage points so that the total no-show probability matches the NP data (see the no-show line in Figure 11(a)). Similar to the cancellation probabilities, the no-show probabilities are used for (a) randomizing no-shows on day T and (b) making acceptance/blocking decisions for the ATT and LL policies. We apply the average no-show probabilities over time to the OB and NOB policies (see Section 6.2).

Using these no-show and cancellation probabilities, with constant blocking costs, yields a decreasing f function. Therefore, we use Algorithm 2 to simulate the performance of the ATT policy.

6.3.2. Cost and Demand Inputs We implement a 3×11 experimental design, including three combinations of over-costs and eleven combinations of total demand. The three combinations of over-costs are designed such that $S_1 = 0$, $S_1 \in \{1, T-1\}$, or $S_1 = T$. $S_1 = 0$ when the maximum of f_t is less than c^o, $S_1 = T$ when the minimum of f_t is greater than c^o, and $S_1 \in \{1, T-1\}$ when c^o is between the maximum and minimum f_t values. In practice, we change only the value of c^o to accomplish these three combinations (see Table 2). The mean of the total demand values throughout the time horizon ($T = 15$ days) ranges between 2000 and 7000 in jumps of 500 entry permits (eleven combinations). The demand of each simulation replication is drawn from a normal distribution with the above means and an SD that is 1% of the mean. These mean values of the total demand are designed so that the mean of the total number of reservations after cancellations varies between 1529 and 5351 but the mean of the total number of arrivals after no-shows drops to between 1173 and 4105. Compared to the target load $\Lambda = 1095$, these demand values create moderate- to high-load scenarios. (Note that this number of reservations represents a realistic demand for one NP, as seen in Figure 6.) We ran 100 replications for each over-cost and total demand combination. The total demand is distributed over time according to the demand function presented in Figure 1.

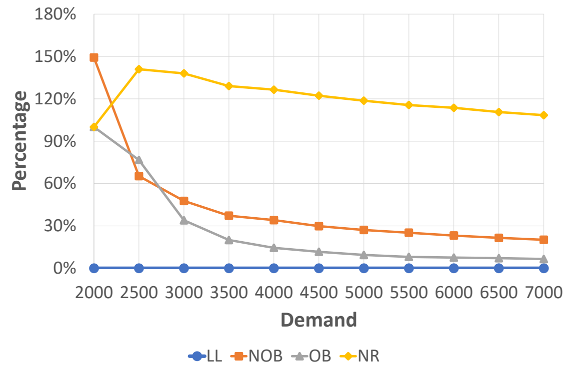
Combination No.	T	c^b	c^u	c^o	Max f_t	Min f_t	S_1
1	15	4	3	12	11.05	4.66	0
2	15	4	3	7	11.05	4.66	13
3	15	4	3	4	11.05	4.66	$T = 15$

Table 2 Parameter combinations for simulation study.

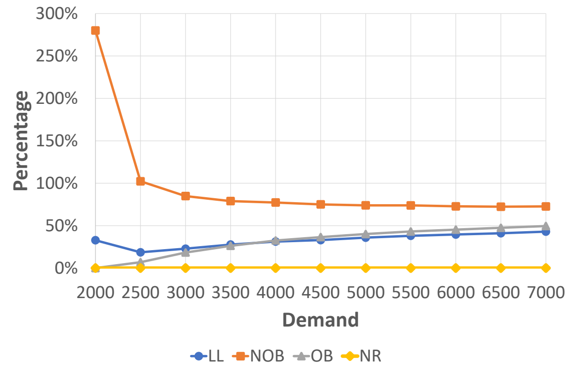
6.4. Simulation Study Results

Figures 12–13 present the results of our simulation study (additional results are presented in Appendix EC.5). Figure 12 presents the cost gap, in percentage, between the cost of a specific policy π, V_π, and the average cost of our proposed policy, $\mathbb{E}[V_{ATT}]$, as a function of demand, for each of the parameter combinations presented in Table 2. Figure 13 shows the gap, in percentage, between arrivals to the park on the focal day and the target number of visitors (Λ) as a function of demand. Figures EC.7–EC.9 present the same information using boxplots of the 100 replications, where we present three specific total demand values for 2000, 3000, and 7000 reservations, each representing a different situation of the relationship between the policies' performance.

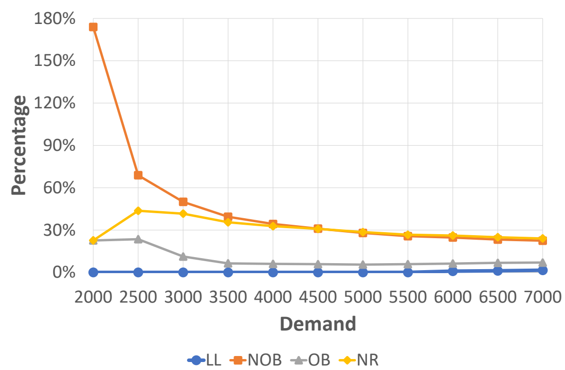
Examining Figure 12, we first observe that ATT is the only policy that provides a minimum cost in all the combination cases (because the cost gap of the other policies is always greater than 0%). The second-best policy is the LL policy. Recall that the LL policy takes into account the time-varying dynamics of both cancellation and no-show probabilities. In analyzing the performance of the LL policy, we observe an interaction between demand and S_1 (which is determined by the relationship between f_t and c^o). As S_1 increases as it goes to T (i.e., moving from Combinations 1 to 3), the differences between the ATT and LL policies become more apparent. This is because the ATT policy accepts all reservations before S_1, while LL rejects some reservations; therefore, when $S_1 > 0$, the two policies diverge. Yet, when $0 < S_1 < T$ (Combination 2), we start seeing



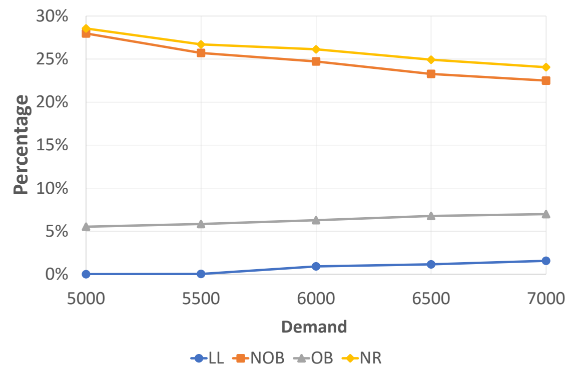
(a) Combination 1, Demand: 2000–7000



(b) Combination 3, Demand: 2000–7000



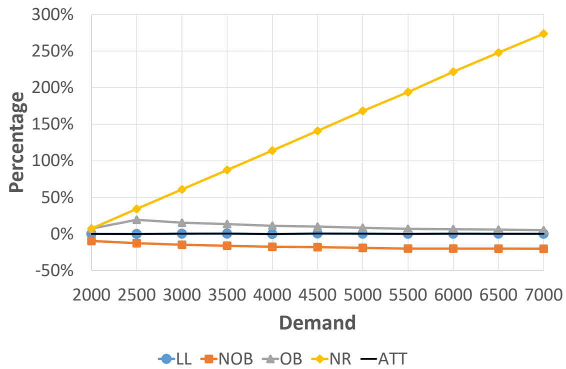
(c) Combination 2, Demand: 2000–7000



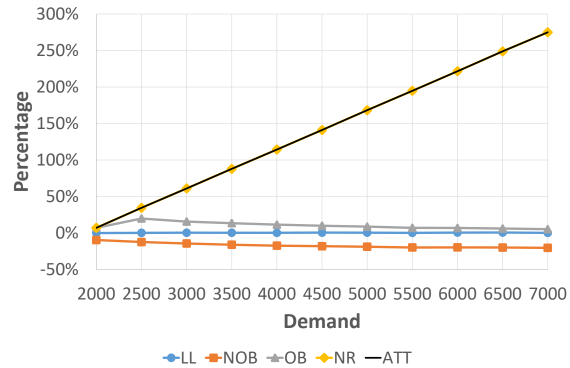
(d) Combination 2, Demand: 5000–7000

Figure 12 **Comparison of the average total cost of each policy, $\mathbb{E}[V_\pi]$, to the average total cost of the ATT policy, $E[V_{ATT}]$, as a function of total demand.**

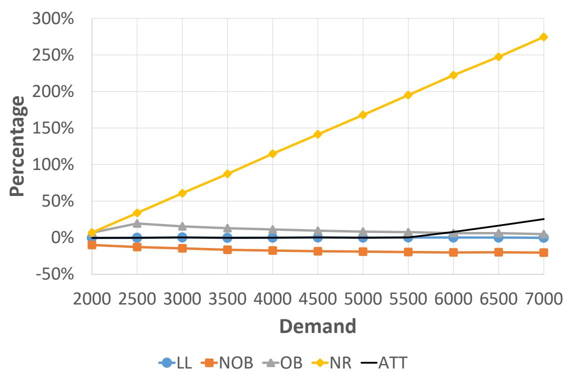
differences only when the total demand is high (see Figure 12(d)). This is because when the demand is moderate, the probability that more than Λ effective reservations are accepted before S_1 is low, especially considering the pattern of lower demand and higher cancellation probabilities at the beginning of the time horizon. Therefore, we see little to no difference between these policies when demand is lower than 5500 reservations. As S_1 increases, these differences manifest at lower demand. Specifically, when $S_1 = T$, the gap reaches 30% when the demand is 2000 (see Figure 12(b)). In contrast to the LL policy, the NR policy performs better as S_1 and total demand increase. The NR policy accepts all reservations. Therefore, it shows the same performance as the ATT policy in Combination 3 regardless of the demand. However, large differences appear in Combinations 1 and 2, where it is optimal to apply some sort of limitation on the amount of accepted reservations by taking into account the time-varying dynamics of both cancellation and no-show probabilities. Hence, this policy makes sense only when blocking costs are greater than or equal to the over-costs (leading to Combination 3 scenarios).



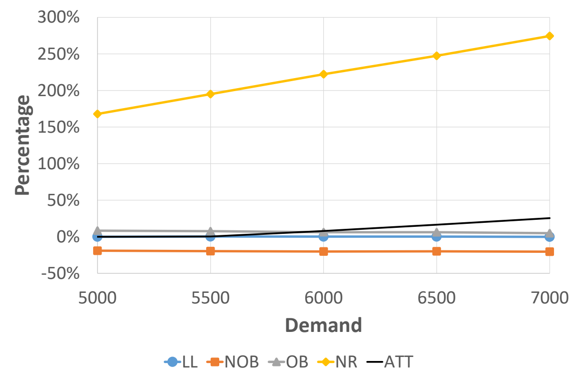
(a) Combination 1, Demand: 2000–7000



(b) Combination 3, Demand: 2000–7000



(c) Combination 2, Demand: 2000–7000



(d) Combination 2, Demand: 5000–7000

Figure 13 **Comparison of average arrivals at the focal day between each policy to Λ as a function of total demand.**

The OB policy performance is somewhere in between that of the LL and NR policies. The OB policy only performs well in Combination 3 when the load is very low (2000 reservations). In the simulation study, the OB policy allowed for more reservations than LL does in most scenarios (see Figure 13). This is because the OB policy uses average no-show and cancellation probabilities that are decreasing in practice and demand that is increasing over time. Therefore, it overbooks reservations later on in the time horizon by overestimating their cancellation and no-show probabilities, and exceeds the target number of visitors.

The NOB policy has lower performance across all simulated scenarios. One can interpret this policy as assuming that the total demand is high enough so that any cancelled reservation for a permit will be replaced by a new reservation before the end of the time horizon. Therefore, this policy takes in account only no-shows. The problem is that most cancellations are done in the last three days before the focal day, and by that time it is hard to fill in the gap between the number of active reservations and the target number of visitors. Therefore, this policy always results in a shortage of visitors (see Figure 13), which in turn results in under-costs.

Remark 5. Note that our experiment design assumed a decreasing f_t function. The acceptance order of periods in the ATT and the LL policy is identical; therefore, if the load is not too high the LL policy performs well. But if f_t were in any other shape, differences might also be observed in light load.

7. Conclusion and Future Research

This paper addressed the optimization of NP workload. Taking a hierarchical approach, we first determine the optimal number of visitors in the park, and then developed an optimization model that reach that target by managing daily reservation quotas. The first step analysis is build on fluid analysis of the time-varying load in the park. We based our second stage model on analyzing reservation data from INPA. Key findings reveal high rates of cancellations and no-shows, with probabilities varying over time. Additionally, demand fluctuates, with some visitors reserving well in advance and others making last-minute bookings. Incorporating these dynamics, we developed an optimization model that minimizes total costs by managing daily reservation quotas. The model assumes real-time blocking penalties, while over- and under-cost penalties are incurred on the focal day in which visitor enter the park. The framework operates on a rolling finite-time horizon and accounts for behavioral patterns in cancellations and demand variability.

Our approach is adaptable to other high-capacity reservation systems, such as amusement parks, museums, and healthcare systems. Future research could explore stochastic modeling for real-time adaptability, as well as extending the model to account for dependencies between reservation size and no-show probabilities. Adjusting probabilities for group reservations or individual characteristics, such as historical behavior, could further personalize and enhance system efficiency.

Dynamic factors like weather should also be integrated into NP capacity estimations, alongside static features like trails and staffing. Furthermore, strategies to influence demand, such as providing load information, could optimize visitor distribution. Extending the model to a network of parks would allow analysis of how blocking at one site impacts demand for others, with parallels to healthcare systems where resource management across interconnected facilities is critical.

Finally, our model can incorporate dependencies between days and sites by refining the blocking cost to reflect visitor flexibility. For example, multi-park visits may entail higher blocking costs if denied access to one park disrupts entire itineraries. Data indicates 6.8% of reservations include multiple sites, underscoring the need for further analysis of such bundling behaviors.

References

Ahmadi RH (1997) Managing capacity and flow at theme parks. *Operations Research* 45(1):1–13.

Alexandrov A, Lariviere MA (2020) Are reservations recommended? *Manufacturing & Service Operations Management* 14(2):218–230.

Bitran GR, Gilbert SM (1994) Managing hotel reservations with uncertain arrivals. *Operations Research* 44(1):35–49.

Cayirli T, Veral E (2003) Outpatient scheduling in health care: A review of literature. *Production and Operations Management* 12(4):519–549.

Cole DN (1990) Ecological impacts of wilderness recreation and their management. *Wilderness Management*, 425–466 (North American Press).

Cole DN (2012) Wilderness visitor experiences: A selective review of 50 years of research. *Park Science* 28(3):441–446.

Cole DN (2019) The relationship between amount of visitor use and environmental impacts. *Prepared for the Interagency Visitor Use Management Council* 1–12.

Cole DN, Hall TE (2010) Privacy functions and wilderness recreation: Use density and length of stay effects on experience. *Ecopsychology* 2(2):67–20.

D'Antonio A, Monz C (2016) The influence of visitor use levels on visitor spatial behavior in off-trail areas of dispersed recreation use. *Journal of Environmental Management* 170:79–87.

Dole A (2023) An analysis of hotel booking cancellations and factors affecting revenue generation. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* 11(X):2141–2148.

Eick SG, Massey WA, Whitt W (1993) $M_t/G/\infty$ queues with sinusoidal arrival rates. *Management Science* 39(2):135–253.

Feldman J, Liu N, Topaloglu H, Ziya S (2014) Appointment scheduling under patient preference and no-show behavior. *Operations Research* 62(4):794–811.

Gallucci G, Swartz W, Hackerman F (2005) Brief reports: Impact of the wait for an initial appointment on the rate of kept appointments at a mental health center. *Psychiatric Services* 56(3):344–346.

Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management* 5(2):79–141.

Grant B, Gurvich I, Mutharasan RK, Mieghem JAV (2022) Optimal dynamic appointment scheduling of base and surge capacity. *Manufacturing & Service Operations Management* 24(1):59–76.

Green LV, Savin S (2007) Reducing delays for medical appointments: A queueing approach. *Operations Research* 56(6):1526–1538.

Gupta D, Denton B (2008) Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* 40(9):800–819.

Hammitt WE, Cole DN, Monz CA (2015) *Wildland recreation: Ecology and management* (John Wiley & Sons).

Hassin R, Mendel S (2008) Scheduling arrivals to queues: A single-server model with no-shows. *Management Science* 54(3):565–572.

Huang J, Mandelbaum A, Momčilović P (2022) Appointment-driven service systems with many servers. *Queueing Systems: Theory and Applications* 100(3):529–531.

Karush W (1939) *Minima of functions of several variables with inequalities as side constraints* (University of Chicago).

Kim SH, Tong J, Peden C (2020) Admission control biases in hospital unit capacity management: How occupancy information hurdles and decision noise impact utilization. *Management Science* 66(11):5151–5170.

Kuhn HW, Tucker AW (1951) Nonlinear programming. Neyman J, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, 481–492 (Berkeley, Calif.: University of California Press).

Law R, Li G, Fong DKC, Hand X (2019) Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research* 75:410–423.

Lawrence RD, Hong S, Cherrier J (2003) Passenger-based predictive modeling of airline no-show rates. *KDD '03: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 397–406 (Association for Computing Machinery).

Leeftink G, Martinez G, Hans EW, Sir MY, Pasupathy KS (2022) Optimising the booking horizon in healthcare clinics considering no-shows and cancellations. *International Journal of Production Research* 60(10):3201–3218.

Liu N, Finkelstein SR, Kruk ME, Rosenthal D (2018) When waiting to see a doctor is less irritating: Understanding patient preferences and choice behavior in appointment scheduling. *Management Science* 64(5):1975–1996.

Liu N, Ziya S, Kulkarni VG (2010) Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing & Service Operations Management* 12(2):189–370.

Mandelbaum A, Massey WA, Reiman M (1998) Strong approximations for markovian service networks. *Queueing Systems* 30(1–2):149–201.

Mandelbaum A, Massey WA, Reiman MI, Rider B (1999) Time varying multiserver queues with abandonment and retrials. *Proceddings 16th Intenational Teletraffic Conference*, volume 4, 4–7 (Edinburgh, Scotland).

Mandelbaum A, Momčilović P, Trichakis N, Kadish S, Leib R, Bunnell CA (2019) Data-driven appointment-scheduling under uncertainty: The case of an infusion unit in a cancer center. *Management Science* 66(1):243–270.

Meijles E, de Bakker M, Groote P, Barske R (2014) Analysing hiker movement patterns using GPS data: Implications for park management. *Computers, Environment and Urban Systems* 47:44–57.

Mondschein SV, Weintraub GY (2003) Appointment policies in service operations: A critical analysis of the economic framework. *Production and Operations Management* 12(2):266–286.

Monz CA, Pickering CM, Hadwen WL (2013) Recent advances in recreation ecology and the implications of different relationships between recreation use and ecological impacts. *Frontiers in Ecology and the Environment* 11(8):441–446.

Norris JB, Kumar C, Chand S, Moskowitz H, Shade SA, Willis DR (2014) An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. *Decision Support Systems* 57:428–443.

Peng G, Liu Y, Wang J, Gu J (2017) Analysis of the prediction capability of web search data based on the HE-TDC method — prediction of the volume of daily tourism visitors. *Journal of Systems Science and Systems Engineering* 26(2):163–183.

Pickering CM (2010) Ten factors that affect the severity of environmental impacts of visitors in protected areas. *AMBIO* 39:70–77.

Pinedo M, Zacharias C, Zhu N (2015) Scheduling in the service industries: An overview. *Journal of Systems Science and Systems Engineering* 24:1–48.

Rice WL, Park SY, Pan B, Newman P (2019) Forecasting campground demand in US national parks. *Annals of Tourism Research* 75:424–438.

Schacht M (2018) Improving same-day access in primary care: Optimal reconfiguration of appointment system setups. *Operations Research for Health Care* 18:119–134.

Schütz HJ, Kolisch R (2013) Capacity allocation for demand of different customer-product-combinations with cancellations, no-shows, and overbooking when there is a sequential delivery of service. *Annals of Operations Research* 206:401–423.

Wang J, Fung RY (2015) Dynamic appointment scheduling with patient preferences and choices. *Industrial Management & Data Systems* 115(4):700–717.

Whittaker D, Shelby B, Manning R, Cole D, Haas G (2011) Capacity reconsidered: Finding consensus and clarifying differences. *Journal of Park and Recreation Administration* 29(1):1–20.

Yom-Tov GB, Mandelbaum A (2014) Erlang-R: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing & Service Operations Management* 16(2):167–327.

Yom-Tov GB, Yedidsion L, Xie Y (2021) An invitation control policy for proactive service systems: Balancing efficiency, value, and service level. *Manufacturing & Service Operations Management* 23(5):1077–1095.

Yung L, Cole DN, Graber DM, Parsons DJ, Tonnessen KA (2010) Changing policies and practices: The challenge of managing for naturalness. *Beyond Naturalness: Rethinking Park and Wilderness Stewardship in an Era of Rapid Change*, 67–83 (Island Press, Washington DC).

Zacharias C, Pinedo M (2013) Appointment scheduling with no-shows and overbooking. *Production and Operations Management* 23(5):788–801.

Zychlinski N, Mandelbaum A, Momčilović P, Cohen I (2020) Bed blocking in hospitals due to scarce capacity in geriatric institutions — cost minimization via fluid models. *Manufacturing & Service Operations Management* 22(2):223–428.

EC.1. Demand for NPs

INPA data shows that 25% of all reservations were for five popular NPs, out of more than 70 parks. Figure EC.1 shows the percentage of reservations made to the 15 most visited parks out of the total number of reservations during May–Dec. 2020. (A reservation may include multiple visitors.) According to the INPA data, 6.8% of the reservations include more than one site. Figure EC.2

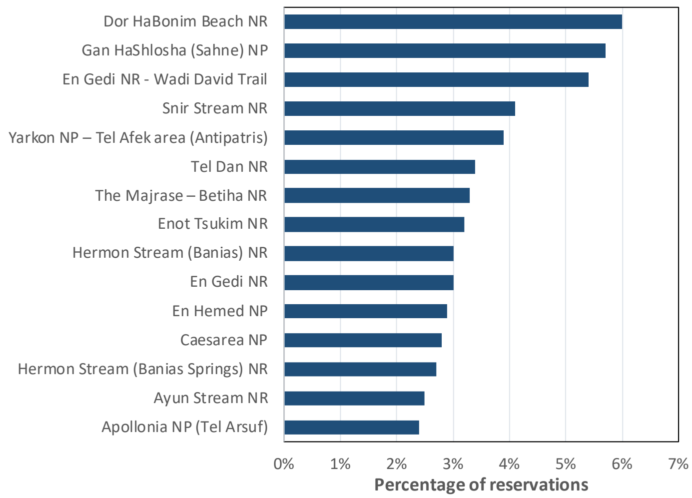


Figure EC.1 Top 15 popular NPs in Israel [May–December 2020].

shows 15 most popular site bundles (of 2 or 3 NPs) in Israel's NPs system during May–Dec. 2020.

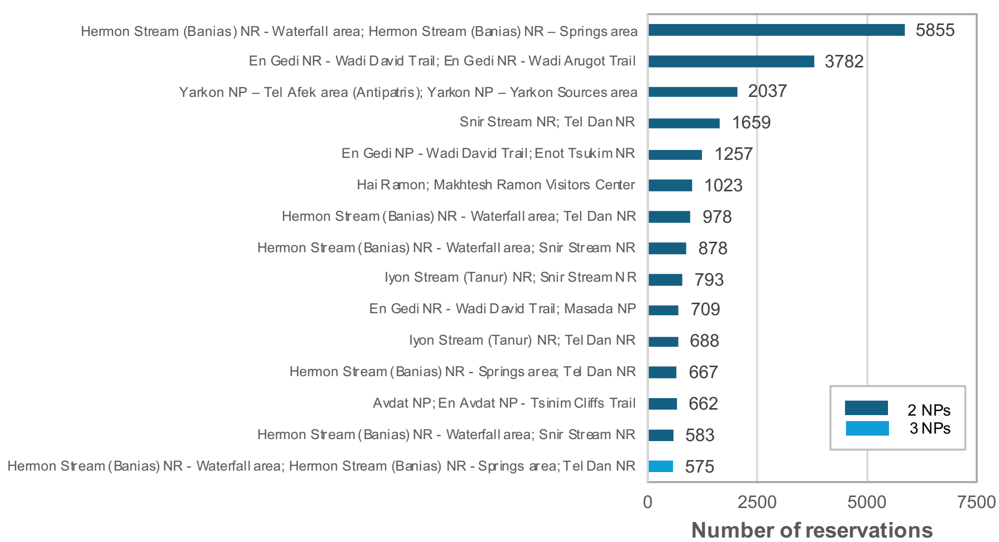


Figure EC.2 Popular NP bundles in Israeli NPs [May–December 2020].

EC.2. Proofs

EC.2.1. Proof of Theorem 1

We can rewrite $C(\Lambda)$ in the following way:

$$\begin{aligned}
C(\Lambda) &= \int_0^T \left(c^o(R(t)-L)^+ + c^u(L-R(t))^+\right) dt \\
&= (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}(R(t)-L)^+ - \frac{c^u}{c^o+c^u}(R(t)-L)^-\right) dt \\
&= (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}(R(t)-L)^+ - \left(1-\frac{c^o}{c^o+c^u}\right)(R(t)-L)^-\right) dt \\
&= (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}(R(t)-L)^+ + \frac{c^o}{c^o+c^u}(R(t)-L)^- - (R(t)-L)^-\right) dt \\
&= (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}(R(t)-L) - (R(t)-L)^-\right) dt
\end{aligned}$$

The first transition is because $(L-R(t))^+ = \max\{0, L-R(t)\} = -\min\{0, R(t)-L\} = (R(t)-L)^-$. We can rewrite this function in the following way:

$$(L-R(t))^+ = \int_{x=R(t)}^{\infty} \mathbb{1}_{\{x\le L\}} dx = -(R(t)-L)^-.$$

Therefore,

$$\begin{aligned}
C(\Lambda) &= (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}(R(t)-L) - (R(t)-L)^-\right) dt \\
&= (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}(R(t)-L) + \int_{x=R(t)}^{\infty} \mathbb{1}_{\{x\le L\}} dx\right) dt \\
&= -c^oLT + (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}R(t) + \int_{x=R(t)}^{\infty} \mathbb{1}_{\{x\le L\}} dx\right) dt \\
&= -c^oLT + c^o\int_0^T R(t)dt + (c^o+c^u)\int_0^T \left(\int_{x=0}^{\infty} \mathbb{1}_{\{x\le L\}} dx - \int_{x=0}^{R(t)} \mathbb{1}_{\{x\le L\}} dx\right) dt \\
&= -c^oLT + c^o\int_0^T R(t)dt + (c^o+c^u)\int_0^T \left(L - \int_{x=0}^{R(t)} \mathbb{1}_{\{x\le L\}} dx\right) dt \\
&= c^uLT + c^o\int_0^T R(t)dt - (c^o+c^u)\int_0^T \int_{x=0}^{R(t)} \mathbb{1}_{\{x\le L\}} dxdt \\
&= c^uLT + (c^o+c^u)\int_0^T \left(\frac{c^o}{c^o+c^u}R(t) - \int_{x=0}^{R(t)} \mathbb{1}_{\{x\le L\}} dx\right) dt \\
&= c^uLT + (c^o+c^u)\int_0^T \int_{x=0}^{R(t)} \left(\frac{c^o}{c^o+c^u} - \mathbb{1}_{\{x\le L\}}\right) dxdt \\
&= c^uLT + (c^o+c^u)\int_0^T \int_{x=0}^{\Lambda R^{\%}(t)} \left(\frac{c^o}{c^o+c^u} - \mathbb{1}_{\{x\le L\}}\right) dxdt \\
&= c^uLT + (c^o+c^u)\int_0^T \int_{x=0}^{\Lambda} R^{\%}(t) \left(\frac{c^o}{c^o+c^u} - \mathbb{1}_{\{xR^{\%}(t)\le L\}}\right) dxdt
\end{aligned}$$

$$
\begin{aligned}
&= c^u LT + (c^o + c^u) \int_{x=0}^{\Lambda} \int_{t=0}^{T} R^{\%}(t) \left(\frac{c^o}{c^o + c^u} - \mathbb{1}_{\{xR^{\%}(t) \leq L\}} \right) dt dx \\
&= c^u LT + (c^o + c^u) \int_{x=0}^{\Lambda} \left(\frac{c^o}{c^o + c^u} \int_{t=0}^{T} R^{\%}(t) dt - \int_{t=0}^{T} R^{\%}(t) \mathbb{1}_{\{xR^{\%}(t) \leq L\}} dt \right) dx \\
&= constant + (c^o + c^u) \int_{x=0}^{\Lambda} (Z - g(x)) dx,
\end{aligned}
$$

where

$$
Z = \frac{c^o}{c^o + c^u} \int_{t=0}^{T} R^{\%}(t) dt \quad \text{and} \quad g(x) = \int_{t=0}^{T} R^{\%}(t) \mathbb{1}_{\{xR^{\%}(t) \leq L\}} dt. \tag{EC.1}
$$

We need to prove that the function $G(\Lambda) = \int_{x=0}^{\Lambda} (Z - g(x))\, dx$ is minimized by Λ^*.

Assuming T is one cycle (i.e., one day) and that the system works with infinite identical cycles $\{(0,T),(T,2T),...\}$, by the definition of $\lambda^{\%}$ as the proportion of arrivals throughout the hours of the day, for any random variable X (with $f(x)$ as its PDF function) the sample path $\int_{t=0}^{T} \lambda^{\%}(t-X)dt = \int_{t=-X}^{T-X} \lambda^{\%}(t)dt = 1$. Therefore,

$$
\begin{aligned}
&\int_{t=0}^{T} E\left[\lambda^{\%}(t-X)\right] dt = \int_{t=0}^{T} \int_{x=0}^{\infty} \lambda^{\%}(t-x) f(x) dx dt = \int_{x=0}^{\infty} f(x) \int_{t=0}^{T} \lambda^{\%}(t-x) dt dx \\
&= \int_{x=0}^{\infty} f(x) dx = 1.
\end{aligned}
$$

Therefore, the function

$$
Z = \frac{c^o}{c^o + c^u} \int_{t=0}^{T} R^{\%}(t) dt = \frac{c^o}{c^o + c^u} E[S] \int_{t=0}^{T} E\left[\lambda^{\%}(t - S_e)\right] dt = \frac{c^o}{c^o + c^u} E[S]. \tag{EC.2}
$$

Z does not depend on x (i.e., Λ) and gets values in the range $[0, E[S]]$.

We note that

$$
g(x) = \int_{t=0}^{T} R^{\%}(t) \mathbb{1}_{\{xR^{\%}(t) \leq L\}} dt = \int_{t=0}^{T} R_i^{\%}(t) \mathbb{1}_{\{xR_i^{\%}(t) \leq L\}} dt \tag{EC.3}
$$

When $x = 0$, the function $g(x)$ equals $\int_{t=0}^{T} E\left[\lambda^{\%}(t-X)\right] E[S] dt = E[S]$ and when $x \to \infty$, $g(x)$ goes to 0, i.e., $g(x)$ is non-increasing non-negative function going from $E[S]$ to 0.

The function $G(\Lambda)$, for Λ starting from 0, is first an integral of a non-positive integrand, and thus is decreasing in Λ. Then, after the first Λ for which $g(\Lambda) = Z$, it is increasing. This proves that $G(\Lambda)$ is minimized (globally) at point Λ^*, where $g(\Lambda^*) = Z$, and that a solution for this equation exists. One can find Λ^* numerically by solving:

$$
\int_{t=0}^{T} R_i^{\%}(t) \mathbb{1}_{\left\{\Lambda^* R_i^{\%}(t) \leq L\right\}} dt = \frac{c^o}{c^o + c^u} E[S].
$$

Since $R_i^{\%}(t)$ is the increasing rearrangement of $R^{\%}(t)$, it is nondecreasing in t. As x increases, the set of t-values for which $R_i^{\%}(t) \leq x$ can only grow (or stay the same), which implies $\Phi(x)$ is a nondecreasing function of x. It is strictly increasing if and only if $R_i^{\%}(t)$ is strictly increasing in t almost everywhere.

Then, (3) is equivalent to

$$\Phi\left(\frac{L}{\Lambda^*}\right) = \frac{c^o}{c^o + c^u} E[S].$$

Provided Φ is invertible on the relevant range, we obtain

$$\frac{L}{\Lambda^*} = \Phi^{-1}\left(\tfrac{c^o}{c^o+c^u} E[S]\right) \implies \Lambda^* = \frac{L}{\Phi^{-1}\left(\tfrac{c^o}{c^o+c^u} E[S]\right)}.$$

□

Note that this solution is similar but not identical to the solution of Zychlinski et al. (2020). They found L^* that minimize similar cost function, and proved that $L^* = R_d(c^u T/(c^u + c^o))$, where $R_d(t)$ is a decreasing sorted version of $R(t)$.

EC.2.2. Proof of Theorem 2

To prove Theorem 2, we define Lemma EC.1 and Theorem EC.1, each of which treats a different condition on the value of f at day t.

Equivalently, we can express Eq. (5) using the percentage of accepted reservations Q_t out of total demand D_t, denoted as z_t; that is, $z_t = \frac{Q_t}{D_t}$, and $\mathbf{z} = (z_1, ..., z_T)$ the vector of z's. Then, we can rewrite Eq. (5) as

$$\begin{aligned}
&\min_{\mathbf{z}} \sum_{t=1}^{T} c_t^b (D_t - z_t \cdot D_t) + c^u (\Lambda - I_T^e)^+ + c^o (I_T^e - \Lambda)^+ \qquad \text{(EC.4)}\\
&s.t.\\
&I_t^e = I_{t-1}^e + z_t \cdot D_t (1 - p_t)(1 - \varphi_t), \qquad \forall t \in 1, ..., T;\\
&I_0^e = 0;\\
&z_t \in [0, 1], \qquad \forall t \in 1, ..., T.
\end{aligned}$$

Note that $I_T^e = \sum_{t=0}^{T} z_t D_t (1 - p_t)(1 - \varphi_t)$, therefore, (EC.4) could be simplified to:

$$\begin{aligned}
&\min_{\mathbf{z}} \sum_{t=1}^{T} c_t^b (D_t - z_t \cdot D_t) + c^u \left(\Lambda - \sum_{t=0}^{T} z_t D_t (1 - p_t)(1 - \varphi_t)\right)^+ + c^o \left(\sum_{t=0}^{T} z_t D_t (1 - p_t)(1 - \varphi_t) - \Lambda\right)^+\\
&s.t.\\
&z_t \in [0, 1], \quad \forall t \in 1, ..., T. \qquad \text{(EC.5)}
\end{aligned}$$

Lemma EC.1. *Assume that for a specific day t, $f_t > c^o$ and that I_{t-1}^e has some arbitrary value. Then, the optimal policy at day t is $z_t = 1$, that is, $Q_t^* = D_t$.*

Proof of Lemma EC.1: We first incorporate the constraints of the optimization problem (EC.4) into the objective function using the Lagrange multipliers method using the simplified version EC.5. The Lagrangian function is

$$\mathcal{L}(z_t, \beta_t, \gamma_t) = \sum_{t=1}^{T} c_t^b (D_t - z_t \cdot D_t) + c^u \left(\Lambda - \sum_{t=0}^{T} z_t D_t (1 - p_t)(1 - \varphi_t)\right)^+ \qquad \text{(EC.6)}$$

$$+c^o\left(\sum_{t=0}^{T} z_t D_t(1-p_t)(1-\varphi_t)-\Lambda\right)^+ - \sum_{t=1}^{T}\beta_t z_t + \sum_{t=1}^{T}\gamma_t(z_t-1),$$

where β_t, and γ_t are the Lagrange multipliers.

The first-order necessary conditions (which are the Karush–Kuhn–Tucker (KKT) conditions (Karush 1939, Kuhn and Tucker 1951)) of Eq. (EC.6) are:

$$\frac{\partial \mathcal{L}}{\partial z_t} = -c_t^b D_t - c^u D_t(1-p_t)(1-\varphi_t)\mathbb{1}_{\left\{\sum_{\tau=0}^{T} z_\tau D_\tau(1-p_\tau)(1-\varphi_\tau)\le\Lambda\right\}}$$
$$+c^o D_t(1-p_t)(1-\varphi_t)\mathbb{1}_{\left\{\sum_{\tau=0}^{T} z_\tau D_\tau(1-p_\tau)(1-\varphi_\tau)>\Lambda\right\}} - \beta_t + \gamma_t = 0, \qquad \forall t \in 1,\ldots,T; \qquad \text{(EC.7)}$$

$$z_t \ge 0, \qquad \forall t \in 1,\ldots,T; \qquad \text{(EC.8)}$$

$$z_t \le 1, \qquad \forall t \in 1,\ldots,T; \qquad \text{(EC.9)}$$

$$\beta_t \ge 0, \qquad \forall t \in 1,\ldots,T; \qquad \text{(EC.10)}$$

$$\gamma_t \ge 0, \qquad \forall t \in 1,\ldots,T; \qquad \text{(EC.11)}$$

$$\beta_t z_t = 0, \qquad \forall t \in 1,\ldots,T; \qquad \text{(EC.12)}$$

$$\gamma_t(z_t-1) = 0, \qquad \forall t \in 1,\ldots,T. \qquad \text{(EC.13)}$$

Next, we will review all possible cases for β_t and γ_t when $\alpha_t \in \mathbb{R}$. Reconciling the necessary condition (EC.12) with condition (EC.13), we get that if $\beta > 0$ then $\gamma = 0$ and if $\beta = 0$ then $\gamma \ge 0$. The combination of $\beta_t > 0$ and $\gamma_t > 0$ yields $z_t = 0$ by (EC.12) and $z_t = 1$ by (EC.13), which is a contradiction.

We divide the case of $\beta > 0$ and $\gamma = 0$ into two subcases: (a) a "Below Λ case", where $\sum_{\tau\neq t} z_\tau D_\tau(1-p_\tau)(1-\varphi_\tau) + D_t(1-p_t)(1-\varphi_t) \le \Lambda$ and (b) an "Above Λ case", where $\sum_{\tau\neq t} z_\tau D_\tau(1-p_\tau)(1-\varphi_\tau) + D_t(1-p_t)(1-\varphi_t) > \Lambda$. In the Below Λ case, condition (EC.7) is not applicable, because Eq. (EC.7) reduces to $-c_t^b D_t - c^u D_t(1-p_t)(1-\varphi_t) = \beta_t$ which contradicts the case assumption that $\beta > 0$. In the Above Λ case, condition (EC.7) is also not applicable, because Eq. (EC.7) reduces to $-c_t^b D_t + c^o D_t(1-p_t)(1-\varphi_t) = \beta_t > 0$, which implies that $-c_t^b + c^o(1-p_t)(1-\varphi_t) > 0$ and contradicts the assumption that $f_t > c^o$. Hence, $\beta > 0$ and $\gamma = 0$ cannot be a feasible solution to (EC.4) by Lemma EC.1's conditions.

Similarly, the case of $\beta = 0$ and $\gamma = 0$ has no feasible solution to Eq. (EC.4) when $f_t > c^o$. To prove this, we again divide case into two subcases. In the Below Λ case, condition (EC.7) is not applicable because Eq. (EC.7) reduces to $-c_t^b D_t - c^u D_t(1-p_t)(1-\varphi_t) = \beta_t$, which contradicts the assumption that $\beta = 0$. In the Above Λ case, Eq. (EC.7) reduces to $-c_t^b D_t + c^o D_t(1-p_t)(1-\varphi_t) = 0$, which implies that $f_t = c^o$, which contradicts Lemma EC.1.

Therefore, the only combination of the Lagrangian multipliers with feasible solutions to Lemma EC.1 is when $\beta_t = 0, \gamma_t > 0$. In this case, by (EC.12) and (EC.13), $z_t = 1$. This combination case can be divided into two subcases to confirm that Eq. (EC.7) holds true under it:

(a) **[Below Λ case]** $\sum_{\tau \neq t} z_\tau D_\tau(1-p_\tau)(1-\varphi_\tau) + D_t(1-p_t)(1-\varphi_t) \leq \Lambda$:

Applying the Below Λ case to Eq. (EC.7) results in $-c_t^b D_t - c^u D_t(1-p_t)(1-\varphi_t) + \gamma_t = 0$. Since $\gamma_t > 0$, it follows that $c_t^b D_t + c^u D_t(1-p_t)(1-\varphi_t) > 0$, which can be satisfied by any set of parameters (as all of the parameters are nonnegative).

(b) **[Above Λ case]** $\sum_{\tau \neq t} z_\tau D_\tau(1-p_\tau)(1-\varphi_\tau) + D_t(1-p_t)(1-\varphi_t) > \Lambda$:

Applying the Above Λ case to Eq. (EC.7) results in $-c_t^b D_t + c^o D_t(1-p_t)(1-\varphi_t) + \gamma_t = 0$. Since $\gamma_t > 0$, it follows that $c_t^b D_t - c^o D_t(1-p_t)(1-\varphi_t) > 0$. The latter implies that $c^o < f_t$, which always holds under Lemma EC.1.

We summarize the case of $\beta = 0$ and $\gamma > 0$: since $z_t = 1$, by Eq. (EC.8) it follows that $I_t^e = I_{t-1}^e + D_t(1-p_t)(1-\varphi_t)$. In other words, for all t when $f_t > c^o$, all demand will be accepted: the number of reservations will be $Q_t^* = D_t$. This concludes the proof of Lemma EC.1. □

We now move to a complementary assumption that $f_t \leq c^o$.

THEOREM EC.1. *Assume that $f_t \leq c^o$ for all t, and that $I_0 + \sum_{t=1}^{T} D_t(1-p_t)(1-\varphi_t) > \Lambda$. Let, $\hat{f}_s$ be the sorted vector of f_t from the highest to the lowest value. Then, the optimal sorted solution to (EC.4) is $\hat{\mathbf{Q}}^* = \{\hat{D}_1, \hat{D}_2, ..., \hat{D}_{S-1}, \frac{\Lambda - \hat{I}_{S-1}^e}{(1-\hat{p}_S)(1-\hat{\varphi}_S)}, 0, ..., 0\}$, where S is the index in which $I_0 + \sum_{s=1}^{S-1} \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) < \Lambda$ and $I_0 + \sum_{s=1}^{S} \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) \geq \Lambda$.*

In order to prove this theorem, we use the following lemma. In Lemma EC.2, we show that the sorting of f_t needs to be done in order to prioritize the days from which we accept reservations. Then, we go back to Theorem EC.1 and prove that one would accept exactly Λ *effective reservations* (i.e., the number of reservations that will ensure that the number of arrivals after cancellations and no-shows will be exactly Λ).

LEMMA EC.2. *For every two periods i and j, where $c^o \geq f_i > f_j$, the marginal cost of accepting one effective unit in period i is lower than that of accepting one effective unit in period j.*

Proof of Lemma EC.2: Assume that we have an optimal policy π with total costs V_π, where in period i we block one *effective reservation* and in period j we accept one *effective reservation*. We compute the total cost of conversely one effective reservation. By blocking one effective reservation of period j, we incur a cost of $c_j^b/((1-p_j)(1-\varphi_j))$, and by not blocking one effective reservation of period i, the blocking cost $c_i^b/((1-p_i)(1-\varphi_i))$ is avoided. There is no change in the number of visitors arriving to the park, since the total number of *effective* reservations remains unchanged. Therefore, there is no change in the under- and over-costs. Hence, the total cost after the reversal is $V_\pi + \frac{c_j^b}{(1-p_j)(1-\varphi_j)} - \frac{c_i^b}{(1-p_i)(1-\varphi_i)} = V_\pi + f_j - f_i$. This cost is less than V_π (since $f_i > f_j$), which contradicts the assumption that this policy, π, was optimal. □

According to Lemma EC.2, the PA prefers to accept reservations with higher f_t. Recall that for every period with $f_t > c^o$ we accept all reservations by Lemma EC.1. Therefore, without loss of generality, we assume that for all the periods in Theorem EC.1 $f_t \leq c^o$ and that they are sorted from highest to lowest values.

We now go back to proving Theorem EC.1.

Proof of Theorem EC.1: We can rewrite Eq. (5) using the sorted indexes $s \in 1, ..., T$ as

$$\min_{\hat{z}_s} \sum_{s=1}^{T} \hat{c}_s^b(\hat{D}_s - \hat{z}_s\hat{D}_s) + c^u(\Lambda - \hat{I}_T^e)^+ + c^o(\hat{I}_T^e - \Lambda)^+ \tag{EC.14}$$

$$\begin{aligned} s.t.& \\ &\hat{I}_s^e = \hat{I}_{s-1}^e + \hat{z}_s \cdot \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s), && \forall s \in 1, ..., T; \\ &\hat{I}_0^e = 0; \\ &\hat{z}_s \in [0,1], && \forall s \in 1, ..., T. \end{aligned}$$

We again use Lagrangian multipliers with similar KKT conditions to a simplified version:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \hat{z}_s} &= -\hat{c}_s^b\hat{D}_s - c^u\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)\mathbb{1}_{\left\{\sum_{\tau=0}^{T}\hat{z}_\tau\hat{D}_\tau(1-\hat{p}_\tau)(1-\hat{\varphi}_\tau)\leq\Lambda\right\}} \\ &\quad + c^o\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)\mathbb{1}_{\left\{\sum_{\tau=0}^{T}\hat{z}_\tau\hat{D}_\tau(1-\hat{p}_\tau)(1-\hat{\varphi}_\tau)>\Lambda\right\}} - \beta_s + \gamma_s = 0, && \forall s \in 1, ..., T; && \text{(EC.15)} \\ &\hat{z}_s \geq 0, && \forall s \in 1, ..., T; && \text{(EC.16)} \\ &\hat{z}_s \leq 1, && \forall s \in 1, ..., T; && \text{(EC.17)} \\ &\beta_s \geq 0, && \forall s \in 1, ..., T; && \text{(EC.18)} \\ &\gamma_s \geq 0, && \forall s \in 1, ..., T; && \text{(EC.19)} \\ &\beta_s\hat{z}_s = 0, && \forall s \in 1, ..., T; && \text{(EC.20)} \\ &\gamma_t(\hat{z}_s - 1) = 0, && \forall s \in 1, ..., T. && \text{(EC.21)} \end{aligned}$$

Since $\hat{f}_s$ is sorted, according to Lemma EC.2 we accept quantities sequentially. As a result, if we accept demand in an s-index period, then under the optimal policy we also accept all reservations in indexes $\{1, ..., s-1\}$. In the same way, if we rejected demand in index s, then under the optimal policy we reject all the demand in indexes $\{s+1, .., T\}$ as well. Therefore, under the optimal policy, the indicator of $\mathbb{1}_{\left\{\sum_{\tau=0}^{T}\hat{z}_\tau\hat{D}_\tau(1-\hat{p}_\tau)(1-\hat{\varphi}_\tau)>\Lambda\right\}}$ is equivalent to the indicator $\mathbb{1}_{\left\{\hat{I}_{s-1}^e+\hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)>\Lambda\right\}}$. Hence, we can rewrite Eq. (EC.15) by

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \hat{z}_s} &= -\hat{c}_s^b\hat{D}_s - c^u\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)\mathbb{1}_{\left\{\hat{I}_{s-1}^e+\hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)\leq\Lambda\right\}} \\ &\quad + c^o\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)\mathbb{1}_{\left\{\hat{I}_{s-1}^e+\hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)>\Lambda\right\}} - \beta_s + \gamma_s = 0, && \forall s \in 1, ..., T. && \text{(EC.22)} \end{aligned}$$

Define the Below Λ case such that $\hat{I}_{s-1}^e + \hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) \leq \Lambda$ and the Above Λ case such that $\hat{I}_{s-1}^e + \hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) > \Lambda$. We will analyze the Below Λ and Above Λ cases separately for each index s, $s \in 1, ..., T$.

1. **[Below Λ case]** Assume that $\hat{I}^e_{s-1}+\hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)\leq\Lambda$ for all $\hat{z}_s\in[0,1]$. In this case, Eq. (EC.22) reduces to $-c^b_s\hat{D}_s-c^u\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)-\beta_s+\gamma_s=0$. We now check four combinations of β_s and γ_s:
   (a) $\beta_s=0,\ \gamma_s=0$: Since $\beta_s=0$ and $\gamma_s=0$, it follows that $-\hat{c}^b_s\hat{D}_s-c^u\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)=0$, which cannot be satisfied by any set of parameters (because all the parameters are nonnegative).
   (b) $\beta_s>0,\ \gamma_s=0$: This combination implies that $-\hat{c}^b_s\hat{D}_s-c^u\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)>0$, which cannot be satisfied by any set of parameters (because all the parameters are nonnegative).
   (c) $\beta_s=0,\ \gamma_s>0$: This combination implies that $\hat{c}^b_s\hat{D}_s+c^u\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)=\gamma_s>0$, which is applicable when $\hat{c}^b_s+c^u(1-\hat{p}_s)(1-\hat{\varphi}_s)>0$, that is, when $\hat{f}_s+c^u>0$ which can be satisfied by any set of parameters. In this case, $\hat{z}_s$=1 by Eq. (EC.21); that is, the policy is to accept all demand for reservations.
   (d) $\beta_s>0,\ \gamma_s>0$: This combination of multiplier values cannot hold for any $\hat{z}_s$ due to (EC.20) and (EC.21). Therefore, this combination is not applicable.

   We conclude that when $\hat{f}_s\leq c^o$ in the Below Λ case, $\hat{z}_s=1$.
2. **[Above Λ case]** Assume that $\hat{I}^e_{s-1}+\hat{z}_s\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)>\Lambda$ for all $\hat{z}_s\in[0,1]$. In this case, Eq. (EC.22) reduces to $-\hat{c}^b_s\hat{D}_s+c^o\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)-\beta_s+\gamma_s=0$. We check four combinations of β_s and γ_s:
   (a) $\beta_s=0,\ \gamma_s=0$: This combination results in $-\hat{c}^b_s\hat{D}_s+c^o\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)=0$, which implies that $\hat{f}_s=c^o$ (which meets the conditions of Theorem *EC.*1). $\hat{f}_s=c^o$ implies that we are indifferent between accepting and blocking reservations; therefore, all $\hat{z}_s$ are optimal, that is, $\hat{z}^*_s=\hat{z}_s\in[0,1]$. Here, we are indifferent between solutions and, therefore, $\hat{z}^*_s=0$ (block all demand for reservations) is an optimal solution.
   (b) $\beta_s>0,\ \gamma_s=0$: This combination results in $-\hat{c}^b_s\hat{D}_s+c^o\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)>0$, which can be satisfied by any set of parameters (because all the parameters are nonnegative). This implies that $\hat{f}_s<c^o$, fitting the lemma's assumptions. In this case, $\hat{z}_s=0$ by Eq. (EC.20). Hence, all demand is blocked.
   (c) $\beta_s=0,\ \gamma_s>0$: This combination results in $-\hat{c}^b_s\hat{D}_s+c^o\hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s)<0$, which implies that $\hat{c}^b_s>c^o(1-\hat{p}_s)(1-\hat{\varphi}_s)$, that is, $\hat{f}_s>c^o$. This contradicts Theorem EC.1. Hence, this combination is not applicable.
   (d) $\beta_s>0,\ \gamma_s>0$: These combinations of multiplier values cannot hold for any $\hat{z}_s$ due to conditions (EC.20) and (EC.21). Therefore, this combination is not applicable.

   We conclude that in the Above Λ case, there exist two solutions: (1) $\hat{z}^*_s=0$ if $\hat{f}_s<c^o$ and (2) $\hat{z}^*_s$ can be any value in its range $[0,1]$ if $\hat{f}_s=c^o$. Combining the two solutions, we conclude that for any Above Λ case that holds $\hat{f}_s\leq c^o$, the optimal solution is to block all demand for reservations, that is, $\hat{Q}^*_s=0$ ($\hat{z}^*_s=0$).

3. **[Transition case]** We conclude this proof by considering an index day in which for some $\hat{z}_s$ we get the Below Λ case, $\hat{I}^e_{s-1} + \hat{z}_s \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) \leq \Lambda$, and for higher values of $\hat{z}_s$, we get the Above Λ case, $\hat{I}^e_{s-1} + \hat{z}_s \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) > \Lambda$. This is precisely index S. Theorem EC.1 suggests accepting reservations for part of the demand, such that we reach exactly Λ effective reservations, and block the rest, that is, $\hat{Q}^*_S = \frac{\Lambda - \hat{I}^e_{S-1}}{(1-\hat{p}_S)(1-\hat{\varphi}_S)}$.

   Without loss of generality, we divide the index-S period into subperiods $s' \in S$, such that in each subperiod the demand equals one effective reservation (where one effective reservation is defined by the number of reservations that will result in one visitor arriving to the park). This means that $\hat{D}_{s'}(1-\hat{p}_{s'})(1-\hat{\varphi}_{s'}) = 1$ for all $s' \in S$. According to the Below Λ case analysis above, we will accept reservations until we reach exactly Λ effective reservations at some time s'', that is, $\hat{I}^e_{s''} = \Lambda$. Thereafter, all reservations are blocked, that is, $z_{s'} = 0, \forall s' > s''$, based on the Above Λ case proof.

   Combining the subperiods, we get that the number of s' indexes when all reservations are accepted is such that $\hat{Q}^*_S = \sum_{s'=1}^{s''} \hat{D}_{s'}(1-\hat{p}_S)(1-\hat{\varphi}_S) = s'' = \frac{\Lambda - \hat{I}^e_{S-1}}{(1-\hat{p}_S)(1-\hat{\varphi}_S)}$.

Summarizing Theorem EC.1 : Since we assume that $\hat{f}_s \leq c^o$ for all s, it follows that in the Below Λ case (where $\hat{I}^e_{s-1} + \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) < \Lambda$) all reservations are accepted, that is, $\hat{Q}^*_s = \hat{D}_s$ ($\hat{z}^*_s = 1$), and that in the Above Λ case (where $\hat{I}^e_{s-1} + \hat{D}_s(1-\hat{p}_s)(1-\hat{\varphi}_s) \geq \Lambda$) all reservations are blocked, that is, $\hat{Q}^*_s = 0$ ($\hat{z}^*_s = 0$). Thus, the optimal value for $\hat{I}^e_T$ is $\hat{I}^e_T = \Lambda$. In other words, in the case where $\hat{f}_s \leq c^o$ for all s, reservations are accepted sequentially up to Λ. □

Summarizing Theorem 2 : Theorem 2 first order the periods according to f in decreasing order. For the first periods in the sorted vector, where f_s is larger than c^o, Lemma EC.1 is applied. Then for the rest of the periods, where f_s is smaller or equal c^o, Theorem EC.1 is applied. □

EC.2.3. Proofs of Propositions 1 and 2

Proof of Proposition 1: If φ_t and p_t decrease over time, then f_t decreases over time too (i.e., $\hat{f} = f$). Therefore, according to Theorem 2, the PA should accept reservations from the beginning of the time horizon onwards. By Lemma EC.1, the PA should accept all demand when $f_t > c^o$ and, by Lemma EC.2, should prefer periods with larger f, that is, from earlier in the horizon when $f \leq c^o$ (until the capacity threshold is reached). That means that regardless of the ratio between f and c^o, the PA should start accepting units as early as possible. □

Proof of Proposition 2: By Lemma EC.2, the PA prefers periods with larger f values. Specifically, if f_t is increasing in time, then the indexes of $\hat{f}_s$ are in reverse order of days in f_t. Therefore, the larger values of f_t are the latest in the horizon, and these should be prioritized. The question is when to start accepting reservations.

According to Theorem 2, we accept all reservations of periods where $f_t > c^o$. Since there are S_1 such periods (by definition), the latest we start accepting reservations is at day $T - S_1 + 1$. Then, by

Theorem 2, if $\sum_{t=T-S_1+1}^{T} D_t(1-p_t)(1-\varphi_t) < \Lambda$, we want to accept reservations in earlier periods specifically at $T-S_2+1$. Recall the definition of S_2 as the time we reach the capacity threshold. Hence, $T-S_2+1$ is the maximal time such that $\sum_{t=T-S_2+1}^{T} D_t(1-p_t)(1-\varphi_t) \geq \Lambda$. Finally, again according to Theorem 2, during period $T-S_2+1$, we accept $\frac{\Lambda-\sum_{t=T-S_2+2}^{T} D_t(1-p_t)(1-\varphi_t)}{(1-p_{T-S_2+1})(1-\varphi_{T-S_2+1})}$ reservations.
□

EC.3. Algorithm for Increasing f

The ATT algorithm for an increasing f function is defined similarly to one for a decreasing f function, except for two differences: (a) by Proposition 2, we only open and accept reservations from S_2 days from the end of the horizon (and in the first period partially), and (b) we have to keep enough permits at the end of every day for reservations made at later days. Let $D^e_{t_1\to t_2}$ denote the effective expected demand from period t_1 to t_2, such that $D^e_{t_1\to t_2} = \int_{t=t_1}^{t_2} D_t(1-p_t)(1-\varphi_t)dt$. We will use this quantity as our guidance to the number of permits that need to be reserved for later time.

Algorithm 3: The Adaptive Two-Threshold (ATT) Algorithm for Increasing f Function

1. Calculate all f_t values for every t, such that $f_t = \frac{c_t^b}{(1-p_t)(1-\varphi_t)}$.
2. Identify time $S_1 = \min\{t|f_t > c^o\}$.
3. If $D^e_{S_1\to T} \geq \Lambda$, set time $S_2 = S_1$. Otherwise, identify time $S_2 = \max\{t|D^e_{t\to T} \geq \Lambda\}$.
4. Set $t = 0$.
5. If $t \geq S_1$, accept all the demand for reservations on that time, i.e., $Q_t = D_t$.
6. If $t < S_2$, block all the demand for reservations on that time, i.e., $Q_t = 0$.
7. If $S_2 \leq t < S_1$, calculate (a) $D^e_{t^+\to T} = \int_{j=t^+}^{T} D_j(1-p_j)(1-\varphi_j)dj$, the effective demand expected to arrive in the future, and (b) $\bar{I}^e_{t^-} = \int_{i<t} \bar{Q}^t_i(1-\mathbb{H}^t_i)(1-\varphi_i)di$, the number of active reservations predicted to arrive at the NP.
   If $\bar{I}^e_{t^-} \geq \Lambda$, accept no reservations (i.e., $Q_t = 0$). Otherwise, accept $Q_t = \min\{D_t, (\Lambda - \bar{I}^e_{t^-} - D^e_{t^+\to T})/((1-p_t)(1-\varphi_t))\}$ reservations.
8. Set $t = t + dt$. If $t \leq T$, go to Step 5; otherwise, stop.

EC.4. Case Study: The Optimal Target Arrivals to a NP

In this section, we analyze the impact of different parameters on the optimal target arrivals, Λ^*. Here, we take a more realistic assumption that the service time is time-varying, ensuring that visitors exit the NP before its closing time. We also assume that the service time is given by an exponential distribution, which allow us to compute $R(t)$ by the following differential equation:

$$\frac{\partial R(t)}{t} = \lambda(\partial t) - \mu(t)R(t).$$

We then compute Λ^* by finding a numerical solution to the following optimization problem:

$$\min_{\Lambda} C(\Lambda) = \int_0^T c^o(R(t)-L)^+ + c^u(L-R(t))^+ dt \tag{EC.23}$$
$$s.t. \ \frac{\partial R(t)}{\partial t} = \Lambda\lambda^{\%}(t) - \mu(t)R(t).$$

EC.4.1. The Impact of the Cost Ratio on Optimal Target Arrivals

Figure EC.3 shows the impact of the cost ratio, $\frac{c^o}{c^o+c^u}$, on the recommended target arrivals (blue) and compares it to the maximal capacity (yellow) and the maximal number of visitors, $\max_t R(t)$, (red). As the cost ratio, $\frac{c^o}{c^o+c^u}$, increases, the target arrivals (Λ) reduces and so does the maximal number of visitors at the park. Hence, with high values of c^o compared to c^u, the visitor load ($R(t)$), under optimal values of Λ, will not exceed the maximal capacity for long periods of time. As the cost ratio approaches 1, the maximal visitor load will approach L but will not cross it. This was the situation during the COVID-19 pandemic when healthcare guidance was strict.

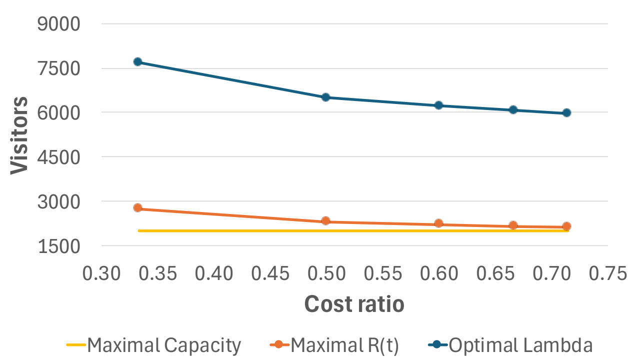


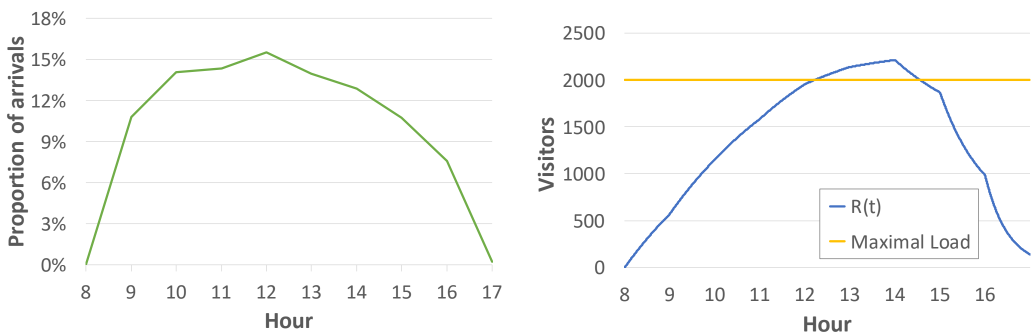
Figure EC.3 Effect of the cost ratio $\left(\frac{c^o}{c^u+c^o}\right)$ on optimal target arrivals and visitor load.

As an illustrative example, assume that $c^u = 20$ and $c^o = 30$, and that the recommended target arrivals to the park is $\Lambda = 6217$. The resulting visitor load over the day (assuming an empty park at the start of the day) is presented using a blue line in Figure 4(b).

EC.4.2. The Impact of the Maximal Capacity on Optimal Target Arrivals

Next, we vary the maximal capacity (L) allowed in the park between 1000 and 3000, with a fixed cost ratio of 0.6. Figure EC.5 shows the visitor load (solid lines) throughout the day ($R(t)$) for four values of the maximal capacity (dashed).

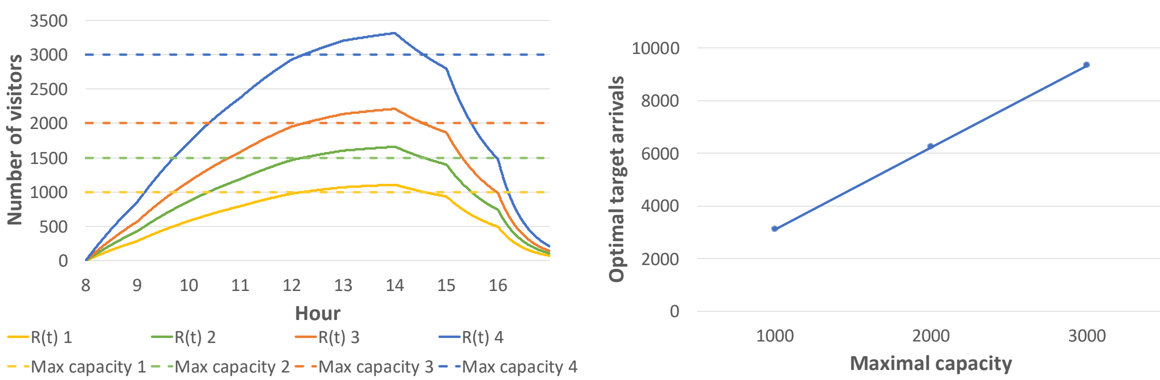
The impact of L is opposite to the impact of the cost ratio—as the maximal capacity increases, the optimal target arrivals increases. Therefore, as the maximal capacity increases, the number of visitors at any given time observed in Figure EC.5 increases too. An increase of the maximal capacity from 1000 to 2000 (200%) increased the optimal target arrivals from 3108 to 6217 (also 200%), and when the maximal capacity increases from 2000 to 3000 (another 150%), the optimal target arrivals increases from 6217 to 9325 (also 150%) (see Figure 5(b)).



(a) $\lambda^{\%}$ in a typical summer day (Aug 2020)

(b) Average visitor load

Figure EC.4 En Gedi NP: Average dynamics during a typical day with optimal target arrivals of $\Lambda = 6217$ and maximum capacity of $L = 2000$ (cost ratio 0.6).



(a) Maximal capacity and visitor load

(b) Optimal target arrivals as a function of maximal capacity

Figure EC.5 Effect of the maximal capacity on visitor load and optimal target arrivals [Cost ratio 0.6].

EC.4.3. The Impact of LOS on Optimal Target Arrivals

The visitors' average LOS is also a significant contributor to the NP visitor load. As visitors stay longer at the park, the number of visitors at the park at any given time increases. In previous analyses, we saw that for a cost ratio of 0.6 and a maximal capacity of 2000, the optimal target arrivals is $\Lambda = 6217$, when the LOS was 3 hours in general (see Table **??**). Here we change the visitor's LOS—either increasing it to 5 hours or decreasing it to 1 hour. See Table **??** for the exact time-varying visiting duration we use for this test.

We find that when we increase LOS to 5 hours, for the same given cost ratio and maximal acceptable load capacity, the optimal target arrivals decreases to $\Lambda = 5165$. That is, an increase of 67% in visit duration (from 3 to 5 hours) translated to a 17% decrease in the recommended target number of visitors entering the park (from 6217 to 5165). This is quite intuitive and expected, since in general queueing systems, an increase in service duration increases system load. Using a 1-hour LOS, we find consistent results. A decrease of 67% in visit duration (from 3 to 1) translates to a

139% increase in the recommended target number of visitors to the park (from 6217 to 14,832). To summarize, the LOS has a significant impact on the optimal number of total visitors that should be allowed to enter the NP during a day.

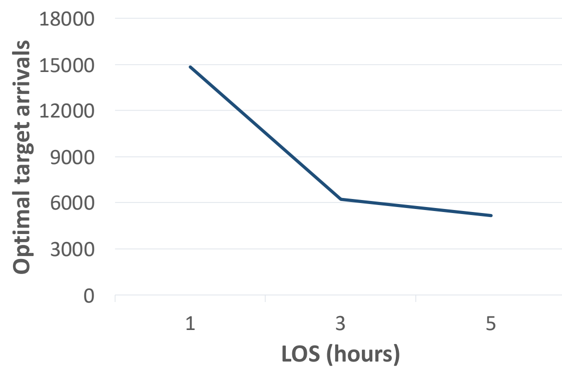
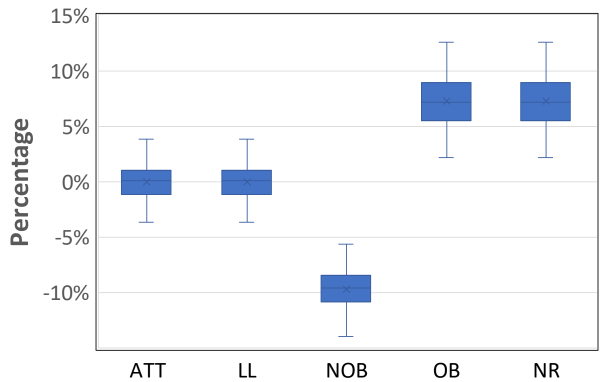


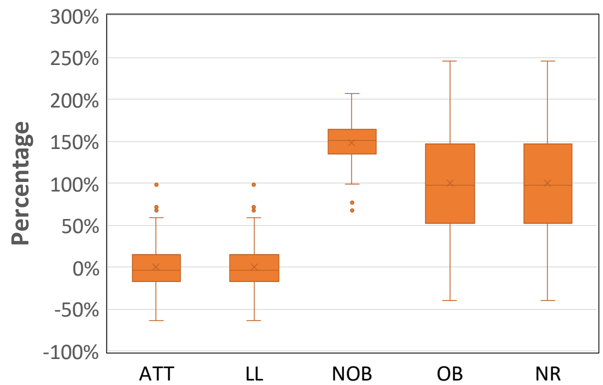
Figure EC.6 **Effect of the LOS on optimal target arrivals.**

EC.5. Simulation Study: Comparison of Policy Performance

Combination 1 ($c^b = 4,\ c^u = 3,\ c^o = 12$)

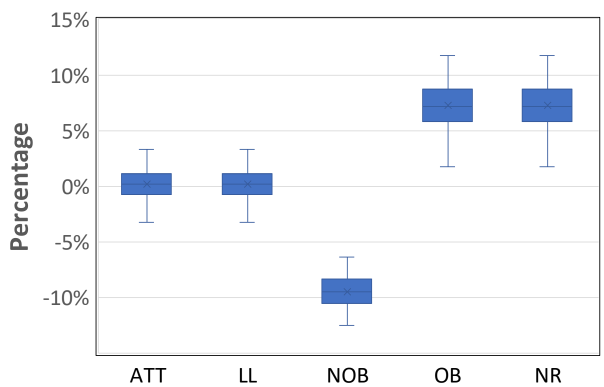


(a) Gap percentage between arrivals and Λ

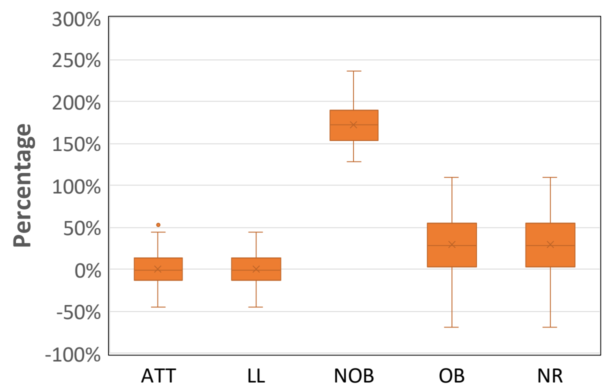


(b) Gap percentage between V_π and $E(V_{ATT})$

Combination 2 ($c^b = 4,\ c^u = 3,\ c^o = 7$)

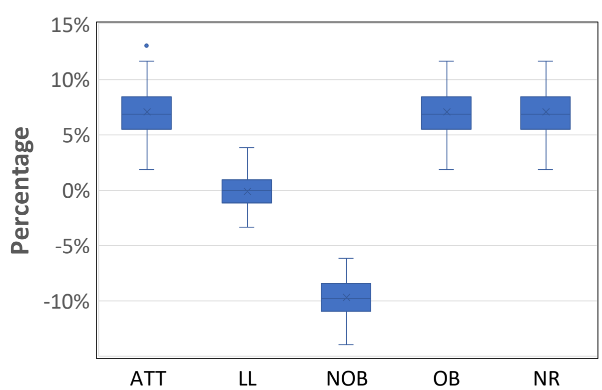


(c) Gap percentage between arrivals and Λ

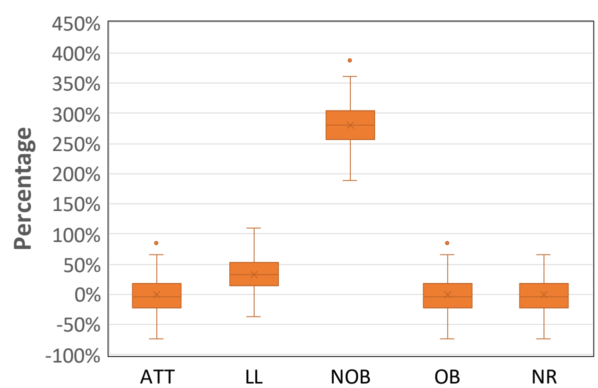


(d) Gap percentage between V_π and $E(V_{ATT})$

Combination 3 ($c^b = 4,\ c^u = 3,\ c^o = 4$)

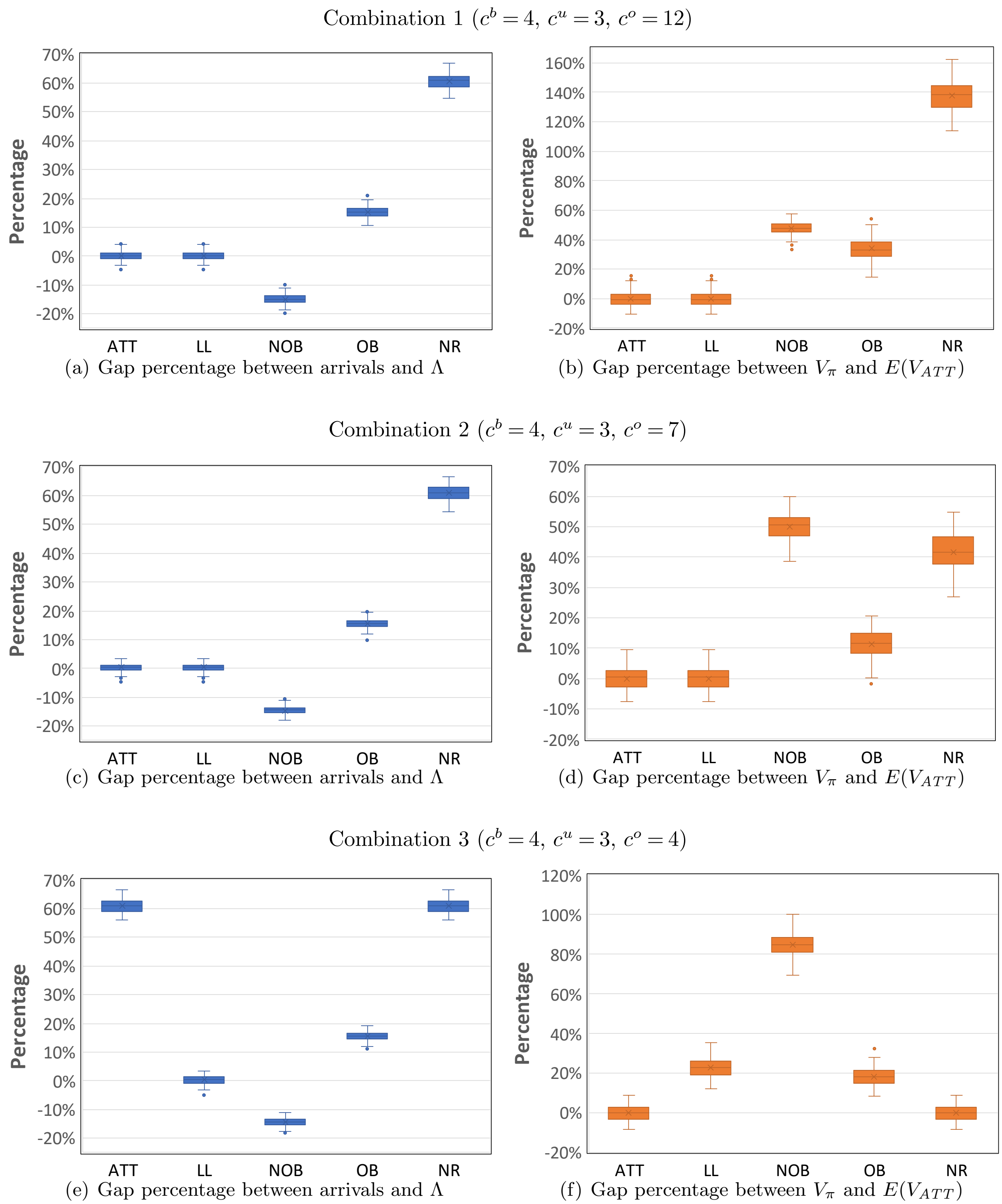


(e) Gap percentage between arrivals and Λ



(f) Gap percentage between V_π and $E(V_{ATT})$

Figure EC.7 Comparison of policy performance by combinations (total demand $\sim N(2000, 20)$).



(a) Gap percentage between arrivals and Λ

(b) Gap percentage between V_π and $E(V_{ATT})$

(c) Gap percentage between arrivals and Λ

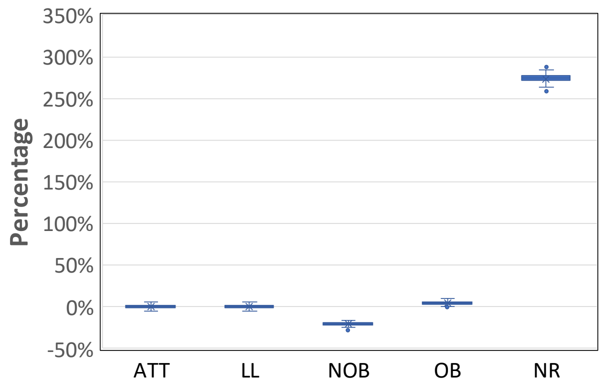
(d) Gap percentage between V_π and $E(V_{ATT})$

(e) Gap percentage between arrivals and Λ

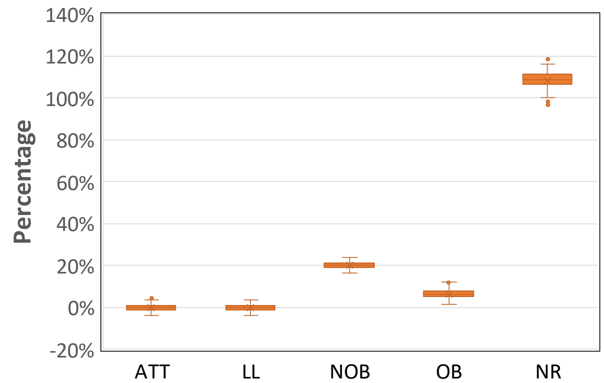
(f) Gap percentage between V_π and $E(V_{ATT})$

Figure EC.8 Comparison of policy performance by combinations (total demand $\sim N(3000, 30)$**).**

Combination 1 ($c^b = 4,\ c^u = 3,\ c^o = 12$)

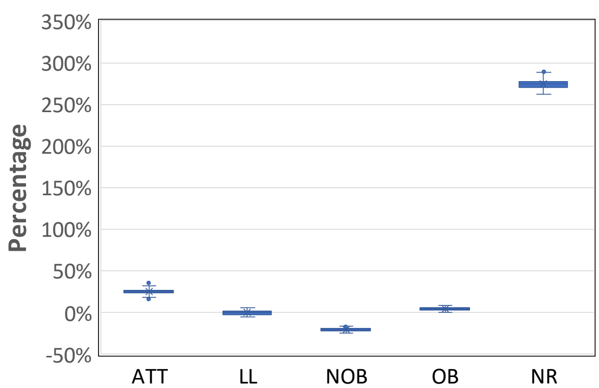


(a) Gap percentage between arrivals and Λ

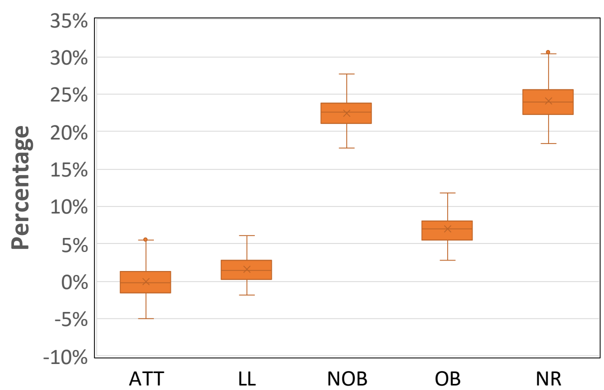


(b) Gap percentage between V_π and $E(V_{ATT})$

Combination 2 ($c^b = 4,\ c^u = 3,\ c^o = 7$)

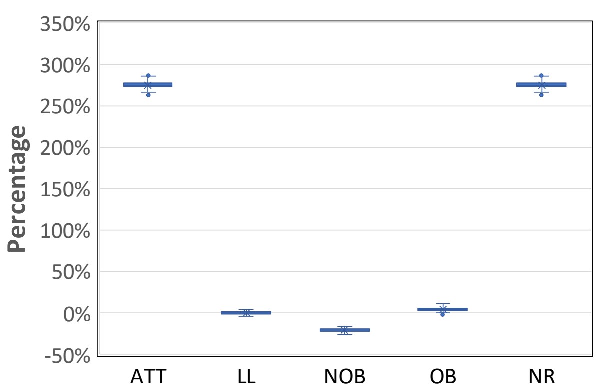


(c) Gap percentage between arrivals and Λ

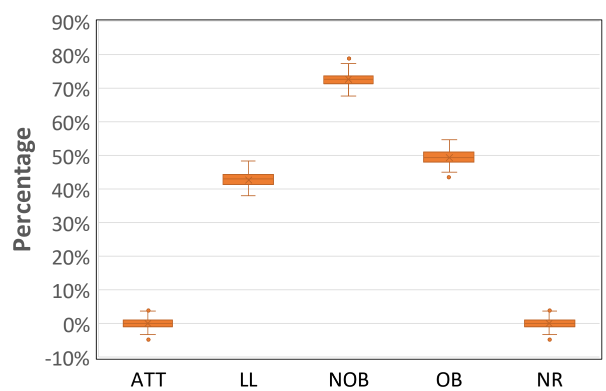


(d) Gap percentage between V_π and $E(V_{ATT})$

Combination 3 ($c^b = 4,\ c^u = 3,\ c^o = 4$)



(e) Gap percentage between arrivals and Λ



(f) Gap percentage between V_π and $E(V_{ATT})$

Figure EC.9 **Comparison of policy performance by combinations (total demand $\sim N(7000, 70)$).**