Optimizing Peak Hour Rolling Stock

Dominik Baus (614414), Iryna Matsiuk (598762),
Anja Petric (598370), and Boris Rine (613746)

Abstract

To minimize the costs of public transport systems, it is crucial to have programs in place that automate the scheduling and allocation process to make it more efficient. This paper explores the 8 o'clock rolling stock assignment problem, focusing on minimizing rolling stock costs while considering varying passenger demands. Mirroring common practice for rolling stock assignment, this study takes the train schedule with starting and ending locations as given. We propose several formulations that model deterministic and stochastic passenger demand and analyze them using data provided by Erasmus University Rotterdam. By allowing rolling stock to be reused and by introducing a two class system, we extend our model making it more applicable in practice. Our results suggest that introducing stochastic components and time-varying demand increases operational fixed costs, underscoring the importance of robust scheduling strategies. The findings of this research offer valuable insights into the efficient allocation of rolling stock.

Date: April 19, 2024

1 Introduction

This paper addresses the 8 o'clock rolling stock assignment problem (8RSAP), optimizing train scheduling and rolling stock management to accommodate fluctuating passenger demands in the Netherlands. Since the onset of the pandemic, there has been a significant shift in peak travel times associated with an increase in remote work, affecting traditional commuting patterns. This has led to variations in passenger load, particularly noticeable on Tuesdays and Thursdays, which have emerged as the most popular office days, contrasting with lower demand on other weekdays.

Given these dynamics, our study aims to develop cost-efficient and flexible solutions for train services leveraging different types of rolling stock, each with distinct cost, capacity, and physical dimensions. Our challenge is to minimize the operational costs while ensuring adequate capacity during peak hours without exceeding the platform constraints. We will study deterministic as well as stochastic and time-varying passenger demand and test the performance of various formulations against each other. The stochastic model will be extended to accommodate the reusing of trains and the introduction of two separate passenger classes.

In the context of optimizing train scheduling and rolling stock management, our study specifically focuses on minimizing operational costs. Previous research looks at similar rolling stock problems but often with different objectives. Alfieri et al. (2006) and Schrijver (1993) aim to reduce the number of train units and carriage kilometers and Peeters and Kroon (2008) maximize the service to the passengers. Similarly, Abbink et al. (2004) focus on minimizing seat shortages during rush hours with an integer linear programming (IP) formulation, directly informing our approach to balancing operational costs and passenger demand. While these studies tackle problems similar to ours, particularly in managing rolling stock efficiently, they do not centre on cost minimization as the primary goal. Our work fills this gap by concentrating on cost efficiency in response to fluctuating passenger demands, a challenge accentuated by recent changes in commuting patterns.

Our approach involves formulating the problem using IP, examining both a direct allocation model and a composition-based model as per the frameworks proposed by Abbink et al. (2004) and adapted from Fioole et al. (2006). We first use both formulations in a setting of deterministic demand equal for each day and then extend the framework by Fioole et al. (2006) to include stochastic and day-varying demand, reflecting real-world uncertainties in daily ridership. Furthermore, other work looks at reusing trains for the same lines, like (Fioole et al., 2006) and (Alfeiri et al., 2006). Therefore, we extend the model by using a new formulation approach, circulating the rolling stock within lines. Lastly, public debate in the Netherlands has arisen concerning allocating first- and second-class train seats (Verbeek & Larkens, 2023). We therefore include a last analysis that considers different seats and the demand for first- and second-class passengers.

The remainder of this paper is organized as follows. Section 2 outlines the context of the problem and the specifics of the rolling stock constraints. Section 3 describes the IP formulations used for the deterministic and stochastic models. Section 4 presents the computational results and discusses the insights gained from the numerical experiments. Finally, Section 5 summarizes our findings and concludes the paper with a discussion on potential future extensions and applications of this research.

2 Problem description

The fundamental objective of the 8RSAP is to determine an optimal allocation and scheduling of rolling stock to minimize the total operational costs while meeting the service requirements across a railway network. This entails allocating various types of train units to scheduled train services in a manner that accommodates expected passenger volumes, adheres to operational constraints, and optimizes the use of available resources.

A rolling stock unit in the 8RSAP can be any designated rolling stock with specific operational characteristics, including cost, capacity, and physical dimensions. The problem can involve various train unit types, each with distinct attributes. Each scheduled train service has fixed starting and ending times and operates between designated stations. The primary challenge is to ensure that each service is covered exactly once by an appropriate configuration of train units. Furthermore, each train service must be provided with sufficient rolling stock capacity to handle the expected number of passengers, ensuring that all passengers can be accommodated comfortably and safely, and the total length of the rolling stock assigned to any service must not exceed the length of the shortest platform on its route to avoid operational disruptions. The composition of the rolling stock fleet should maintain a balanced use of available train unit types, often requiring that the proportions of different types remain within specified limits to ensure flexibility and cost efficiency. The key is to minimize the total cost associated with the rolling stock utilization, which may include fixed and variable costs associated with each type of train unit.

For analytical tractability, several assumptions are typically made when formulating the 8RSAP. Firstly, it may be assumed that any additional rolling stock required can be made available instantly, ignoring the lead times usually associated with procuring new units. Secondly, initial models often treat passenger demand as deterministic, simplifying the optimization problem. This assumption can later be relaxed to incorporate stochastic elements reflecting real-world variability. Each train unit is initially allocated to only one train service during critical peak periods, precluding its use in multiple consecutive services within short time frames.

3 Methodology

We will now present the general notation and formulations used to obtain our results. The formulations will be adapted to our specific dataset, and the relevant parameters will be initialized in the results section. Once a parameter is introduced in a section, its meaning remains unchanged in subsequent sections unless explicitly specified otherwise.

For sections Section 3.1 and Section 3.2, we use the assumption of constant demand, while in the other sections we assume stochastic demand which can be represented by a randomly distributed variable.

3.1 Basic integer linear programming formulation

We first denote the most basic formulation of the rolling stock problem and afterwards denote necessary changes for the linear programming (LP) relaxation.

3.1.1 Basic formulation

For the formulation of the basic IP problem, we rely on the notation used by Abbink et al. (2004). Let U denote the set of all types of rolling stock units and let C denote the set of different cross-section trains.

Furthermore, we denote the fixed costs of a type of train unit as f_u, the length of a type of rolling stock unit as l_u, and the capacity of a type of train unit in terms of seats as s_u. We then denote the demand of a cross-section train as d_c and the length of the platforms as p_c. k is a parameter that denotes the maximum allowed relative difference between the most and least preferred rolling stock type.

We define the decision variable as $N_{u,c}$, where $N_{u,c}$ is equal to the number of train units of type u allocated to cross-section train c. The formulation of the problem then becomes:

$$
\begin{aligned}
&\min_{N} \sum_{u} \sum_{c} f_u N_{u,c} && && (1)\\
&\textbf{s.t.} \sum_{u} s_u N_{u,c} \geq d_c && \forall c \in C && (2)\\
&\sum_{u} l_u N_{u,c} \leq p_c && \forall c \in C && (3)\\
&\sum_{c} (N_{i,c} - N_{j,c}) \leq k \sum_{c} N_{j,c} && \forall i, j \in U, i \neq j && (4)\\
&N_{u,c} \in \mathbb{N} \cup \{0\} && \forall c \in C, \quad \forall u \in U && (5)
\end{aligned}
$$

where (1) models the objective to minimize the rollings stock costs. Constraints (2) specify that the demand in seats is satisfied for each cross-section train. Additionally, constraints (3) model that each cross-section train should not be longer than the smallest platform along their route. Furthermore, constraints (4) specify that the difference between the number of units of the most desired type and the number of units of the least desired rolling stock type cannot be more than k% of the number of units of the least desired type. Lastly, constraints (5) are required in order to exclude the possibility that the number of train units is not a natural number.

3.1.2 Linear programming relaxation

To compare the performance of the formulations proposed in Section 3.1 and Section 3.2, we briefly introduce the LP relaxations of these problems. For this formulation, we relax constraints (5) and get the following new constraints:

$$
N_{u,c} \geq 0 \quad \forall c \in C, \quad \forall u \in U \tag{6}
$$

3.2 Basic set covering formulation

We now introduce the set covering formulation. The main difference to the formulation presented in Section 3.1 is that in this formulation we use train compositions as main decision variables instead of the integer amount of rolling stock type units. The compositions can consists of multiple rolling stock units and different rolling stock unit types.

Let P denote the set of all possible compositions of rolling stock and let C be the set of different cross-section trains. We now denote the fixed costs per year of a composition by f_u. For the second formulation, we introduce a binary decision variable $X_{c,p}$, which is defined as

$$
X_{c,p} = \begin{cases} 1, & \text{if composition p is used for cross-section train c} \\ 0, & \text{otherwise} \end{cases} \tag{7}
$$

For the set covering formulation, we have come up with two different formulations. They will be described in detail and formalized in Section 3.2.1 and Section 3.2.2. The main difference is that the first formulation uses subsets of possible compositions for each cross-section so that length and capacity constraints do not have to be included in the formulation, while the second formulation constrains this in a pre-computed parameter matrix.

3.2.1 Formulation with subsets for every cross section

In this formulation, P_c is a set of possible compositions for cross-section c. Composition is only included in P_c if it satisfies the platform length constraint for cross-section c (i.e. is shorter than the minimum platform length encountered) and its capacity constraint (i.e. demand does not exceed the capacity). We first present the basic set covering formulation with different possible combination sets for every cross-section and then discuss its meaning.

$$\min_X \sum_c \sum_{p \in P_c} X_{c,p} f_p \tag{8}$$

$$\textbf{s.t.} \sum_{p \in P_c} X_{c,p} = 1 \qquad \forall c \in C \tag{9}$$

$$\sum_c \sum_{p \in P_c} X_{c,p}(t_{i,p} - t_{j,p}) \leq k \sum_c \sum_{p \in P_c} X_{c,p} t_{j,p} \qquad \forall i, j \in U, i \neq j \tag{10}$$

$$X_{c,p} \in \mathbb{B} \qquad \forall c \in C, \quad \forall p \in P_c \tag{11}$$

where (8) models the objective to minimize the rolling stock costs. Then constraints (9) ensure that every cross-section train has exactly one composition p. Furthermore, constraints (10) specify that the difference between the number of units of the most desired type and the number of units of the least desired rolling stock type cannot be more than k% of the number of units of the least desired type. Here, $t_{i,p}$ stands for number of train units of type i in composition p. Lastly, constraints (11) are required to ensure that $X_{c,p}$ is a binary variable.

3.2.2 Formulation with parameter matrix

We now discuss a second formulation for the set covering problem formulation. Reasons for this will be explored in the results section , where the performance of both formulations will be compared. This composition is almost equal to the one presented in Section 3.2.1, however, there are some important differences. First of all, p again refers to the whole set of possible compositions, not only the feasible ones for that particular cross section c. Furthermore, for the additional constraints (14), we introduce the parameter matrix $v_{c,p}$, which is defined as

$$v_{c,p} = \begin{cases} 1, & \text{if composition p fulfills the capacity and length constraint for cross-section train c} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

The new formulation with parameter matrix is presented below:

$$\min_X \sum_c \sum_p X_{c,p} f_p \tag{13}$$

$$\textbf{s.t.} \sum_p X_{c,p} v_{c,p} = 1 \qquad \forall c \in C \tag{14}$$

$$\sum_p X_{c,p} = 1 \qquad \forall c \in C \tag{15}$$

$$\sum_c \sum_p X_{c,p}(t_{i,p} - t_{j,p}) \leq k \sum_c \sum_p X_{c,p} t_{j,p} \qquad \forall i, j \in U, i \neq j \tag{16}$$

$$X_{c,p} \in \mathbb{B} \qquad \forall c \in C, \quad \forall p \in P \tag{17}$$

For explanation of the objective function and constraints (13), (15) - (17), please refer to section Section 3.2.1. The additional constraints (14) ensure that only feasible compositions in terms of length and capacity are chosen in each cross section.

3.2.3 Linear programming relaxation

We rewrite the constraints (17) from the parameter matrix formulation in the following way to relax the IP problem:

$$0 \leq X_{c,p} \leq 1 \quad \forall c \in C, \quad \forall p \in P \tag{18}$$

Constraint (11) from the subset version is relaxed in a similar fashion. All other constraints stay the same when we refer to the LP relaxations in subsequent sections.

3.3 Stochastic and time-varying demand

Previously, we restricted demand to be deterministic and required it to be fully satisfied for all cross-sectional trains. We now relax both of these assumptions and expand our model. Firstly, demand can now be stochastic, following a known continuous distribution. Secondly, there is now a target of overcrowded trains for the whole cross section of trains, and single targets per cross section have to be satisfied.

We are building the following formulation on the basic set covering formulation that has been introduced before. There are several reasons for this. First and foremost, the set covering formulation proves to perform better than the basic IP formulation. Second, the model is easier to implement, especially when dealing with the new stochastic constraints. Lastly, this model is more flexible in the sense that it allows us to compare the performance of multiple equivalent formulations in practice.

Following Section 3.2, we first introduce two formulations: one formulation where the set of possible compositions is pre-determined and restricted to the length requirement and a second one where this constraint is controlled for via a parameter matrix. Afterwards, we show how varying demands on different days can be modeled.

3.3.1 Stochastic formulation with subsets for every cross section

In this formulation, sets P_c are constructed only based on the length requirement. The parameter r represents the percentage of cross section trains where the number of passengers should be lower than the capacity of the train. In subsequent sections, we will refer to this as passenger satisfaction. The number of cross sections is given by m. w_c^{th} denotes the percentile of the passenger demand distribution for cross section c.

$$\min_X \sum_c \sum_{p \in P_c} X_{c,p} f_p \tag{19}$$

$$\textbf{s.t.} \sum_{p \in P_c} X_{c,p} = 1 \qquad \forall c \in C \tag{20}$$

$$\sum_c \sum_{p \in P_c} X_{c,p}(t_{i,p} - t_{j,p}) \leq k \sum_c \sum_p X_{c,p} t_{j,p} \qquad \forall i, j \in U, i \neq j \tag{21}$$

$$\sum_c \sum_{p \in P_c} X_{c,p} \mathbb{P}(d_c \leq s_p - 0.5) \geq rm \tag{22}$$

$$\sum_{p \in P_c} X_{c,p} s_p \geq \ w_c^{\text{th}} \qquad \forall c \in C \tag{23}$$

$$X_{c,p} \in \mathbb{B}, \qquad \forall c \in C, \quad \forall p \in P_c \tag{24}$$

Objective function (19) and constraints (20) - (24) are interpreted the same way as in Section 3.2.1. Constraint (22) guarantees that in proportion r of m cross-section trains the number of passengers should be lower than the train's capacity. 0.5 factor appears in the expression as a continuity correction, in order to ensure the strict inequality when expressing discrete distribution in terms of continuous one. Constraint (23) ensures that for each cross section, all passengers should have a seat on at least share w of the time, where w_c^{th} denotes the w_c^{th} percentile of the demand distribution for each cross section

3.3.2 Stochastic formulation with parameter matrix

We now present the formulation with parameter matrix, which has a similar discrepancy to the subset version as the models presented earlier:

$$
\begin{aligned}
&\min_X \sum_c \sum_p X_{c,p} f_p && && (25)\\
&\textbf{s.t.} \sum_p X_{c,p} = 1 && \forall c \in C && (26)\\
&\sum_p X_{c,p} l_{c,p} = 1 && \forall c \in C && (27)\\
&\sum_c \sum_p X_{c,p}(t_{i,p} - t_{j,p}) \leq k \sum_c \sum_p X_{c,p} t_{j,p} && \forall i, j \in U, i \neq j && (28)\\
&\sum_c \sum_p X_{c,p} \mathbb{P}(d_c \leq s_p - 0.5) \geq rm && && (29)\\
&\sum_p X_{c,p} s_p \geq \ w_c^{\text{th}} && \forall c \in C && (30)\\
&X_{c,p} \in \mathbb{B}, && \forall c, \quad \forall p \in P && (31)
\end{aligned}
$$

The difference between this and the previous formulation is that here we always draw p from set P of all possible combinations and account for the length restriction through constraints (27), where

$$
l_{c,p} = \begin{cases} 1, & \text{if composition p fulfills the length constraint for cross-section train c} \\ 0, & \text{otherwise} \end{cases} \tag{32}
$$

3.3.3 Time-varying demand over several days, allowing for different compositions on different days

In this subsection, we are going to consider demand that varies over time. To do this, consider set S, consisting of subsets of days with potential demands, such that each $s \in S$ denotes a subset of days exhibiting the same demand levels. For example, if S = { { Tue, Thu }, { Mon, Wed, Fri } }, then the demand on Tuesday and Thursday is the same and it differs from demand on Monday, Wednesday and Friday. Here we will also allow train compositions to vary from day to day, depending on the demand. To formulate this, we need to modify our decision variable a bit. Let

$$
X_{c,p,s} = \begin{cases} 1, & \text{if composition p is used for cross-section train c on day-type s} \\ 0, & \text{otherwise} \end{cases} \tag{33}
$$

Our goal is to minimize the number of overcrowded trains by changing compositions depending on demand, but keeping the number of train units, and therefore costs, fixed. The IP formulation with different compositions subsets for different cross sections of this problem is shown below:

$$
\begin{aligned}
&\min_X \sum_c \sum_{p \in P_c} \sum_s X_{c,p,s} \mathbb{P}(d_{c,s} \geq s_p) && && (34)\\
&\textbf{s.t.} \sum_{p \in P_c} X_{c,p,s} = 1 && \forall c \in C, \quad \forall s \in S && (35)\\
&\sum_c \sum_{p \in P_c} X_{c,p,s} t_{i,p} = t_i && \forall s \in S, \quad \forall i \in U && (36)\\
&X_{c,p,s} \in \mathbb{B} && \forall c \in C, \quad \forall p \in P_c, \quad \forall s \in S && (37)
\end{aligned}
$$

In this formulation, we again create different sets of possible compositions for each cross-section c beforehand, based on the platform length of train c. The objective function (34) models the minimization of overcrowded trains, and constraints (35) ensure that each train has exactly one composition on each type of day. Constraints (36) guarantees the fixed number t_i of train units of each type $i \in U$ on each kind of day $s \in S$. As previously, we have also formulated this problem using the same set of possible compositions P for each train c, but controlling the length with an additional constraint

$$\sum_{p} X_{c,p,s} l_{c,p} = 1 \qquad \forall c \in C, \quad \forall s \in S \tag{38}$$

The full formulation can be found in the Appendix. To see whether and by how much allowing for daily schedule changes decrease the number of overcrowded trains, we compare these models to basic models. These models keep the same composition each day, despite varying demand. Their formulations can also be found in the Appendix.

3.4 Reusing trains in multiple cross sections

In this subsection, to enhance cost efficiency and reflect real-world scenarios more closely, we assume that trains can be reused. Specifically, a train that travels from one station to another can return to the original station for reuse, provided that the time allows it. We group such cross-section trains $c \in C$ in the same group $q \in Q$, for which the same train is used. The decision variable for this problem is

$$X_{q,p} = \begin{cases} 1, & \text{if composition p is used for cross-section trains in group q} \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

The formulation of this a problem is very similar to the formulation from Section 3.3.1, the sets of possible compositions for each group q are determined beforehand. The formulation can be seen below:

$$\min_X \sum_q \sum_{p \in P_q} X_{q,p} f_p \tag{40}$$

$$\textbf{s.t.} \sum_{p \in P_q} X_{q,p} = 1 \qquad \forall q \in Q \tag{41}$$

$$\sum_c \sum_{p \in P_q} X_{q_c,p}(t_{i,p} - t_{j,p}) \leq k \sum_c \sum_p X_{q_c,p} t_{j,p} \qquad \forall i,j \in U, i \neq j \tag{42}$$

$$\sum_c \sum_{p \in P_q} X_{q_c,p} \mathbb{P}(d_c \leq s_p - 0.5) \geq rm \tag{43}$$

$$\sum_{p \in P_q} X_{q,p} s_p \geq \ w_c^{\text{th}} \qquad \forall c \in C \tag{44}$$

$$X_{q,p} \in \mathbb{B} \qquad \forall c \in C, \quad \forall p \in P_q \tag{45}$$

Here q_c refers to the group to which the cross-section train c belongs. The structure of these constraints and the objective function is similar to (19) to (24).

3.5 Introducing a two-class system

To reflect the real-world scenario more closely, we now impose the introduction of different passenger classes within a train. In practise, usually two classes exist, but our model is applicable to any integer amount of classes. We denote the different classes $z \in Z$. This model is constructed in the parameter version and closely resembles the basic parameter formulation of the stochastic demand model from Section 3.3.2.

$$\min_X \sum_c \sum_p X_{c,p} f_p \tag{46}$$

$$\textbf{s.t.} \sum_p X_{c,p} = 1 \qquad \forall c \in C \tag{47}$$

$$\sum_p X_{c,p} l_{c,p} = 1 \qquad \forall c \in C \tag{48}$$

$$\sum_c \sum_p X_{c,p}(t_{i,p} - t_{j,p}) \leq k \sum_c \sum_p X_{c,p} t_{j,p} \qquad \forall i, j \in U, i \neq j \tag{49}$$

$$\sum_c \sum_p X_{c,p} \mathbb{P}(d_{c,z} \leq s_{p,z} - 0.5) \geq rm \qquad \forall z \in Z \tag{50}$$

$$\sum_p X_{c,p} s_{p,z} \geq \ w_{c,z}^{\text{th}} \qquad \forall c \in C \quad \forall z \in Z \tag{51}$$

$$X_{c,p} \in \mathbb{B}, \qquad \forall c \in C, \quad \forall p \in P \tag{52}$$

The only difference compared to constraints (25) to (31) lies in constraints (50) and (51), which restrict the overall and cross-sectional satisfaction levels now per class instead of the passengers as a whole. Therefore, different customer satisfaction targets can be set depending on the class, which provides valuable insights for practitioners on the capacity per class.

4 Numerical results

In this section, we present the results of our numerical experiments. For each formulation presented in Section 4, we provide the results of running it on data from Erasmus University Rotterdam for 200 cross-section with normally distributed demand trains representing the 8RSAP (see Abbink et al., 2004). All experiments are run on an Apple M1 Pro 10 core 3.2 GHz computer with 16 GB of RAM. All optimization problems are implemented and solved in Python using the Gurobipy package.

Before presenting our results, we introduce the important properties of our dataset. Two types of rolling stock can be acquired. Their fixed costs per year, capacity and length are given in Table 1.

Table 1: Rolling stock types

Type	Fixed costs	No of seats	Length
OC	€260.000	620	100
OH	€210.000	420	70

The basic IP problem is concerned with finding the optimal integer amount of OC and OH rolling stock units for each of the 200 cross-sectional trains. For the basic set covering formulation, we pre-compute the possible compositions of OC and OH rolling stock units constraining it by the maximum platform length of 300 meters. The possible compositions are presented in the Table 2. For clarity, OC is abbreviated by C and OH by H in Table 2. The order of rolling stock units is assumed to be irrelevant. Table 2 presents possible compositions, fixed costs per year, the number of available seats, the length in meters, the number of OC and OH units.

Table 2: Composition information

Composition	Fixed costs	No of seats	Length	OC units	OH units
C	€260000	620	100	1	0
H	€210000	420	70	0	1
CC	€520000	1240	200	2	0
CH	€470000	1040	170	1	1
HH	€420000	840	140	0	2
CCC	€780000	1860	300	3	0
CCH	€730000	1660	270	2	1
CHH	€680000	1460	240	1	2
HHH	€630000	1260	210	0	3
HHHH	€840000	1680	280	0	4

We present the results regarding basic formulations with deterministic demand in Section 4.1, the results while considering stochastic demand in Section 4.2. The basic problem is strained with many unrealistic assumptions. In our extensions, we are relaxing some of these assumptions to make the model more realistic and applicable to real-world scenarios. In Section 4.3 we investigate the extension of reusing trains in multiple cross sections and in Section 4.4 we explore sensitivity analysis for a system with two passenger classes.

4.1 Basic formulations

We present the results of the optimization problems described by the formulations in Section 3.1 and Section 3.2 in Table 3. Table 3 contains information about the formulation used, the type of optimization used, the total costs, the number of OC and OH rolling stock units and the running time. The running time is split up into the formulation time, which includes setting up the constraints and computing possible subsets and parameter matrices. The optimization time is the time used to actually optimize the problem.

Table 3: Results for different basic formulations

Formulation	Type	Total costs	OC units	OH units	For time	Opt time
Basic	IP	€72,940,000	140.00	174.00	0.089	0.062
Basic	LP	€61,147,590	142.87	114.29	0.092	0.002
Set cov subsets	IP	€72,940,000	140.00	174.00	0.724	0.012
Set cov subsets	LP	€72,917,778	139.56	174.44	0.727	0.003
Set cov param	IP	€72.940.000	140.00	174.00	0.134	0.014
Set cov param	LP	€72,917,778	139.56	174.44	0.116	0.004

From Table 3 we see that the objective value for all three IP formulations are equal. This was to be expected as all formulations cover the same optimization problem.

The value of the LP relaxation of the basic problem is substantially lower than the optimal LP solutions of both set covering formulations. This indicates that the set covering formulations are more efficient as the lower bound given by the LP relaxation is closer to the integer solution. In particular, when looking at the results at cross-section level, only one cross-section is non-integer.

We found two interesting results when considering formulation and optimization times. First, we confirm our previous statement that the set covering problem is performing better as the optimization time is roughly one-fifth of that of the the basic formulation. This holds for both the parameter and subset version of the problem. Second, we find that the total time, formulation and optimization time together are substantially lower for the parameter version than for the subset version. This finding was a surprise, as it could be expected that pre-computing a subset of the allowed composition for each cross-section performs better than constraining this with 200 additional constraints. We will expand on this notion in Section 4.2.

4.2 Stochastic demand

Next, we examine the case involving stochastic demand, where demands follow a normal distribution with specified means and variances. In Table 4 we present the results of IP problems discussed in Section 3.3.1 and Section 3.3.2. Similarly to Table 3, it contains information about the formulation and type of optimization used, the total costs, the number of OC and OH rolling stock units, and the computation time. We notice that the number of required train units and, consequently, total costs have increased with stochastic demand despite the relaxation of the demand satisfaction constraint.

Table 4: Results for stochastic demand

Formulation	Type	Total costs	OC units	OH units	For time	Opt time
Set cov subsets	IP	€75,530,000	175	143	1.159	0.126
Set cov param	IP	€75,530,000	175	143	0.238	0.127

Regarding the running time, the results resemble those of Section 4.1, as version with different subsets for various cross-sections has a slightly lower optimization time, but significantly higher formulation time. One potential explanations for these differences in running times could be the practical intricacies in the implementation in Python, for example, the handling of sets of unequal size.

We now check how allowing for different compositions on different days for the time-varying demand model, discussed in Section 3.3.3 affects the amount of overcrowded trains. In Table 5 we present results. The second column describes the type of day by nature of demand: high, low and on average over all week days.

Table 5: Results for time-varying demand

Formulation	Overcrowdedness	Same comp	Diff comp	Reduction
Subsets	Mean over	38.272	33.241	15.758%
	High over	51.109	48.355	5.388%
	Low over	29.714	23.165	22.040%
Param	Mean over	38.272	33.241	15.758%
	High over	51.109	48.355	5.388%
	Low over	29.714	23.165	22.040%

We can observe that allowing for different compositions on such days reduces the number of overcrowded trains by 15.758% on average, compared to the basic case with the same daily compositions. The running times are presented in Table 6:

Table 6: Runtime comparison

Formulation	Time	Same Comp	Diff Comp
Subsets	For time	0.683	0.879
	Opt time	0.016	0.024
Param	For time	0.200	0.214
	Opt time	0.017	0.027

We can see that optimisation time is slightly better for the subsets formulation again but at the expense of a significantly higher formulation time.

4.3 Reusing trains

In this section, we are relaxing the assumption that each cross section needs to have separate rolling stock to be acquired. This is not very realistic, as in practise, there is no need to buy new rolling stock units for every cross section train. After consulting the given dataset, we propose to use certain rolling stock units for multiple cross sections. For this, we require that each train has at least five minutes idle time in the station before starting the next route. In our case, we only allow for a train to go back to the location where it came from, but a possible interesting extension would be to also relax that constraint.

Figure 1 shows the minimum costs for different customer satisfaction levels. The baseline stochastic model from Section 3.3 and our proposed extension are included.

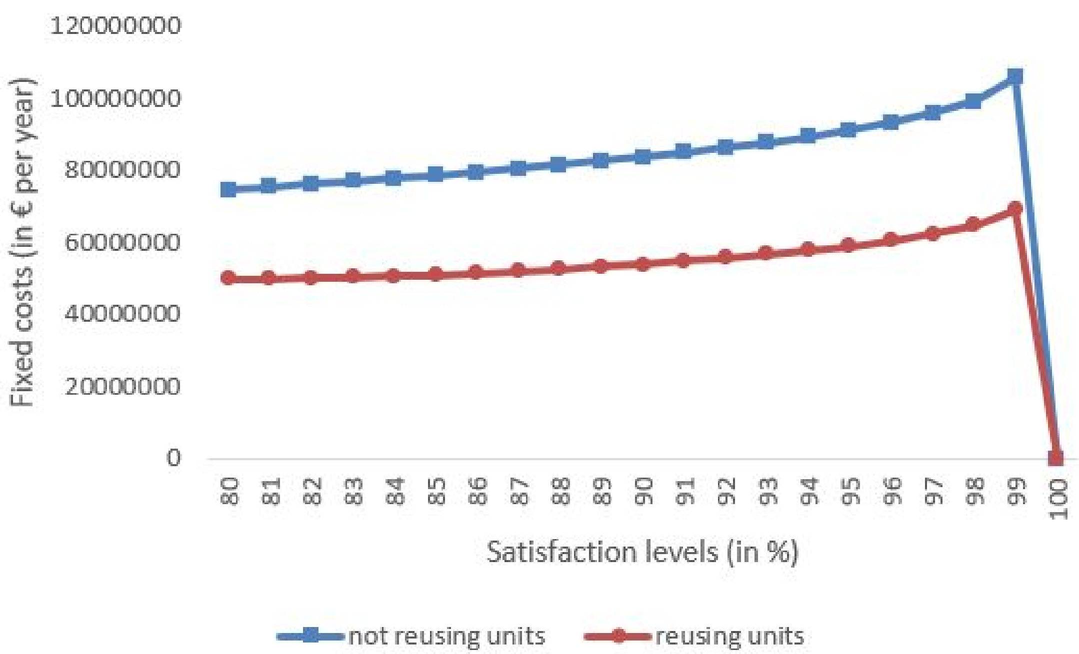

Figure 1: Performance comparison between basic model and extension

Figure 1 illustrates that the proposed extension to reuse trains significantly reduces costs, from €24,750,000 for a customer satisfaction level of 80% up to €36,570,000 for a customer satisfaction level of 99%. At a satisfaction level of 100%, both models are infeasible, which is represented by costs of 0 in the figure. This is due to the properties of the normal distribution.

4.4 Two-class system

We now study the case of two different classes within the train. This means there are two different sections within a train, where first- and second-class passengers sit separately. This extension greatly enhances its applicability in real-world scenarios and provides interesting dynamics to study and understand. We still impose some restricting assumptions on our model. Firstly, it is assumed that each rolling stock unit also includes a first-class section, and that first-class seats are equally distributed across rolling stock units in terms of the relative number of seats. Secondly, if the first-class section of a train is overcrowded, the train is counted as not satisfying first-class demand, even if these first-class passengers would still find a seat in the second-class section of the cross-section train.

As we are given only one normally distributed demand per cross section, we have to split up demand into first-class passenger demand and second-class passenger demand. After researching on the percentage of first-class passenger demand in the Netherlands, but also more generally in Europe, it is hard to pin it down to a certain percentage. We, therefore, assume that first-class demand accounts for 10% of total demand, implying that it is normally distributed with 10% of the mean of that cross section and 10% of the variance of that cross-section. However, we also conduct additional sensitivity analysis for different percentages in the Appendix. The remaining 90% are allocated to the second-class passenger demand. We assume that both resulting normal distributions are independent. It is well known that the sum of two independently distributed normal variables is also normally distributed, with mean and variance being equal to the sum of the means and variances of the separate normal distributions. Therefore, total demand is still in line with the numbers given in our dataset.

Figure 2 shows the minimum costs given that the satisfaction level for second-class demand must be at least 81% and first- and second-class demand must be satisfied for at least 125 out of 250 weekdays per year. We vary the satisfaction level for the first class and the first class capacity per cross-section train to find the optimal configuration for each satisfaction level. Infeasible data points are excluded from the figure for the sake of clarity.

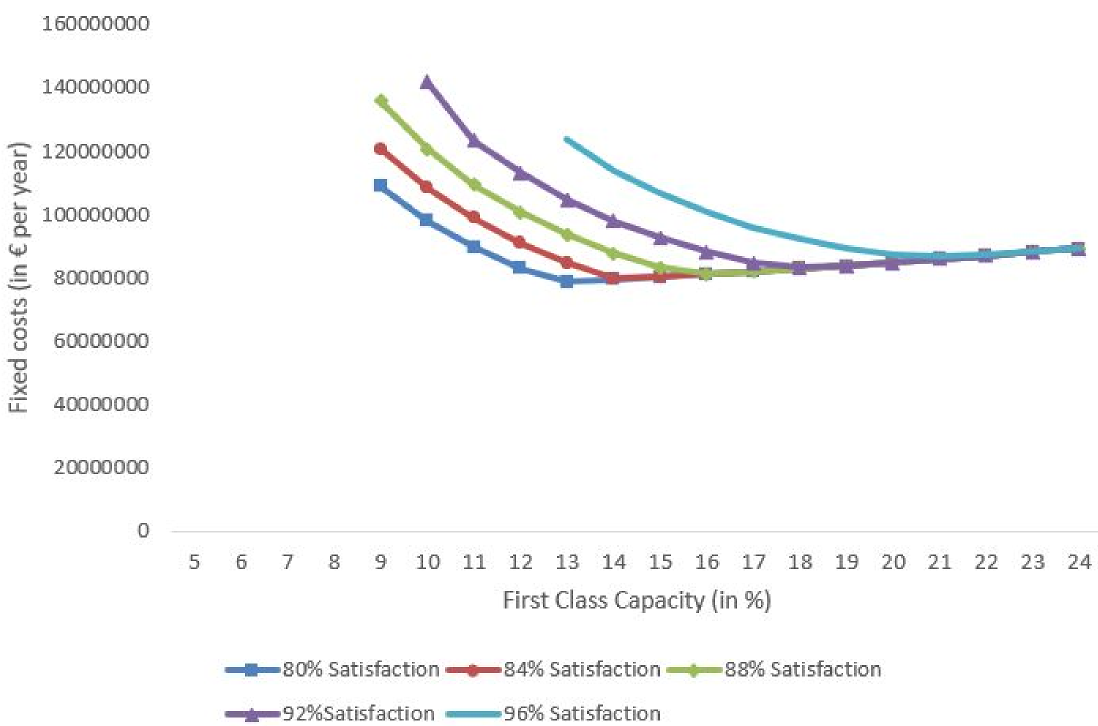

Figure 2: Sensitivity analysis for first class capacity and satisfaction levels at 10% first class demand

The shape of each satisfaction level line can be reconciled with a U-shape. The initial decrease as first-class capacity rises can be explained by the fact that fewer rolling stock units are required to satisfy first-class satisfaction constraints. As first-class capacity increases further, even though first-class satisfaction constraints are met, second-class seats are fewer and therefore, more rolling stock units are required to satisfy second-class demand. Given satisfaction constraints, this leaves us with an optimal point for first-class capacity. If first-class satisfaction level was set to 96%, for example, it would be optimal to let 21% of the capacity be the first class to satisfy all constraints. This would come with an increase in yearly fixed costs of €11,330,000 compared to the basic stochastic model.

The Appendix also includes more sensitivity analysis we conducted to check for 5% and 15% first class demand. The resulting picture has a very similar structure and is just shifted to the left and to the right, respectively. The flexibility of our model makes it easily applicable in real-world scenarios to determine the optimal distribution of seats between classes depending only on initializing parameters based on real-world data and objectives.

5 Conclusion

This paper tackles the 8 o'clock rolling stock assignment problem, focusing on the minimization of operational costs for train scheduling under fluctuating passenger demands, particularly due to the shifting work patterns post-pandemic. We explored two primary formulations: a basic integer linear programming model and a set covering model with two distinct approaches—one using subsets specific to each cross-section and another utilizing a predefined parameter matrix to manage constraints. We have also extended the basic 8RSAP problem enabling the reuse of rolling stock units across different cross-sections and incorporating multiple classes within the trains, which makes the models more applicable to real-world scenarios.

Our computational results indicate that the set covering formulation with a pre-computed constraining parameter matrix generally offers superior computational efficiency, managing faster solution times. This finding underscores its potential for practical application in real-world scenarios, where daily passenger volumes can vary significantly. Moreover, introducing stochastic elements to model these fluctuations revealed that accommodating uncertainty typically increases operational costs. We have also found that using different train compositions of days with different demands will help reduce the number of overcrowded train, highlighting the need for robust scheduling strategies that can adapt to daily and seasonal changes in passenger behaviors. Additionally, reusing trains in multiple cross sections greatly reduces costs and allows much higher customer satisfaction levels. Introducing two classes raises costs, but our model can be used to determine optimal first class capacity when first class demand and satisfaction levels are given.

It could be of interest for future research to apply these models to larger and more complex datasets could reveal insights into scalability and efficiency of our formulations. On the other hand, from a theoretical standpoint, considering larger datasets would allow to study the performance differences that subset and parameter matrix specifications yield in more depth. Since our two class extension provides a model to determine the optimal first class capacity given demand, future work could study the impact of offering two passenger classes on revenue and profit.

This study contributes to the ongoing discussion about optimizing public transport in response to evolving urban mobility patterns, providing a framework that could be applied to other public transport systems facing similar challenges worldwide.

6 References

Abbink, E., van den Berg, B., Kroon, L., & Salomon, M. (2004). Allocation of Railway Rolling Stock for Passenger Trains. *Transportation Science, 38*(1), 33-41. doi: https://doi.org/10.1287/trsc.1030.0044

Alfieri, A., Groot, R., Kroon, L., & Schrijver, A. (2006). Efficient Circulation of Railway Rolling Stock. *Transportation Science, 40*(3), 378–391. doi: https://doi.org/10.1287/trsc.1060.0155

Fioole, P.-J., Kroon, L., Maro´ti, G., & Schrijver, A. (2006). A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research, 174*, 1281–1297. doi: https://doi.org/10.1016/j.ejor.2005.03.032

Peeters, M., & Kroon, L. (2008). Circulation of railway rolling stock: A branch-and-price approach. *Computers & Operations Research, 35*(2), 538–556. doi: https://doi.org/10.1016/j.cor.2006.03.019

Verbeek, J., & Lalkens, P. (2023, October 23). Eerste klas in de trein moet ruimte inleveren, maar verdwijnt nog niet. *Het Financieel Dagblad*. Retrieved from: https://fd.nl/bedrijfsleven/1493688/eerste-klas-in-de-trein-verdwijnt-niet-maar-moet-wel-ruimte-inleveren

Schrijver, A. (1993). Minimum Circulation of Railway Stock. *CWI Quarterly, 6*, 205–217.

7 Appendix

7.1 Additional formulations for time-varying demand

7.1.1 Time-varying demand over several days with same daily compositions, formulation with different composition subsets for every cross section

$$\min_X \sum_c \sum_{p \in P_c} \sum_s X_{c,p} \mathbb{P}(d_{c,s} \geq s_p) \tag{53}$$

$$\textbf{s.t.} \sum_{p \in P_c} X_{c,p} = 1 \qquad \forall c \in C \tag{54}$$

$$\sum_c \sum_{p \in P_c} X_{c,p} t_{i,p} = t_i \qquad \forall i \in U \tag{55}$$

$$X_{c,p} \in \mathbb{B} \qquad \forall c \in C, \quad \forall p \in P_c \tag{56}$$

7.1.2 Time-varying demand over several days with same daily compositions, formulation with parameter matrix

$$\min_X \sum_c \sum_p \sum_s X_{c,p} \mathbb{P}(d_{c,s} \geq s_p) \tag{57}$$

$$\textbf{s.t.} \sum_p X_{c,p} l_{c,p} = 1 \qquad \forall c \in C \tag{58}$$

$$\sum_p X_{c,p} = 1 \qquad \forall c \in C \tag{59}$$

$$\sum_c \sum_p X_{c,p} t_{i,p} = t_i \qquad \forall i \in U \tag{60}$$

$$X_{c,p} \in \mathbb{B} \qquad \forall c \in C, \quad \forall p \in P \tag{61}$$

7.1.3 Time-varying demand over several days, allowing for different compositions on different days, formulation with parameter matrix

$$\min_X \sum_c \sum_p \sum_s X_{c,p,s} \mathbb{P}(d_{c,s} \geq s_p) \tag{62}$$

$$\textbf{s.t.} \sum_p X_{c,p,s} l_{c,p} = 1 \qquad \forall c \in C, \quad \forall s \in S \tag{63}$$

$$\sum_p X_{c,p,s} = 1 \qquad \forall c \in C, \quad \forall s \in S \tag{64}$$

$$\sum_c \sum_p X_{c,p,s} t_{i,p} = t_i \qquad \forall s \in S, \quad \forall i \in U \tag{65}$$

$$X_{c,p,s} \in \mathbb{B} \qquad \forall c \in C, \quad \forall p \in P, \quad \forall s \in S \tag{66}$$

7.2 Additional sensitivity analysis for a two class system

7.2.1 Sensitivity analysis assuming 5% first class demand

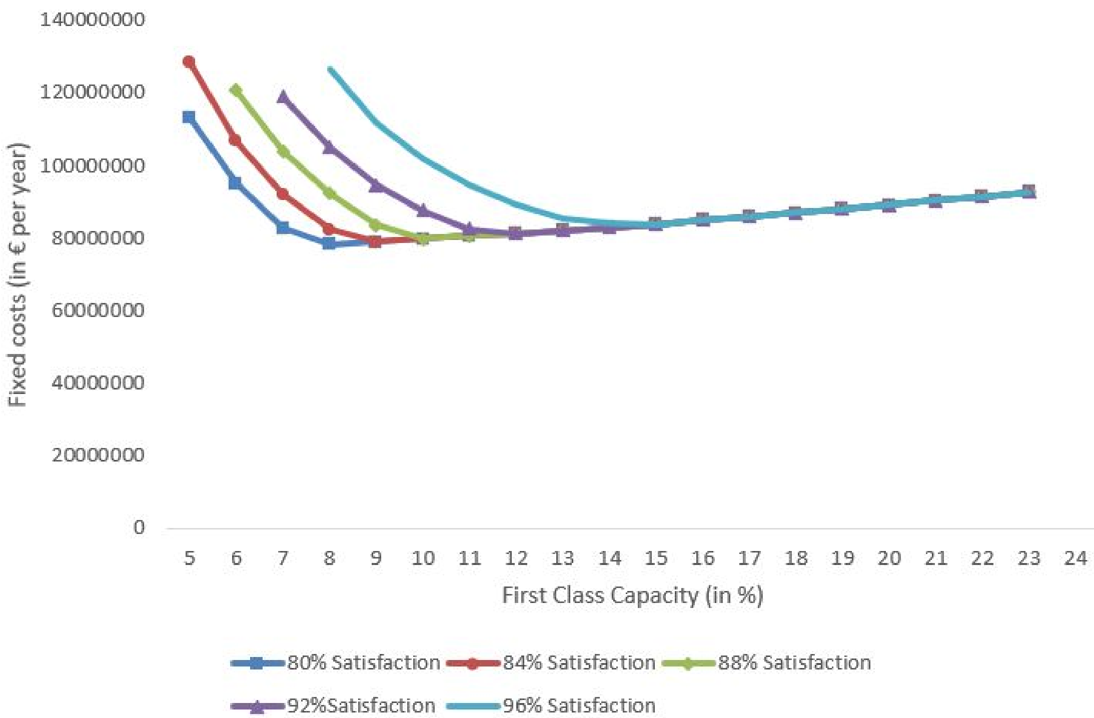

Figure 3: Sensitivity analysis for first class capacity and satisfaction levels at 5% first class demand

7.2.2 Sensitivity analysis assuming 15% first class demand

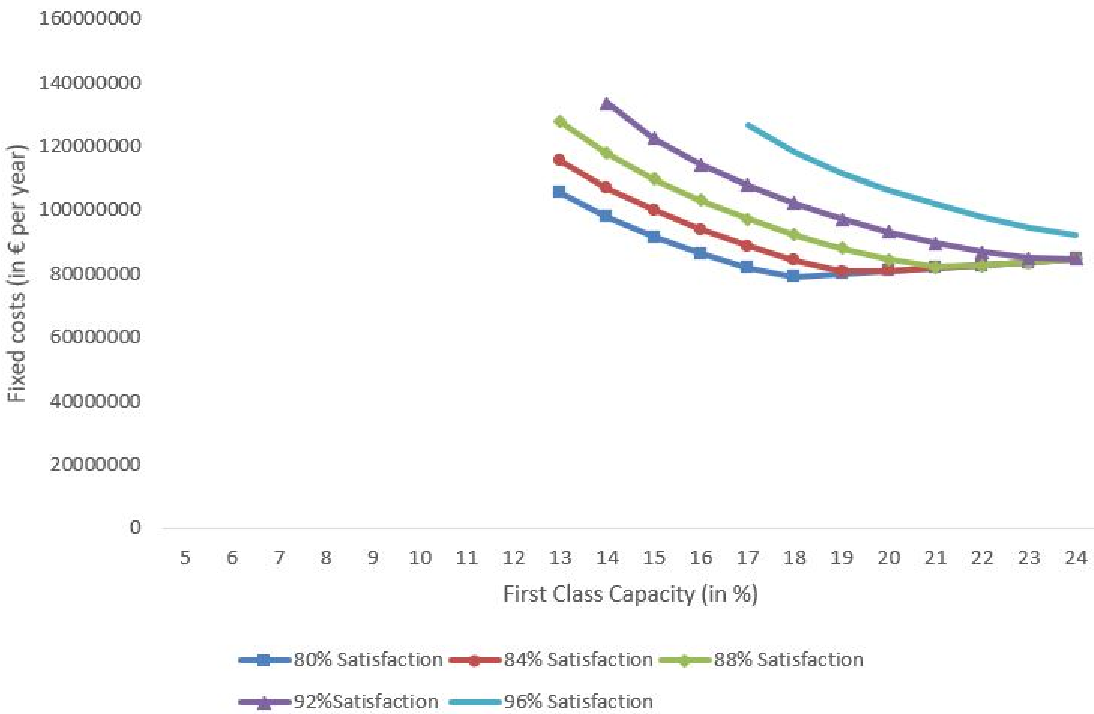

Figure 4: Sensitivity analysis for first class capacity and satisfaction levels at 15% first class demand