BSc Thesis Applied Mathematics

Cost optimising railway allocation in the Netherlands

Sanne Oude Veldhuis

Supervisor: N. van Dijk

June, 2021

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Preface

This article is made as a finalising Bachelors assignment for the bachelor Applied Mathematics at the University of Twente.
I want to thank Nico van Dijk for the help he gave me during our meetings when I was stuck and all the useful insight in how railway companies operate. I also want to thank him for reading my drafts, giving feedback and answering all my questions. Next to that, I want to thank my family and friends for listening to me complaining about my struggles to them in order to be able to figure out a solution.
Finally, I want to say that I really enjoyed doing this assignment and I am very happy with the way it turned out. I hope you as reader will also enjoy reading this article.

Abstract

The Nederlandse Spoorwegen exploits a lot of railway and transports more than a million passengers every day. It has a very high customer satisfaction rate that is possibly due to the low amount of transfers needed on average because of the long lines throughout the Netherlands. This is however not the least expensive way to allocate trains. In this article, a mathematical program is implemented to minimize the operational costs by deciding on the lines to use, with which frequency and with which amount of cars. First, a general model is constructed by making use of an example system. This model is then made into a linear model and subsequently expanded to decide the allocation for a part of the south-east of the Netherlands. For this part of the Netherlands, the model shows a large reduction of the costs and amount of empty seats in trains. It thus seems worthwhile for the NS to look into cost optimising their railway allocation.

Keywords: Railway allocation, Optimisation, Cost optimal

Contents

1 Introduction **3**

2 Restrictions and assumptions **4**
2.1 Dutch system 4
2.2 Model assumptions and data 4

3 Mathematical model **6**
3.1 Notation 6
3.2 Calculating the costs 7
3.3 General model 8
3.4 Linearisation 10
3.5 Implementation 10
3.5.1 Excel 10
3.5.2 AIMMS 11

4 Results **12**
4.1 Example system 12
4.2 South-east of the Netherlands 14

5 Discussion **18**
5.1 Differences between models 18
5.2 Comparing to NS allocation 18

6 Conclusions **20**

7 Recommendations **21**

8 References **22**

9 List of symbols **23**
9.1 Example system 23
9.2 Southeast of the Netherlands 24

A Appendices **27**

1 Introduction

The Nederlandse Spoorwegen is the largest railway company in the Netherlands. They have about 2100 kilometers of railway, more than a million passengers every day and almost 3000 trains with up to 260.000 seats [1]. To cater to all of these passengers, an allocation of all the trains has to be made. This is done while keeping a few important factors in mind, such as the number of transfers a passenger has to make, the chance that a passenger will have a seat in the train, the maximal capacity of the railway and some more. The current strategy of the NS mostly uses long lines throughout the Netherlands. This leads to fewer transfers for passengers in general and hence a higher customer satisfaction. When looking at efficiency however, the length of the train will be decided for the entire line, which will also lead to quite a high number of empty seats in parts of that line. The aim of this project is to look at cost efficiency, so to optimise the allocation of trains such that the costs are minimised while still catering to all passenger demands.
This is an important way to look at it, since the government is the only shareholder of NS, which means that a third of the money that will be saved can go back into the Dutch economy [2]. On the other hand, it will be interesting to see whether there is a lot of difference between the current method and the cost optimal method with regards to the length of lines and possibly passenger satisfaction. This will then give the NS another way to look at their current line allocation and possibly consider to change it such that it becomes more cost effective.

Operations research is a part of mathematics that has many uses in railway operations [3] [4]. It is used for planning the personnel on and around trains, the cleaning and maintenance of trains and of course for creating a timetable [5]. Next to that, it is also used for scheduling railway maintenance [6]. In all these situations, operations research is used to find an optimal solution for the problems and to be able to easily implement possible changes such as new trains or different demands or goals. In this article, operations research is also used to optimise the amount and length of trains for part of the Netherlands railway system.

In chapter 2, a general outline is given of all the assumptions that were made to create a working model and the restrictions regarding the Dutch railway network. Next, in chapter 3 is explained how the model is built up and how a smaller example was used to create the model, which was then expanded to fit every part of the railway network. The build up and explanation of the model can also be found in chapter 3, including the implementation in both Excel and AIMMS. The model is then implemented for the small example and for an existing part of the railway network of the Netherlands. These results can be found in chapter 4, with an explanation of differences between the different models as well as differences to the current allocation in chapter 5.

2 Restrictions and assumptions

When making a line allocation, there are some restrictions that have to be met in order to be able to implement the results in the real world. At the same time, in order to create a workable model, certain assumptions have to be made. The model has to be as realistic and true-to-life as possible. Unfortunately, when working with a mathematical model, it is not possible to create a model that will represent the situation 100% accurately. This is because every restriction or component that has to be added in, will increase the computations required to solve the program. The choices that were made can be found in the following subsections, with some elaboration on their consequences for the model.

2.1 Dutch system

There are some factors that are characteristic for the Dutch railway system and are given by the NS. These factors will be taken into account in this model, since a workable model has to be created and these are some conditions that the NS has on their line allocation.
The first condition to keep in mind is the fact that the NS uses a cyclic timetable. This means that the system repeats itself every hour during the day. This causes the fact that an optimisation for this system will not depend on the time. Therefore in this model, the line allocation will be done for one hour and it is assumed that it will repeat itself every hour.
As Peeters said [7], using a cyclic system will increase the customer satisfaction, as it is very clear for passengers when a train will depart. It is however also the case that creating this timetable can be very difficult. Since creating a timetable is not part of this report, this can be neglected and is possibly a subject for further research.

Another thing that is typical for the Dutch system is that the system is symmetric. This means that the same amount of passengers and trains will travel in each direction, causing the fact that is does not matter whether the model is taken in both directions or just one direction and then duplicated. As Liebchen found [8], including symmetry will speed up the optimisation process, while it can also create suboptimality in the results. This means that there are advantages as well as disadvantages to assuming the system is symmetric, but since the NS prefers to include symmetry, it will also be used in this report.

2.2 Model assumptions and data

One of the parts to keep in mind when making an allocation of trains is the number of passengers that want to ride each track. The number of passengers will change every hour, with certain peak hours in the morning and evening with home-work commute. In this model the passenger arrival is taken as a deterministic variable and is considered to be a constant amount. This variable is taken as a strict constraint, such that all passengers will be served.
In this case, a differentiation will be made between two different train types used, namely intercity (IC) and regional train (AR). These types will both have certain properties, such as passenger capacity, speed, costs and more. Furthermore, every train station will be able to accommodate the regional train, while the intercity will only stop at certain stations. If this model is to be expanded, another train type can easily be added, but in this case we use these two types.
Also, the costs are split up in two different parts. First, the costs that depend on the amount of kilometers travelled, such as the deprecation of trains and tracks, maintenance

and gas. Next to that, there are also time dependent costs, such as the salaries of the personnel, cleaning costs and energy costs.

There are also some smaller assumptions, that are listed below:

- No distinction is made between first and second class passengers
- The speed of the trains is taken as a constant speed, independent of where it is travelling
- The frequency of trains on a certain track has an upper bound in order to make sure the track does not exceed its capacity

For the small example system, certain input parameters, such as the amount of passengers or maximum frequency, were arbitrarily chosen in order to be able to work with the system. For the part of the Netherlands that is integrated, different sources were used to find the needed data. The length from one station to another is taken [9] and can be found in table 10. For the passenger data, the data was taken from the travellers dashboard of the NS [10], with corresponding percentages of passengers that were travelling during morning peak hours. This is done because on average, the largest percentage of passengers were travelling during these hours. Next to this input, the intercity stations are found on the website of the NS. The area in the south east of the Netherlands that is implemented can be seen in figure 1.

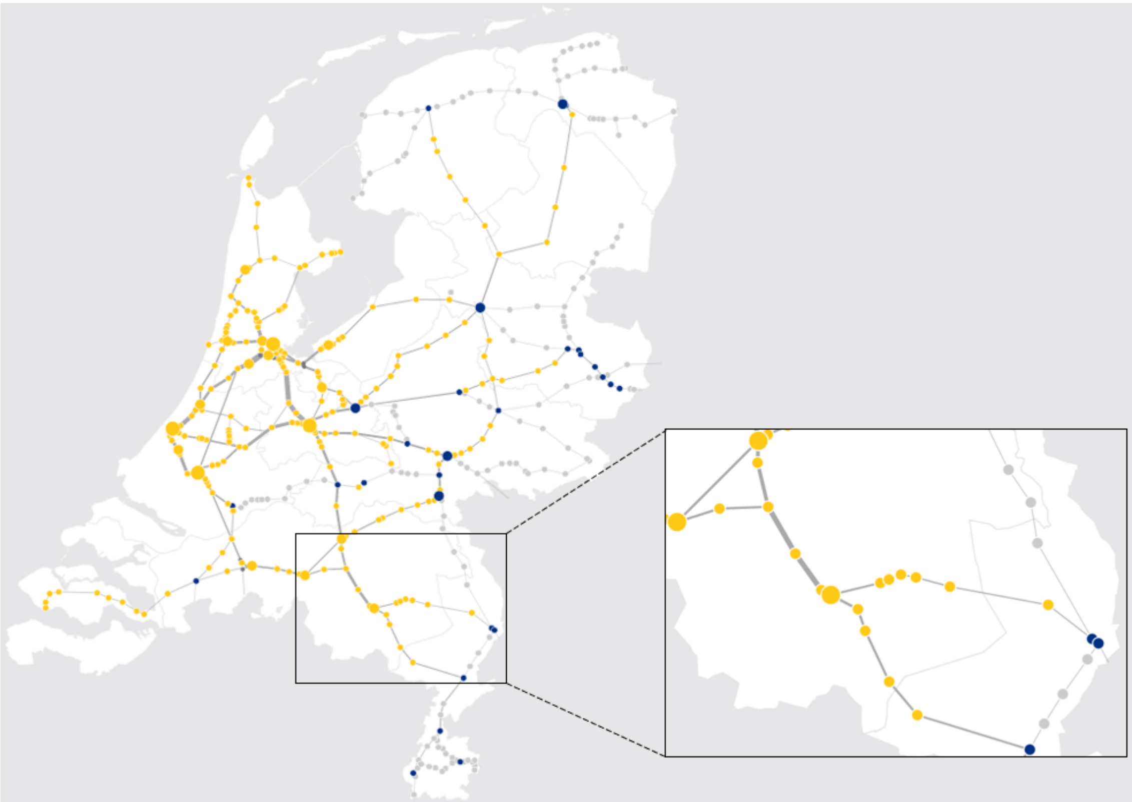

FIGURE 1: Part of the Netherlands that is implemented [10]

3 Mathematical model

In order to make a model that can be used to determine which lines to use and how long the trains have to be, it is useful to start with a small scale instructive example. In this case, an example railway network with five different stations is considered. A schematic depiction of this system can be seen in figure 2. The stations are named A to E and the tracks are numbered 1 to 5. These names will be used in the rest of the article. Also, station A, B and D are IC stations and C and E are AR stations. The number of passengers per track are also given in table 1.

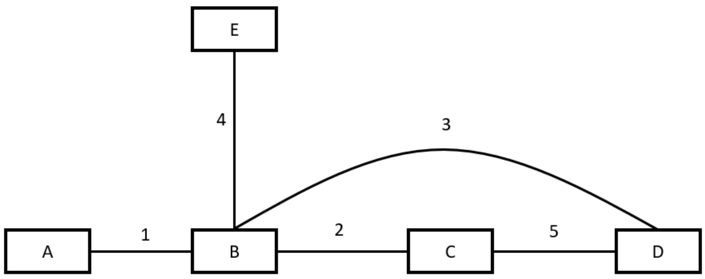

FIGURE 2: Example system

Track	Passengers
1	475
2	250
3	262
4	355
5	163

TABLE 1: Passengers per track

3.1 Notation

In the general problem, there are a few indices that are to be used:

i	start station	$i = 1, \ldots, n-1$
j	end station	$j = i+1, \ldots, n$
t	train type (IC, IR, AR)	$t = 1, 2, 3$
k	track	$k = 1, \ldots, m$

Next to these indices, there are also some input parameters to take into account:

r_k	track load for track k	$k = 1, \ldots, m$
β^t	car capacity for car of type t	$t = 1, 2, 3$
cf^t	costs per hour for car of type t	$t = 1, 2, 3$
ckm^t	costs per kilometer for car of type t	$t = 1, 2, 3$
$trkm^t$	costs per kilometer for train of type t	$t = 1, 2, 3$
I_{ij}	distance from station i to j	$i = 1, \ldots, n-1$ $j = i+1, \ldots, n$
f_k^{min}	minimal frequency for track k	$k = 1, \ldots, m$
$min(t)$	minimal number of cars for train of type t	$t = 1, 2, 3$
$max(t)$	maximal number of cars for train of type t	$t = 1, 2, 3$
s^t	speed for train of type t	$t = 1, 2, 3$
t^t	turnaround times for train of type t	$t = 1, 2, 3$

Finally, there are still some decision variables chosen over which we can optimise:

a_{ij}^t	number of cars of type t from station i to j	$t = 1, 2, 3$ $i = 1, \ldots, n-1$ $j = i+1, \ldots, n$
f_{ij}^t	frequency of train of type t from station i to j	$t = 1, 2, 3$ $i = 1, \ldots, n-1$ $j = i+1, \ldots, n$

3.2 Calculating the costs

When calculating the costs for an allocation, the costs for every line have to be calculated and then have to be all added up. So first, for example line AB of type 1 (IC). The number of cars a^1_{AB} has to be multiplied with the costs per car per kilometer. This is then multiplied by the belonging frequency f^1_{AB} and the amount of kilometers travelled I_{AB}. Next, we add the costs of the train per kilometer, this is also multiplied by the frequency and distance. Then, the time dependent costs have to be added to this. They are calculated by first calculating the time that the train takes to travel this line. This is done by taking the distance travelled, dividing it by the speed of the train and then adding this to the turnaround time in hours. Then, the travelling time is multiplied by the costs per hour, the amount of cars and the frequency. So the formula that is obtained for this is:

$$a^1_{AB} \cdot ckm^1 \cdot f^1_{AB} \cdot I_{AB} + trkm^1 \cdot f^1_{AB} \cdot I_{AB} + (\frac{I_{AB}}{s^1} + \frac{t^1}{60}) \cdot a^1_{AB} \cdot f^1_{AB} \cdot cf^1 \tag{1}$$

This can then be done for all possible combinations of begin and end station and train type. As an example, the allocation in table 2 is taken:

$$6 \cdot 1,26 \cdot 1 \cdot 150 + 2,535 \cdot 1 \cdot 150 + (\frac{150}{147} + \frac{15}{60}) \cdot 6 \cdot 1 \cdot 46 = 1864,88$$

$$1 \cdot 1,633 \cdot 1 \cdot 150 + 2,547 \cdot 1 \cdot 150 + (\frac{150}{111} + \frac{8}{60}) \cdot 1 \cdot 1 \cdot 50 = 701,23$$

$$4 \cdot 1,633 \cdot 1 \cdot 90 + 2,547 \cdot 1 \cdot 90 + (\frac{90}{111} + \frac{8}{60}) \cdot 4 \cdot 1 \cdot 50 = 1005,94$$

$$4 \cdot 1,633 \cdot 1 \cdot 140 + 2,547 \cdot 1 \cdot 140 + (\frac{140}{111} + \frac{8}{60}) \cdot 4 \cdot 1 \cdot 50 = 1549,98$$

Total costs: $1864,88 + 701,23 + 1005,94 + 1549,98 = 5122,03$

Type	Line	Frequency	Cars
IC	AD	1	6
AR	AD	1	1
AR	AE	1	4
AR	DE	1	4

Table 2: Allocation general solution

So in this case, the total costs are € 5122,03. In this situation, the other lines are also calculated, but since their frequency is zero, formula 1 becomes zero for these lines. All possible line combinations can be seen in matrix 2. So, to calculate the costs, formula 1 has to be filled in for all different possible lines. Then, all outcomes are added to each other in order to get the total costs.
It also becomes clear that the optimisation will determine the values of f^t_{ij} and a^t_{ij} for every possible combination of i, j and t.

To find out how many trains ride on each track, an incidence matrix is created. This matrix has the possible lines on the vertical axis and the possible tracks on the horizontal axis. When a track is used, the corresponding matrix entry reads 1, else it is zero. The matrix for this system is given below:

$$M_{ijt}^k = \begin{array}{c} \\ AB1 \\ AD1 \\ BD1 \\ AB3 \\ AC3 \\ AD3 \\ AE3 \\ BC3 \\ BD3 \\ BE3 \\ CD3 \\ CE3 \\ DE3 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array} \tag{2}$$

One of the constraints that has to be taken into account is the minimum amount of passengers that has to travel each track. When determining the amount of passengers on for example track 1, the incidence matrix is multiplied with the frequency, the corresponding amount of cars and the capacity of that train type. So for track 1 with the example combination AB1 it is calculated as follows:

$$f_{AB}^1 \cdot a_{AB}^1 \cdot \beta^1 \cdot M_{AB1}^1 \geq r_1 \tag{3}$$

Again, this can be calculated for the example allocation in table 2:
Track 1: $1 \cdot 6 \cdot 45 \cdot \mathbf{1} + 1 \cdot 1 \cdot 50 \cdot \mathbf{1} + 1 \cdot 4 \cdot 50 \cdot \mathbf{1} + 1 \cdot 4 \cdot 50 \cdot \mathbf{0} = 520 \geq 475$
Track 2: $1 \cdot 6 \cdot 45 \cdot \mathbf{0} + 1 \cdot 1 \cdot 50 \cdot \mathbf{1} + 1 \cdot 4 \cdot 50 \cdot \mathbf{0} + 1 \cdot 4 \cdot 50 \cdot \mathbf{1} = 250 \geq 250$
Track 3: $1 \cdot 6 \cdot 45 \cdot \mathbf{1} + 1 \cdot 1 \cdot 50 \cdot \mathbf{0} + 1 \cdot 4 \cdot 50 \cdot \mathbf{0} + 1 \cdot 4 \cdot 50 \cdot \mathbf{0} = 270 \geq 262$
Track 4: $1 \cdot 6 \cdot 45 \cdot \mathbf{0} + 1 \cdot 1 \cdot 50 \cdot \mathbf{0} + 1 \cdot 4 \cdot 50 \cdot \mathbf{1} + 1 \cdot 4 \cdot 50 \cdot \mathbf{1} = 400 \geq 355$
Track 5: $1 \cdot 6 \cdot 45 \cdot \mathbf{0} + 1 \cdot 1 \cdot 50 \cdot \mathbf{1} + 1 \cdot 4 \cdot 50 \cdot \mathbf{0} + 1 \cdot 4 \cdot 50 \cdot \mathbf{1} = 250 \geq 163$
As can be seen in the bold parts, vertically, the entries in matrix 2 can be seen for the corresponding lines that are chosen. Again, all other entries can also be calculated, but since their frequency is zero, it will not add anything to the outcome.

3.3 General model

As can be seen in formula 1, there are two main parts when determining the total costs for the trains, the distance dependent costs and the time dependent costs. The costs for the whole system are determined to exist of the costs for all lines added to each other. So in order to get a general expression for this, the sum over all possible lines has to be taken. This is done by summing over i, j and t. First, the distance dependent costs are given:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{t \in (1,3)} f_{ij}^t \cdot I_{ij}(a_{ij}^t \cdot ckm^t + trkm^t)$$

Here, the number of cars on a line (a_{ij}^t) is multiplied with their costs per kilometer (ckm^t) and then the costs per kilometer for the train ($trkm^t$) is added to that. Next, this is

multiplied by the number of kilometers the train travels (I_{ij}) and by its frequency of travel (f_{ij}^t).
Secondly, there are also the costs that depend on how long the train is in use:

$$\sum_{i=1}^{n}\sum_{j=i+1}^{n-1}\sum_{t\in(1,3)}(\frac{I_{ij}}{s^t}+\frac{t^t}{60})\cdot a_{ij}^t\cdot f_{ij}^t\cdot cf^t$$

For these costs, again the sum over all possible lines with different train types is taken. The amount of time that a train takes to complete the line is calculated by dividing the distance by the speed of the train ($\frac{I_{ij}}{s^t}$), the turnaround time in hours ($\frac{t^t}{60}$) is then added to this. This is then multiplied by the amount of cars on that line(a_{ij}^t), the frequency on that line (f_{ij}^t) and the hourly costs per car (cf^t). This is how we get the time dependent costs.

To get the total costs, the distance dependent costs have to be added to the time dependent costs. The total costs can then be minimised by changing the amount of cars per line and the frequency per line. There are also some constraints that have to be met. There is a minimal track load (r_k) that has to hold in order to be able to serve all travelers. In order to calculate the realised capacity per track, the incidence matrix 2 is multiplied by the capacity of a car (β^t), the amount of cars and the frequency. This is again summed over all start stations, end stations and train types. Another constraint is the minimal frequency on a certain track (f_k^{min}), to calculate the real frequency, the same matrix (M_{ijt}^k) is used and multiplied by the frequency. This is summed over the start stations, end stations and train types as well. Next to that, f_{ij}^t and a_{ij}^t have to be integer, non negative and not exceed their maximum.
The complete program formulation for this is as follows:

$$\min_{a_{ij}^t,f_{ij}^t}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t\in(1,3)} f_{ij}^t\cdot I_{ij}(a_{ij}^t\cdot ckm^t+trkm^t)+(\frac{I_{ij}}{s^t}+\frac{t^t}{60})\cdot a_{ij}^t\cdot f_{ij}^t\cdot cf^t$$

Subject to:

$$\begin{aligned}
&\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t\in(1,3)} f_{ij}^t\cdot a_{ij}^t\cdot\beta^t\cdot M_{ijt}^k\geq r_k && \forall k &&\\
&\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t\in(1,3)} f_{ij}^t\cdot M_{ijt}^k\geq f_k^{min} && \forall k &&\\
&0\leq f_{ij}^t\leq max(f) && \forall i,j>i,t && f_{ij}^t\ \text{integer}\\
&0\leq a_{ij}^t\leq max(t) && \forall t && a_{ij}^t\ \text{integer}
\end{aligned}\tag{4}$$

As can be seen from this program, it is not a linear program since the optimisation variables are f_{ij}^t and a_{ij}^t, which are multiplied with each other in the objective function, as well as in the constraints. This will lead to problems when using this program on a large train system, since non-linear programs are more complex and take more calculations than linear programs.

3.4 Linearisation

As said before, there are some advantages to having a linear model. For example, lower calculation requirements for the program that solves the problem. Another advantage is the fact that a linear model is more flexible, there are more possibilities to change it when the situation changes and a somewhat different model is required.

To transform the previous program into a linear program, one of the decision variables has to be integrated in another one or the formula has to be changed such that these variables are not multiplied with each other anymore. In this case, the first option is the simplest way to turn this program into a linear program. This is done by making a an index that runs between 0 and 15, as 15 is the maximum amount of cars, and then creating a new variable for the frequency dependent on this a, F^a_{ijt}. So:
$a := \{0, 1, ..., 15\}$
$F^a_{ijt} := \begin{cases} 0 & \text{if line } ij \text{ of type } t \text{ with } a \text{ cars not used} \\ x & \text{if line } ij \text{ of type } t \text{ with } a \text{ cars used x times} \end{cases}$

This new decision variable decides the frequency of a train of type t, length a, from i to j. This will be zero when this combination is not used and will have an integer value when it is used, which represents its frequency. The new linear program we get is as follows:

$$\min_{F^a_{ijt}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{t\in(1,3)} \sum_{a=1}^{15} F^a_{ijt} \cdot I_{ij}(a \cdot ckm^t + trkm^t) + (\frac{I_{ij}}{s^t} + \frac{t^t}{60}) \cdot a \cdot F^a_{ijt} \cdot cf^t$$

Subject to:

$$\begin{aligned} &\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{t\in(1,3)} \sum_{a=1}^{15} F^a_{ijt} \cdot a \cdot \beta^t \cdot M^k_{ijt} \geq r_k && \forall k \\ &\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{t\in(1,3)} \sum_{a=1}^{15} F^a_{ijt} \cdot M^k_{ijt} \geq f^{min}_k && \forall k \\ &0 \leq F^a_{ijt} \leq max(f) && \forall i, j > i, t, a. \end{aligned} \tag{5}$$

3.5 Implementation

Now that a model is made, it can be implemented in a solver software. This is an important step, as implementing the model will possibly make unfeasibility within the model clear or will give us results.

3.5.1 Excel

The first implementation is done in Excel. Excel has a solver add-on that can easily be installed and solve nonlinear as well as linear programs. For a small program, Excel is easy to use and it has a good overview of all of the variables. It does lack some computational power however, as the small example that includes five stations takes about an hour to compute using the linear model. It is expected that increasing the amount of stations and thus the amount of possible lines will lead to a much higher computation time. This is why further computation will not be done in Excel, but another program will be used.
There are also some advantages to Excel. Firstly, the implementation in Excel does not have a set layout. This means that the user can adapt and implement the layout in

whichever way they like. Another big advantage of Excel is that it can also be used as a calculation tool. When changing the input, the model automatically calculates the cost, making it a great tool to test certain inputs. For example when testing the costs of the current allocation or another suggested allocation.

3.5.2 AIMMS

Since Excel does not have enough computational power to solve a bigger problem within an hour, another solver was needed. AIMMS is solver software that has a higher computational power than Excel, so this was chosen to use. First it is used to solve the smaller example. This time, it only takes 0.16 seconds to solve the linear program, opposed to the hour that Excel takes. Since this looks promising, the larger model for part of the Netherlands is also implemented in AIMMS. For this larger model, it takes 4.85 seconds to solve, while the amount of variables is about fifteen times as much as the smaller model. The data screen for matrix M_{ijt}^k and all possible lines in AIMMS can be seen in appendix 8.

It is however a bit more difficult to implement the model in AIMMS. Where Excel is quite intuitive and free in setting up the program, AIMMS has a set layout in which the program can be built up. This makes it more difficult for beginners to set up a complicated program. When some experience is built up however, it turns out to be a straightforward application.

AIMMS does unfortunately not have the possibility to manually manipulate the decision variables, so can not be used as a calculation tool.

After all, Excel is easier to use for beginners and small programs, but when looking at computation time, it lacks quite a bit. With the problem of creating an allocation for part of the Netherlands in mind, AIMMS is a suitable program to solve these kinds of programs.

4 Results

4.1 Example system

For the example system, first a general solution was found by hand, this solutions is similar to how the current allocation by the NS is done. This is done in order to gain a starting point for the model, which can then be compared to the other solutions. The schematic depiction and exact allocation of the trains for the general solution can be found in figure 3 and table 3 respectively. Here, the passengers given in table 1 are taken.

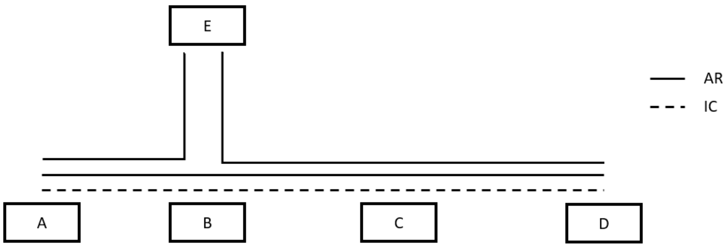

FIGURE 3: General solution

Type	Line	Frequency	Cars
IC	AD	1	6
AR	AD	1	1
AR	AE	1	4
AR	DE	1	4

TABLE 3: Allocation general solution

The total costs amount up to € 5122,04. For all results, it is the case that all the constraints hold and the solutions are feasible solutions.

Next, the results for the nonlinear program are given. This model took approximately 1-2 hours of calculation time. The resulting schematic and specified allocation can be found in figure 4 and table 4.

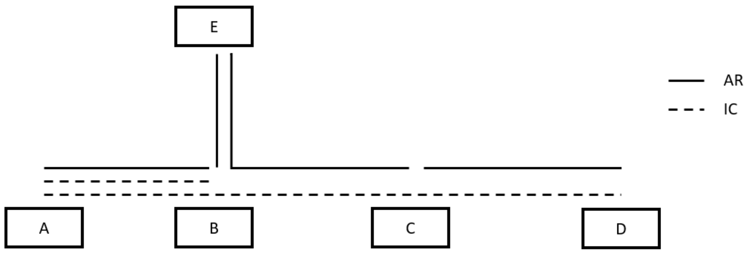

FIGURE 4: Solution nonlinear program

Here, the total costs are € 4753,96. It can be seen that in this case the amount of lines is higher than the general solution, but the lines are shorter.

Type	Line	Frequency	Cars
IC	AB	1	4
IC	AD	1	6
AR	AB	1	1
AR	BE	1	3
AR	CD	1	4
AR	CE	1	5

Table 4: Allocation nonlinear program

Finally for the example system, the results of the linear program are given below. This program is executed by both Excel and AIMMS.

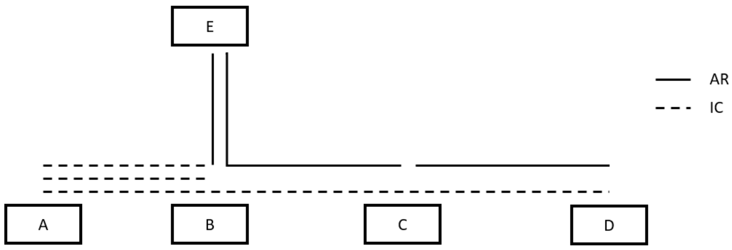

Figure 5: Solution linear program

Type	Line	Frequency	Cars
IC	AB	1	1
IC	AB	1	4
IC	AB	1	6
AR	BE	1	3
AR	CD	1	4
AR	CE	1	5

Table 5: Allocation linear program

Here, the total costs are € 4732.67. In this case, the result of AIMMS was a bit different, since it does not make a difference whether the first two entries for cars are 1 and 4 or 2 and 3. For both situations the total frequency and amount of cars for the IC on line AB are the same, so the costs will also remain the same.

For all three situations, the empty seats per track were also kept track of. An overview of this can be found below in table 6.

Track	General	Nonlinear	Linear
1	45	25	20
2	0	0	0
3	8	8	8
4	45	45	45
5	87	37	37
Total	185	115	110

Table 6: Empty seats per track

4.2 South-east of the Netherlands

For the south-east part of the Netherlands, all tracks that are part of the system have to be numbered in order to work with them. The numbering of the tracks can be seen in figure 6:

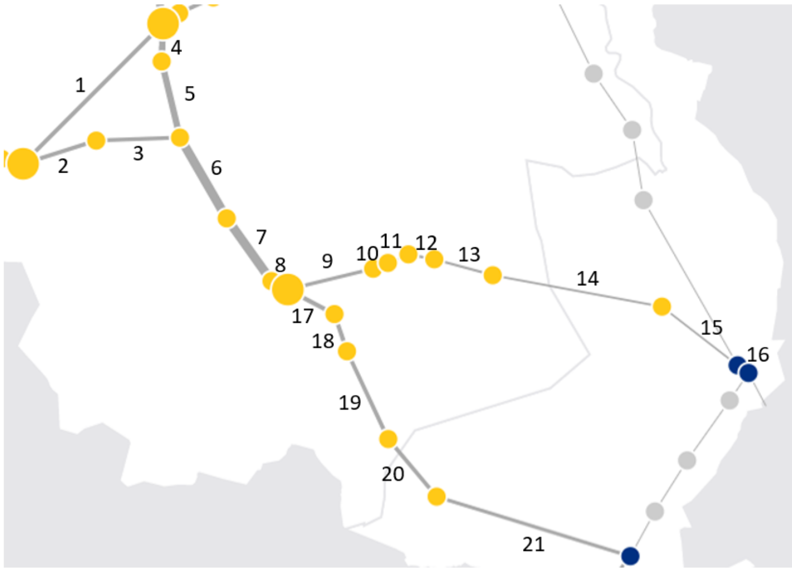

FIGURE 6: Numbering of the tracks

To be able to compare the results of the model, the current allocation of the NS is needed as a reference point. The schematic depiction of this system is given in figure 7 and the exact allocation can be found in table 7. Something to note here is the fact that some lines go on longer than their endpoints in this figure, for example the line from Tilburg to Eindhoven does not originate in Tilburg, but in Den Haag. Since this system does not include all stations in the Netherlands, these lines are taken to end at the station that is still included.

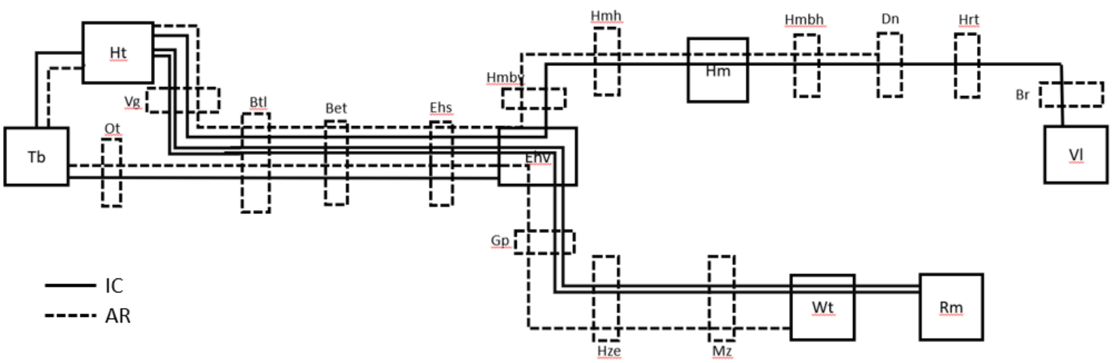

FIGURE 7: Allocation NS south-east Netherlands

This allocation is put into the model in Excel in order to calculate the costs and to get the amount of empty seats. In this case, the costs that the NS makes for this part of the Netherlands is € 19.281,56. The amount of empty seats adds up to 30.776. This is mostly

Type	Begin station	End station	Frequency	Cars
IC	Tilburg	's Hertogenbosch	2	6
IC	Tilburg	Eindhoven	2	9
IC	's Hertogenbosch	Venlo	2	10
IC	's Hertogenbosch	Roermond	2	4
IC	's Hertogenbosch	Roermond	2	8
AR	Tilburg	's Hertogenbosch	2	4
AR	Tilburg	Weert	2	4
AR	's Hertogenbosch	Deurne	2	6

Table 7: Allocation south-east Netherlands

due to the tracks between Boxtel (where Tilburg and 's Hertogenbosch come together) and Eindhoven. These tracks account for half of all empty seats in this system.

For the model, the data that is gathered is put into AIMMS in order to let AIMMS calculate the best allocation. An example of what this looks like in AIMMS can be seen in figure 8. Here, the left side depicts all of the input sets, parameters and variables and constraints. Next to this, part of the matrix M_{ijt}^k can be seen as well as all of the different possible lines.

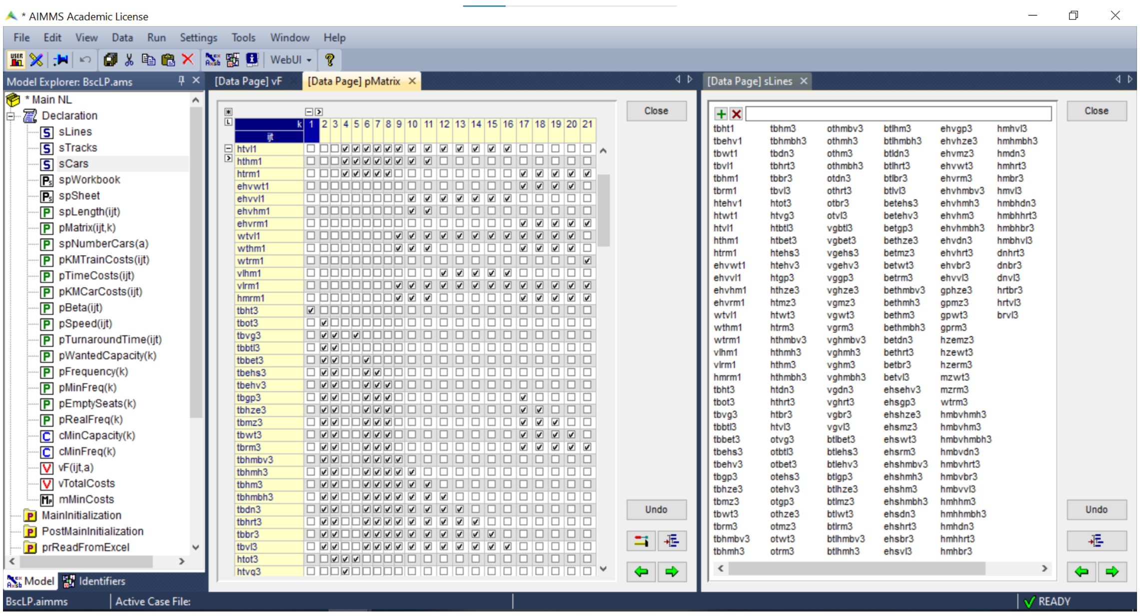

Figure 8: Input screen AIMMS

Then, for the solution of the model, the schematic depiction of the first solution is given in figure 9:

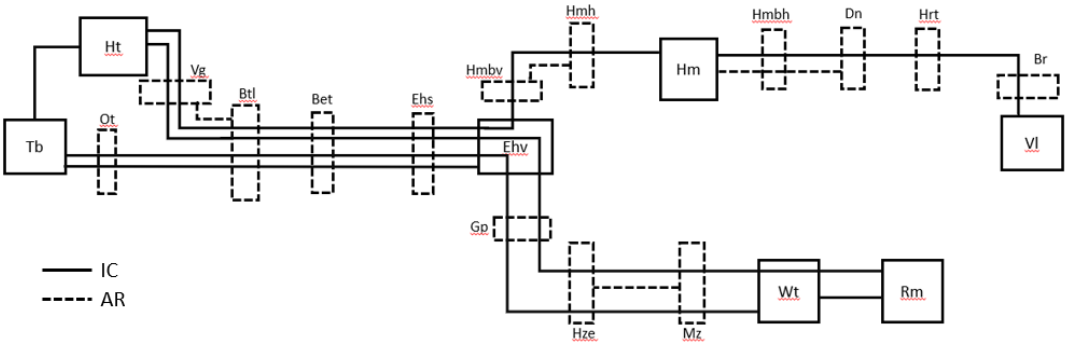

FIGURE 9: First solution model south-east Netherlands

Here, the costs are € 11.840,35 and the amount of empty seats is 6376. It becomes clear that in this case, there are a lot of IC trains. This is a possibility, but it will cause some stations to have no trains stop there, since they are not an IC station. This could be fixed the way NS does right now, by making the IC into an AR after a certain station. This is for example done between Helmond and Venlo in figure 7.
Another option is to make sure the model differentiates between IC and AR passengers. This is done by checking how many passengers travel from one IC station to another and creating a new virtual track for these passengers such that the model knows these trains will not stop at the other stations. The latter one is the method that was chosen for this model. The solution after this adaptation can be seen in figure 10 and the specifics of each line can be found in table 8.

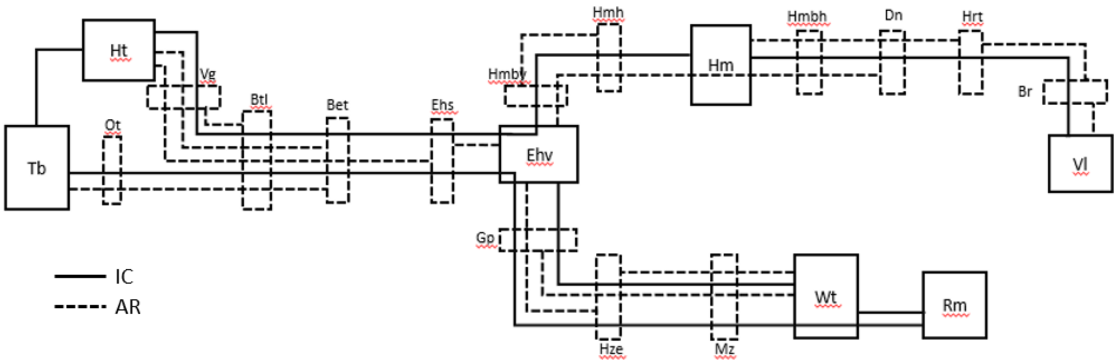

FIGURE 10: Solution model south-east Netherlands

Type	Begin station	End station	Frequency	Cars
IC	Tilburg	's Hertogenbosch	1	12
IC	Tilburg	Roermond	1	13
IC	's Hertogenbosch	Helmond	1	12
IC	Eindhoven	Weert	1	1
IC	Weert	Roermond	1	4
IC	Venlo	Helmond	1	3
AR	Tilburg	Best	1	8
AR	's Hertogenbosch	Best	1	12
AR	's Hertogenbosch	Eindhoven Strijp	1	9
AR	's Hertogenbosch	Eindhoven Strijp	1	15
AR	Vught	Boxtel	1	3
AR	Eindhoven Strijp	Eindhoven	1	1
AR	Eindhoven	Heeze	1	3
AR	Eindhoven	Deurne	1	5
AR	Geldrop	Weert	1	15
AR	Heeze	Weert	1	8
AR	Helmond Brandevoort	Helmond 't Hout	1	7
AR	Helmond	Horst-Sevenum	1	11
AR	Horst-Sevenum	Blerick	1	7
AR	Blerick	Venlo	1	1

Table 8: Allocation south-east Netherlands

For this final solution, the costs of the system are € 13.399,18. It is clear that the amount of AR trains has increased a lot compared to the previous solution. In this system, the amount of empty seats is only 616 seats. Table 9 is added below to compare all three given allocations.

Allocation	Costs (€)	Empty seats
NS	19.821,56	30.776
First solution	11.840,35	6.376
Final solution	13.399,18	616

Table 9: An overview of the allocations

5 Discussion

5.1 Differences between models

The first thing that stands out when looking at the results for the example system is that there is a big difference between the general model and the cost optimal models. This is the case in both the allocation and the height of the costs. In particular, it seems that shorter lines will lead to lower costs. As can be seen in table 6, the amount of empty seats per track decreases a lot when creating smaller lines. This is expected, as longer lines have to have enough seats to cater to all passengers on busy parts of that line, which often leads to seats leftover during other parts of that line. On the other hand, shorter lines lead to more time spent on turning the train around. That is the consideration that has to be made.

Another part that is interesting to see is the fact that the nonlinear and linear solution differ. Beforehand, it was expected that these outcomes would be the same and that only the calculation time would differ. The reason that it is different is because in the nonlinear model it is not possible for a line to have multiple values for the amount of cars. It is only possible to get an amount of cars and a corresponding frequency. In the linear model however, it is possible to have multiple values for the amount of cars. For every amount of cars, an entry for frequency is determined. That is where the difference between the linear and nonlinear model comes from.

5.2 Comparing to NS allocation

When comparing the solution from the model to the current allocation by NS, one of the things that stand out is the fact that the frequency of the NS trains is always 2, while in the model it is always 1. The NS does this to increase customer satisfaction, since a train once every 30 minutes is more convenient for passengers than a train once every hour. When looking at the costs however, it is cheaper to have one longer train than two smaller trains. This was not expected beforehand, but is not actually surprising as the constant costs of a train are counted twice when using two smaller trains instead of one long train. Another thing that catches the eye is the fact that in the first solution (figure 9) the majority of trains that is allocated are intercity trains, while there are a few small AR lines. This would mean that cost wise, having ICs is cheaper than having ARs in this system. However, as said in section 4.2, having many IC trains can have the consequence that some stations do not get visited. NS fixes this by letting an IC train also stop at AR stations after a certain point, but this was too difficult to implement in the model. This could also be considered to be less customer friendly, as they think that they travel with an intercity, while they do have to stop more often and thus have a longer travel time. That is why although the final solution (figure 10) has higher costs than the first solution, it is considered to be the wanted cost optimal allocation for this system.

For the smaller system it was the case that a lot of smaller lines were introduced in order to keep the empty seats at a minimum and thus having less costs. In this system, this is also partly the case, but mostly for the AR lines. It can be seen that there are a few short and a few long IC lines, which is already different from the NS allocation, but not as much as was expected after seeing the example system outcome. When looking at the AR lines, it is clear that there are more and smaller lines included than in the NS allocation. This was expected after seeing the results of the example system and will in this case also lead to a reduction of the amount of empty seats, as was expected.

However, when comparing the first allocation to the final one (table 9), the costs of the first solution are lower, while the amount of empty seats is higher. This implies that decreasing the number of empty seats does not necessarily lead to lower costs. This note is not entirely reliable as the first solution is considered to be an unrealistic solution. It does imply that having ICs is cheaper than having ARs, which is not really surprising because they are faster and thus have less time costs.

6 Conclusions

Altogether, it is clear that including smaller lines leads to less empty seats and thus to lower costs. This is expected as empty seats create unnecessary costs. This will be something for the NS to look into as it can significantly reduce their operating costs while for smaller stations it might not influence the amount of transfers that much.
Another aspect for the NS to consider implementing in some lines is changing the frequency from 2 to 1, since it will also lead to lower costs as it is cheaper to have one longer train than two shorter trains.

In conclusion, looking at the cost optimal approach for railway allocation is something that can be interesting for the NS. It is however not the perfect solution as it does increase the amount of transfers for passengers and can lead to some problems when implementing the allocation in a timetable. So it is a good idea to look at an allocation that is somewhere between the current allocation and the cost optimal allocation in order to reduce the costs and still have a high customer satisfaction rate.

7 Recommendations

Something that was mentioned in sections 4.2 and 5.2 is the fact that the NS sometimes lets an IC train stop at some AR stations when wanted. This would be something to look into in order to improve the model and make it even more realistic and applicable in the real world. Another thing that could be implemented in further research is making it possible for a train to split at a certain station and travelling further in two different directions. This could decrease the costs, while keeping the amount of transfers for passengers the same.

Another thing to keep in mind when doing further research is to cooperate with the NS in order to gain more specific passenger data and thus creating a more reliable situation. In this case, the model can be used, but the input data could be more specific and accurate. This is of course the case for all data, it is the data that the NS put online, but perhaps the NS has even better data to use.

8 References

References

[1] NS Utrecht. Ns jaarverslag 2019. pages 36–41, 2019.

[2] Ministerie van Financiën. Jaarverslag beheer staatsdeelnemingen 2019. page 97, 2019.

[3] Dennis Huisman, Leo G. Kroon, Ramon M. Lentink, and Michiel J. C. M. Vromans. Operations research in passenger railway transportation. *Statistica Neerlandica*, 59(4):467–497, 2005.

[4] Erwin Abbink et al. *Handbook of Optimization in the Railway Industry*, volume 268. 2018.

[5] Leo Kroon, Dennis Huisman, and Gábor Maróti. Railway timetabling from an operations research. *Report / Econometric Institute, Erasmus University Rotterdam*, (EI 2007-22), June 2007.

[6] G. Budai-Balke. *Operations research models for scheduling railway infrastructure maintenance*. Tinbergen Institute research series. Thela Thesis, 2009.

[7] Leon Peeters. *Cyclic Railway Timetable Optimization*. PhD thesis, E, June 2003.

[8] Christian Liebchen. Symmetry for periodic railway timetables. *Electronic Notes in Theoretical Computer Science*, 92:34–51, 2004. Proceedings of ATMOS Workshop 2003.

[9] Stationsweb, www.stationsweb.nl, 2020.

[10] Reizigersgedrag 2019, dashboards.nsjaarverslag.nl/reizigersgedrag, 2019.

9 List of symbols

Symbol	Description	Indices
i	start station	$i = 1, \dots, n-1$
j	end station	$j = i+1, \dots, n$
t	traintype (IC, IR, AR)	$t = 1, 2, 3$
k	track	$k = 1, \dots, m$
r_k	trackload for track k	$k = 1, \dots, m$
β^t	car capacity for car of type t	$t = 1, 2, 3$
cf^t	costs per hour for car of type t	$t = 1, 2, 3$
ckm^t	costs per kilometer for car of type t	$t = 1, 2, 3$
$trkm^t$	costs per kilometer for train of type t	$t = 1, 2, 3$
I_{ij}	distance from station i to j	$i = 1, \dots, n-1$ $j = i+1, \dots, n$
f_k^{min}	minimal frequency for track k	$k = 1, \dots, m$
$min(t)$	minimal number of cars for train of type t	$t = 1, 2, 3$
$max(t)$	maximal number of cars for train of type t	$t = 1, 2, 3$
s^t	speed for train of type t	$t = 1, 2, 3$
t^t	turnaround times for train of type t	$t = 1, 2, 3$
a_{ij}^t	number of cars of type t from station i to j	$t = 1, 2, 3$ $i = 1, \dots, n-1$ $j = i+1, \dots, n$
f_{ij}^t	frequency of train of type t from station i to j	$t = 1, 2, 3$ $i = 1, \dots, n-1$ $j = i+1, \dots, n$
F_{ijt}^a	frequency of train of type t with a cars from station i to j	$t = 1, 2, 3$ $a = 1, \dots, 15$ $i = 1, \dots, n-1$ $j = i+1, \dots, n$

9.1 Example system

Length (I_{ij}) in km:

	A	B	C	D	E
A	-	50	110	150	90
B		-	60	100	40
C			-	40	100
D				-	140
E					-

Symbol	Unit	Value
r_1		475
r_2		250
r_3		262
r_4		355
r_5		163
β^1		45
β^3		50
cf^1	€	46
cf^3	€	50
ckm^1	€	1,26
ckm^3	€	1,633
$trkm^1$	€	2,535
$trkm^3$	€	2,547
f_1^{min}		3
f_2^{min}		1
f_3^{min}		1
f_4^{min}		2
f_5^{min}		1
$max(t)$		15
s^1	km/hour	147
s^3	km/hour	111
t^1	Minutes	15
t^3	Minutes	8

9.2 Southeast of the Netherlands

The length (I_{ij}) in km can be found in table 10 and the numbers of the tracks can be seen in figure 6.

Symbol	Unit	Value
r_1		1188
r_2		1690
r_3		1688
r_4		2875
r_5		3150
r_6		4675
r_7		3655
r_8		2040
r_9		1339
r_{10}		1675
r_{11}		1260
r_{12}		998
r_{13}		1050
r_{14}		800
r_{15}		563
r_{16}		285
r_{17}		1540
r_{18}		2275
r_{19}		2503
r_{20}		2500
r_{21}		1675
β^1		100
β^3		50
cf^1	€	46
cf^3	€	50
ckm^1	€	2,52
ckm^3	€	1,633
$trkm^1$	€	2,535
$trkm^3$	€	2,547
f_k^{min}		1
$max(t)$		15
s^1	km/hour	147
s^3	km/hour	111
t^1	Minutes	15
t^3	Minutes	8

-	Vl	Br	Hrt	Dn	Hmbh	Hm	Hmh	Hmbv	Ehv	Gp	Hze	Mz	Wt	Rm	Ehs	Bet	Btl	Vg	Ht	Tb	Ot
Vl	-	1,35	11,35	29,33	35,35	38,36	40,55	43,00	51,57	57,70	61,87	71,57	81,57	105,73	53,58	61,65	70,49	78,71	82,65	88,34	80,29
Br		-	10,00	27,98	34,00	37,01	39,20	41,65	50,23	56,35	60,53	70,23	80,23	104,38	52,24	60,30	69,14	77,36	81,30	86,99	78,94
Hrt			-	17,98	24,00	27,01	29,20	31,65	40,23	46,35	50,53	60,23	70,23	94,38	42,24	50,30	59,14	67,36	71,30	76,99	68,94
Dn				-	6,02	9,03	11,22	13,67	22,25	28,37	32,55	42,25	52,25	76,40	24,26	32,32	41,16	49,39	53,32	59,01	50,96
Hmbh					-	3,01	5,20	7,65	16,22	22,35	26,52	36,22	46,22	70,38	18,23	26,30	35,14	43,36	47,30	52,99	44,94
Hm						-	2,19	4,64	13,21	19,34	23,51	33,21	43,21	67,37	15,22	23,29	32,13	40,35	44,29	49,98	41,93
Hmh							-	2,46	11,03	17,15	21,33	31,03	41,03	65,19	13,04	21,10	29,94	38,17	42,10	47,79	39,74
Hmbv								-	8,57	14,70	18,87	28,57	38,57	62,73	10,58	18,65	27,49	35,71	39,65	45,34	37,29
Ehv									-	6,13	10,30	20,00	30,00	54,16	2,01	10,08	18,91	27,14	31,07	36,77	28,71
Gp										-	4,17	13,87	23,87	48,03	8,14	16,20	25,04	33,27	37,20	42,89	34,84
Hze											-	9,70	19,70	43,86	12,31	20,38	29,21	37,44	41,37	47,07	39,01
Mz												-	10,00	34,16	22,01	30,08	38,91	47,14	51,07	56,77	48,71
Wt													-	24,16	32,01	40,08	48,91	57,14	61,07	66,77	58,71
Rm														-	56,17	64,23	73,07	81,30	85,23	90,92	82,87
Ehs															-	8,07	16,90	25,13	29,06	34,76	26,70
Bet																-	8,84	17,06	21,00	26,69	18,64
Btl																	-	8,23	12,16	16,98	8,93
Vg																		-	3,94	25,21	17,16
Ht																			-	22,43	21,09
Tb																				-	8,05
Ot																					-

Table 10: Length between stations in km

A Appendices

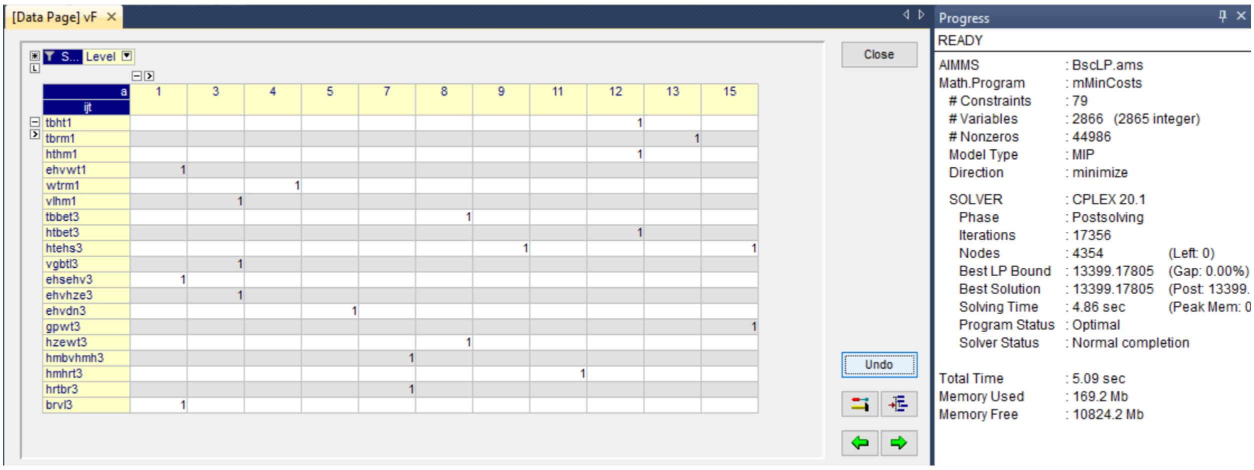

Figure 11: Output screen AIMMS

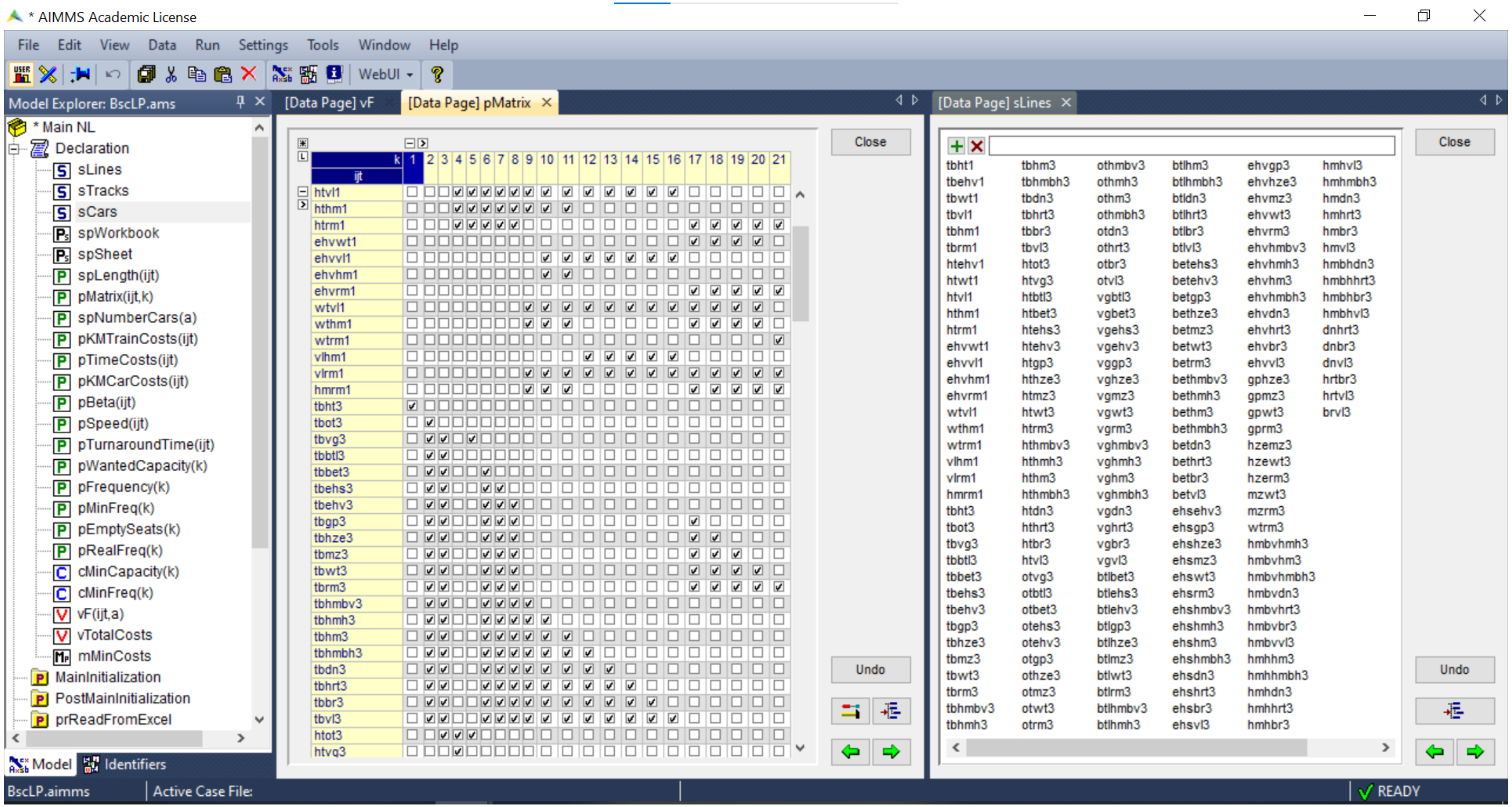

FIGURE 12: Input screen AIMMS