Examining Manual Solutions of the Train Unit Shunting Problem to find Train Type Patterns

Mitchell van Pelt[1]

Supervisors: Prof. dr. Mathijs de Weerdt[1]**, ir. Issa Hanou**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Mitchell van Pelt
Final project course: CSE3000 Research Project
Thesis committee: Prof. dr. Mathijs de Weerdt, ir. Issa Hanou, dr. Jing Sun

Abstract

This paper analyses manually realised solutions to the Train Unit Shunting Problem (TUSP) to find patterns in train type. The parking element is most important for the TUSP. Therefore, this research specifically investigates the presence of train type patterns in parking track and parking time. The difference in the patterns between main train type and train subtype is also analysed. The study uses statistical hypothesis testing to look for biases between individual train types and parking tracks. Kernel density estimation is used to analyse the differences in parking time between the types. The results show that there are strong patterns in type and parking track, but no clear difference in parking time. Considering subtype results in the differences being more specific. It is suspected that the strongly present track pattern is a strategy used by human planners.

1 Introduction

Railway companies base the size of their rolling stock on demand during peak hours. Outside of peak hours and during the night, demand is much lower and fewer trains are used. Trains that are temporarily not in service must be parked somewhere and optionally cleaned or serviced. This happens at so-called shunting yards, locations close to train stations with a large set of tracks. The Train Unit Shunting Problem (TUSP) concerns the challenge of deciding where to park each unused train unit, how to route each unit through the shunting yard and matching multiple units to form a new train that is needed [1].

In the Netherlands, shunting yards are usually located close to major stations, which means they are often enclosed within the city centre with little room for expansion. Because of growing passenger numbers, more trains will be needed in the future, which only makes the TUSP more complex than it already is [2]. To address this issue, Nederlandse Spoorwegen (NS, the Dutch national railway operator) is aiming to automate the planning process by 2028. Although algorithms exist, finding an optimal solution for this problem is computationally infeasible or yields an incomplete result, thus planning is still done manually by experienced planners [3], [4].

To address the flaws of these traditional algorithms, Van de Gevel [2] has suggested improving the problem's feasibility by incorporating concepts (patterns) that are understood by human planners into the algorithms. An example of such a concept used by van de Gevel is the knowledge that a block of train units arriving in a shunting yard is already in the correct order for departure. Hanou et al. [4] propose utilising artificial intelligence methods to help find good solutions to the TUSP and discover new insights that could be used to aid human operators in their planning process.

Human-like insight or patterns in TUSP solutions can thus be very helpful in making complex variants of the problem feasible to solve. However, not much is known about these patterns. While there has been research into how patterns can be used, no formal research has been conducted to find patterns in certain aspects of the shunting problem.

One such aspect is train type. This is a prominent characteristic of modern train units. At first glance, train type may seem easy to define. However, there is one important choice that can be made in defining it. These days, NS mainly operates Multiple Unit (MU) trains. These are train units consisting of multiple permanently joined self-propelled carriages [5]. In their naming system, NS defines the type of each MU on two levels: each unit has a type indicating the name of the series (e.g. ICMm), but can also be listed as the series combined with the number of carriages that the MU consists of (e.g. ICMm-IV). In this paper the first is referred to as the 'type' and the latter as the 'subtype'.

The goal of this study is to address part of the research gap by finding patterns of train types in solutions to the Train Unit Shunting Problem realised by manual planning. To achieve this, the following main research question will be answered: **What patterns of train type can be found in realised solutions of the Train Unit Shunting Problem?**

By answering this question, this study hopes that the knowledge of patterns in manual solutions can be used to improve the feasibility of existing algorithms. Insight into how train type is used by planners in existing solutions can help simplify the problem and construct artificial intelligence systems that can find good TUSP solutions.

Finding patterns is a very broad objective. If one considers the whole TUSP, there might be dozens of patterns that can be looked into. Since no formal work has been done on finding these patterns, this research will focus on the patterns that play the biggest role in the TUSP. According to Kroon et al. [6], the TUSP in its core consists of two main tasks: matching and parking train units. Since trains operated by NS always consist of units of the same type, it is already clear that train type is the key factor in the matching problem and that there is thus no need to look for patterns here. Instead, this research focuses on finding train type patterns in the parking aspect.

In a previous collaboration with NS, Beerthuizen [7] has developed a heuristic that involves parking train units of the same type on the same track. This strategy is in turn based on shipping container stacking [8]. If the heuristic has turned out useful for manual planners, it should be a pattern present in realised solutions. Furthermore, next to the track a unit is parked at, the time it is parked could be a very useful heuristic for parking units of different types on one track. Many shunting yards consist of only Last In First Out (LIFO) tracks. Units that stay parked for longer should be parked further down the track and vice versa. Lastly, the exact definition of a train unit's type could play a role in the patterns. The two hypothesised patterns are likely different when considering train unit subtype instead of just type. In summary, this all leads to the following three subquestions:

- Can a pattern of train unit types be found in the track where the units are parked in a shunting yard?
- Can a pattern of train unit types be found in the duration the units are parked in the shunting yard?
- Are the patterns different when unit subtype is considered?

The research questions are answered by investigating data of realised TUSP solutions constructed by manual planners. The utilised dataset consists of real train movements in a number of shunting yards over a period of ten months. The first two subquestions are answered by conducting two individual experiments with the data. The third subquestion is answered by performing each experiment twice: once for train type and once for train subtype.

Chapter 2 discusses relevant background in the methods used for the experiments. Then, Chapter 3 explains data preparation and the setup for the experiments. Chapters 4 and 5 discuss the experiments for the first and second subquestions respectively. Chapter 6 analyses the difference in results when subtype is considered. Chapter 7 discusses aspects of the reproducibility of this research and chapter 8 discusses the results. Finally, the conclusions are listed in chapter 9.

2 A background on the methods

This section presents the background important for understanding the experiments conducted in chapters 4 and 5. Firstly, section 2.1 briefly explains the dataset used for finding patterns. Then, the next two sections look at methods from literature that are useful for finding patterns. Section 2.2 explains biases and binomial significance, and section 2.3 explains kernel density estimation.

2.1 Dataset

The main objective of this research project is to find train type patterns in good Train Unit Shunting Problem (TUSP) solutions. The data used for finding patterns are real train movements from a dataset provided by ProRail, the Dutch national railway manager, and NS. They obtained the data by placing GPS trackers in a large number of NS train units. The GPS location data was combined with general information about the unit itself to form the initial dataset. Additionally, the companies had already done some important processing of the raw GPS data. Most notably, they combined it with a separate dataset about the railway network to map all original GPS coordinates to the exact track sections the train units were on for each recorded location. They also added some fields to the dataset representing general information about the current movement of a unit and condensed it by combining all records of a unit standing still into one. Finally, they split the data into seven sub-datasets based on the geographical area corresponding to seven shunting yards in the Netherlands.

2.2 Biases and significance

For the experiment with parking tracks in chapter 4 a pattern is determined by analysing the difference between the expected number of train units of a type per track and the actual number. This difference is referred to as the bias between the type and track. It is crucial to determine if a bias is valid, meaning the distribution is truly different than expected, or if it is simply due to variance in the expected distribution. For example, when tossing a coin 50 times and landing on heads 30 times, is it fair to say the coin is biased towards heads? This depends on the significance of the bias. The significance (or p-value) of a bias can be determined using Statistical Hypothesis Testing. While hypothesis testing in general is a standard procedure, how it is applied in the experiment is not straightforward. To make the method of the experiment clear, the general method as described by Emmert-Streib [9] is presented here first.

The first step is defining the test statistic T. This is a mapping between the data sample taken from the population and the test statistic value t which is being tested for bias. Then, the second step is to formally define the null hypothesis H_0 and the alternative hypothesis H_1. The null hypothesis is defined as the situation where no bias is present and that θ, the test statistic value for the entire population, follows the expected distribution. The alternative hypothesis is the opposite, namely that θ does not match what is expected. Specifically, H_1 can be defined either for the number to be less than the expectation or greater (for now only one-sided biases are considered). Formally, this is denoted as:

$$\begin{aligned} &H_0 : \theta = \theta_0 \\ &H_1 : \theta < \theta_0 \text{ or } \theta > \theta_0 \end{aligned} \tag{1}$$

Step three is to determine the sampling distribution of T given H_0 is true. This can be any type of probability distribution and depends on the nature of the test statistic. For the test described in section 4.1, each individual sample comes from a single yes/no experiment called a Bernoulli trial. The test statistic used here is the result of several Bernoulli trails, which form a binomial distribution. The probability mass function for such a distribution is given as [10]:

$$f(k, n, p) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{2}$$

Where k is the number of successes, n is the number of trials, and p is the probability of success in one trial.

The fourth step is determining the significance level α. This can be defined as the probability of making a type I error. That is the probability that at the end of the test H_0 is incorrectly rejected. The α thus determines if the bias is significant enough to be considered real and must be chosen before conducting the experiment.

Step five is to evaluate the test statistic T to obtain its value t from the data. This is explained in section 4.1. The sixth step is determining the p-value. This value is calculated by comparing the observed data t with the hypothetical sampling distribution of H_0. The p-value is the probability of observing the value t or more extreme values. For the binomial distribution, the p-value is calculated as:

$$p = \begin{cases} P(X \le t) = \sum_{i=0}^{t} \binom{n}{i} p^i (1-p)^{n-i} \text{ for } H_1 : \theta < \theta_0 \\ P(X \ge t) = \sum_{i=t}^{n} \binom{n}{i} p^i (1-p)^{n-i} \text{ for } H_1 : \theta > \theta_0 \end{cases} \tag{3}$$

The last step is to make a decision about the null hypothesis. When the p-value is smaller than α, the probability of the number of trains of the type occurring by chance alone is sufficiently small and H_0 is rejected. In this case, H_1 is assumed true and the bias is confirmed to be real.

Finally, next to determining the significance of the bias, its value can also be estimated. Formally, the bias is the difference between a known or expected value and the observed value. For the context of this study, the bias is defined as the fraction of the two. Also, since the actual population value is unknown, the bias can only be estimated using the data. The value of the bias is thus defined as:

$$bias(\theta) = \frac{\theta}{\theta_0} \approx \frac{t}{\theta_0} \tag{4}$$

2.3 Kernel density estimation

Unlike the parking track, the parking duration is a continuous variable. Therefore looking for patterns in duration involves analysing the entire distribution of all data points, as opposed to being able to group them. To obtain a probability distribution from the individual data points Kernel Density Estimation (KDE) is used in the experiment of chapter 5.

Traditionally, to estimate the probability density of a set of points one can use a histogram. However, this approach has several drawbacks. Most significantly, the subjective choice of the bin width greatly influences the shape [11]. To estimate the probability density with a histogram, one divides the space of all points into bins of a set size h and calculates the probability of the points in the bin $p(x)$ as:

$$p(x) = \frac{K}{Nh} \tag{5}$$

Where K is the number of points that fall within the bin at x and N is the total number of points.

KDE improves on this method by having an infinite number of bins centred at the location x of the data space [12]. This creates a continuous function of the probability density. The number of points that fall into each bin is given as:

$$K = \sum_{i=1}^{N} k(\frac{x - x_i}{h}) \tag{6}$$

In this formula, it is determined for each point x_i from the data if it falls within the bin around location x. Here, the function $k(u)$ is the crucial part of the KDE method, called the kernel. The kernel is a function that determines if each point falls within the bin or not. The simplest kernel is similar to the histogram approach. It only counts each point if it falls within the bin:

$$k(u) = \begin{cases} 1 & \text{if } |u| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Combining this with the original formula (5) gives a continues function for the probability density:

$$p(x) = \frac{1}{Nh} \sum_{i=1}^{N} k(\frac{x - x_i}{h}) \tag{8}$$

Furthermore, the power of KDE lies in its different options for the kernel. Where the kernel from (7), called a rectangular kernel, only counts each point completely or not at all, other kernels can give a continuous score based on the distance of the point from the centre. This eliminates the 'blocky' nature caused by the bins by smoothing out the data points [13]. A common function used as a kernel is the Gaussian distribution:

$$k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \tag{9}$$

Finally, KDE is also very suitable for estimating a PDF in multiple dimensions. This is achieved by slightly modifying the kernel.

3 Data preparation and experimental setup

This chapter describes two matters that are important for conducting the experiments described in chapters 4 and 5. Firstly, section 3.1 discusses the initial operations on the data to prepare them for answering both questions. Then, section 3.2 explains the technical setup that is used for computing the results of both experiments.

3.1 Data preparation

While the originally provided data represents train movement data, it does not represent TUSP solutions. To make the data more useful, they are first condensed further by grouping the individual location data points by each train unit's appearances in the shunting yard area. This means that one unit visiting the area once is one row or data point. When the unit leaves the area and returns at a later time it is considered a new data-point. Each appearance of a unit consists of general data about the unit and a list of movements in the area. Each movement consists of a timestamp, a track section and the type of movement. Furthermore, these data are filtered to only include units that enter the shunting yard at one point. In the end, the data are represented as a list of train unit appearances in a given timeframe and location, with each unit's exact movements in the shunting yard and surrounding area. This is a good representation of the real train data as TUSP solutions.

3.2 Experimental setup

The experiments are executed in Python, using Jupyter Notebook to organise the steps of the experiments. The notebooks are added to a small template provided by Prorail. This template already contained functionality to retrieve the data from their storage and some simple helper functions. On top of the provided code, new helper functions are added for the data preparation described in the previous section. The implementation utilises some external packages to make the data processing and visualisation easier. These include Polars, NumPy, SciPy, scikit-learn, Matplotlib and Seaborn.

4 Finding type-track patterns

This chapter explains the experiment that answers the first subquestion. The objective of the experiment is to determine if there is a pattern between train type and parking track and analyse it if present. Section 4.1 describes the method of the experiment and section 4.2 its results.

The hypothesis is that there is indeed a pattern for two reasons: parking units of the same type on the same track 1) makes the matching problem easier to solve and 2) makes finding solutions for the whole TUSP more feasible for human planners. When the matching of units can be done on the same track as they are parked, the cost of movement and using other shunting tracks is saved. Even when matching is not required (i.e. for trains consisting of one unit), human planners might favour clustering the types on specific tracks. With this tactic, when for example an SNG type is required, planners can just take one from the 'SNG track' and no additional movements are needed.

4.1 Method

The first subquestion is answered by looking for a pattern between train type and parking track in the data. Testing if this pattern exists can be done quite exactly: because both train type and parking track are discrete variables with a relatively small number of values, it is feasible to consider every combination and determine patterns on the individual level. This is done using the approach described in section 2.2. Following this approach, firstly the test statistic T is defined. In the context of this experiment, what is being investigated is if some tracks have a stronger 'preference' for certain train types as opposed to all the types being spread equally over all parking tracks. Therefore, what is formally being determined is if the probability p of a unit of a certain type being parked on a certain track is different from the fraction of all trains in the shunting yard that are of that type, which here will be referred to as p_0. These probabilities are formally defined as:

$$p_0 = \frac{u_{type}}{u_{total}} \quad p = \frac{u_{type,track}}{u_{track}} \tag{10}$$

Where u_{total} is the total number of units in the yard, u_{type} is the total number of units of the selected type, u_{track} is the total number of units parked on the selected track, and $u_{type,track}$ is the number of units of the type on the track.

The value of the test statistic t is the estimation of p obtained by using the right formula on the unit counts that can be obtained from the data. Similarly, p_0 can be obtained using the total number of units of a type in the shunting yard.

Since the alternative hypothesis H_1 can be either left or right-handed, the bias is first estimated. Using formula (4) this gives:

$$bias(p) = \frac{p}{p_0} \approx \frac{t}{p_0} \tag{11}$$

Using this bias, the null and alternative hypotheses are defined as:

$$\begin{aligned} &H_0 : p = p_0 \\ &H_1 : \begin{cases} p < p_0 & \text{if } bias(p) < 1 \\ p > p_0 & \text{if } bias(p) > 1 \end{cases} \end{aligned} \tag{12}$$

The significance level α determines how low the p-value needs to be for the bias to be valid. For this statistical test, an α of 0.01 is used. This is lower than the popular 0.05 because for this research it is important that the patterns are certain.

The distribution of the null hypothesis is binomial. This is because every train unit parked on the track can be seen as a single Bernoulli trial. Here, success is defined as the unit being of the type that is being tested for and failure as the unit being from one of the other types. Again, p_0 is the chance of each trial being a success when no bias is present.

Although the formal test statistic T used in this procedure is the probability p, to calculate the p-value in a binomial test the number of successes is used. In this context, this is $u_{type,track}$. This number is obtained from the TUSP solution data described in section 3.1. It involves first filtering the list of movements for each train unit appearance to only contain the 'movements' where the unit is listed as parked. A unit being parked is defined as a period of at least 30 minutes where it is standing still.

A large part of units is listed multiple times as parked during one visit to the shunting yard. This is because units that need servicing are moved to the servicing track at some point, parked there for a while, and then moved back to a parking track. This behaviour creates multiple possibilities for how the number of parked trains can be defined. Most crucially, a unit being parked after returning from the servicing station can be counted as a new unit being assigned a parking spot, or it can be considered a different problem of its own. The data shows that most often, a unit returns to its original parking track. Therefore, for this experiment, each unit is only counted once as being parked on a track. That is, the track it is parked on when it first arrives in the shunting yard.

Each appearance of a train unit is thus converted to a tuple consisting of only the unit type and the track it was first parked at. This track must be one of the tracks that are regularly used for parking trains and not an unusual location where the unit stood still for a longer time. The parking tracks are defined manually for each shunting yard based on its layout and units parked at other tracks are considered outliers and removed. Next, to count the number of parked units, the tuples can be grouped by train type with the parking tracks as a list per type. Finally, for each train type, the occurrences of each parking track can be counted using the list. The other unit counts u in formulae (10) can be calculated using these $u_{type,track}$.

To calculate the p-value (p_v) formula (3) for the binomial test is used. As mentioned, instead of using t directly, $u_{type,track}$ is used (here denoted as just u). Also in the formula below, n is the total number of units on the track u_{track}.

$$p_v = \begin{cases} P(X \leq u) = \sum_{i=0}^{u} \binom{n}{i} p_0^i (1-p_0)^{n-i} \text{ for } H_1 : p < p_0 \\ P(X \geq u) = \sum_{i=u}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \text{ for } H_1 : p > p_0 \end{cases} \tag{13}$$

By calculating the bias and the p-value for every type-track combination, it can be determined exactly which tracks have biases for which types and how strong these biases are.

Lastly, there is one interesting factor left to mention. The bias can be seen and calculated from two perspectives: one could say a track is biased for a particular train type, or a train type is biased for a particular track. Because only one combination of track and type is considered at a time, the bias is always the same regardless of perspective. Note that each type can have a bias in more than one combination with a track. The same is true for each track. While the value of the bias is the same regardless of from which perspective it is calculated, the p-value can be different. This is because the parameters of the binomial distribution are different from each perspective. In the steps mentioned, the process was formulated from the perspective of the track being biased for a type. When looking from the other perspective, p is the fraction of all trains going into the selected track (u_{track}/u_{total}) and n is the total number of units of the selected type u_{type}. While these two p-values are never too far apart for this dataset, it does result in some biases being slightly below α when looking from one perspective and slightly above α when looking from the other. A bias is only considered significant when both p-values are less than α.

4.2 Results

The results are obtained by running the experiment using the setup described in section 3.2. The steps are executed individually for each shunting yard in the data, except Arnhem Goederenstation, because very few units from the data park here. Since the global pattern over all tracks is similar for each shunting yard, this section only covers the results for one yard, Amersfoort Bokkeduinen. The results for this shunting yard are presented as a heat map in table 1. This table excludes the type E1700 and track 365L because for both very few units were in the data.

The table shows for each combination between a parking track and a train type the bias and two values of the significance p_{tr} and p_{ty}. The latter are the p-values corresponding to the biases calculated from the two perspectives. Here p_{tr} is the significance from the perspective of the tracks and p_{ty} is the significance from the perspective of the types. The biases are represented as the fraction of the observed number of units of each type on each track over the expected number. Here a bias greater than 1 represents a positive bias where more units were observed than expected and a bias less than 1 is a negative bias where fewer units were observed than expected. The colours highlight the severity of the bias. Biases that are not significant are coloured in grey.

The table makes it immediately clear that there are indeed significant biases between train types and parking tracks. Significant biases are present for every type and every track in the table. Out of 70 combinations, 51 have a significant bias. Three tracks have a significant bias with every train type. These are tracks 361AV, 381R and 383L. One train type, SNG, has a significant bias for every track. The ICMm type has only one insignificant bias. This can be explained by the observation that these two types are also the most prominent in this shunting yard. Because of the larger sample size for these types, the biases are more easily significant.

The most prominent biases are found in types DDZ and VIRM. These both have two positive biases of 2 or higher. Looking at the tracks there are also differences. Track 361AV is very strongly biased for two train types. 365L has two very strong negative biases. This can be explained by the low number of trains parking on the track.

One interesting observation is the correlation between positively biased track location and train service type. The tracks in the table are ordered as they are structured geographically. That is, 361AV (the first row) is the most northern parking track in the shunting yard and 383L (the last row) is the most southern track. The table shows types DDZ, ICMm and VIRM solely having positive biases for the top half of tracks, and SLT and SNG only having positive biases for the bottom half of the track. These two groups of types correspond with the two types of train service NS units are used for. DDZ, ICMm and VIRM are all used for intercity services and SLT and SNG are both used for sprinter (local) services.

For this shunting yard, there are no biases that are only significant from one perspective. However, when running the experiment on other shunting yards, there are biases like this. These shunting yards are Watergraafsmeer, Carthesiusweg, Hoofddorp, Arnhem West and Dordrecht.

5 Finding type-duration patterns

This chapter explains the second experiment to answer the second subquestion. Firstly, section 5.1 explains the method that is used to find a pattern between train type and duration. Then, section 5.2 presents the results of the experiment.

The hypothesis for this subquestion is that there is a relation between train type parking duration because some types might be needed for passenger service less frequently than others. The parking duration is useful for the shunting problem because the time a unit stays parked can be linked to its position on a First In Last Out (LIFO) track. The unit that is parked the longest should enter the parking track first and vice versa.

5.1 Method

The parking durations can be obtained from the TUSP solutions using the list of movements for each unit. Each movement has two timestamps: one for the beginning of the movement and one for the end. When a unit is moving, the time in between is ten seconds. When a unit is stationary, the time in between the timestamps is the total time the unit has not moved. For this research question, the total duration a unit has been parked is defined as the total time it is listed as parked. The time spent in the shunting yard moving and standing still for short periods is insignificant and irrelevant.

Some units need servicing. In this case, they are usually listed as parked three times: when they first park upon their arrival in the yard, when they stand still in the servicing track for a longer time, and when they are parked again after servicing is done. Unlike the other experiment where only the initial parking spot is considered, here the times for the three separate parking occasions are combined into one. This is because the total time parked in the shunting yard is important.

After calculating the parking times for each unit, the data are converted to a list of just each unit's type and the time that it was parked. Finally, all units that were parked for longer than 18 hours are removed. This is because units parked 'long term' are another special case that is not relevant to this question. Here, only units that are in service each day are considered.

To determine if there are patterns, these type-duration data are analysed on different levels. Firstly, a box plot of the duration with the train types as different groups is used. The box plot can highlight significant differences in the median and variance of the duration between the types.

If the plot does not show significant differences, the probability densities are analysed. The probability density functions (PDFs) are obtained by using Kernel Density Estimation as described in section 2.3. To obtain a smooth distribution from the data points, the Gaussian kernel is used. The bandwidth h that is used for the kernel is 1.0 because it is found that this value smooths out any outliers without losing too much detail. To find patterns, class conditional PDFs are compared for significant differences. Here, the classes are the types. Which classes are worth comparing is determined by the analysis of the box plot. The comparisons can be between two train types, or between one type and all other types.

Track	**DDZ**			**ICMm**			**SLT**			**SNG**			**VIRM**		
	bias	p_{tr}	p_{ty}	bias	p_{tr}	p_{ty}	bias	p_{tr}	p_{ty}	bias	p_{tr}	p_{ty}	bias	p_{tr}	p_{ty}
361AV	2.47	.000	.000	1.19	.000	.001	0.61	.000	.000	0.73	.000	.000	2.24	.000	.000
361BL	1.38	.012	.001	1.54	.000	.000	0.93	.187	.194	0.47	.000	.000	2.00	.000	.000
363R	2.12	.000	.000	1.40	.000	.000	1.02	.424	.426	0.58	.000	.000	0.75	.131	.127
365L	0	.101	.106	2.66	.000	.000	0	.000	.000	0.05	.000	.000	1.59	.293	.294
367L	0.90	.289	.282	1.82	.000	.000	0.71	.000	.000	0.43	.000	.000	1.15	.204	.197
273R	1.04	.433	.432	1.79	.000	.000	0.68	.000	.000	0.47	.000	.000	0.98	.503	.502
375R	1.10	.310	.307	1.47	.000	.000	0.85	.051	.059	0.61	.000	.000	1.66	.001	.001
377R	1.24	.124	.122	1.06	.135	.177	1.46	.000	.000	0.76	.000	.000	1.10	.339	.338
377L	0.77	.118	.114	0.41	.000	.000	1.62	.000	.000	1.34	.000	.000	0.36	.000	.000
379R	0.21	.000	.000	0.36	.000	.000	0.71	.001	.001	1.74	.000	.000	0.66	.051	.048
379L	0.35	.000	.000	0.31	.000	.000	0.44	.000	.000	1.85	.000	.000	0.67	.056	.053
381R	0.38	.000	.000	0.26	.000	.000	0.66	.000	.000	1.82	.000	.000	0.40	.002	.002
381L	0.33	.000	.000	0.30	.000	.000	0.98	.423	.427	1.70	.000	.000	0.22	.000	.000
383L	0.47	.001	.000	0.24	.000	.000	2.46	.000	.000	1.23	.000	.000	0.11	.000	.000

Table 1: Biases between train type and parking track at Amersfoort Bokkeduinen shunting yard. Note how a large majority of the combinations have a significant bias.

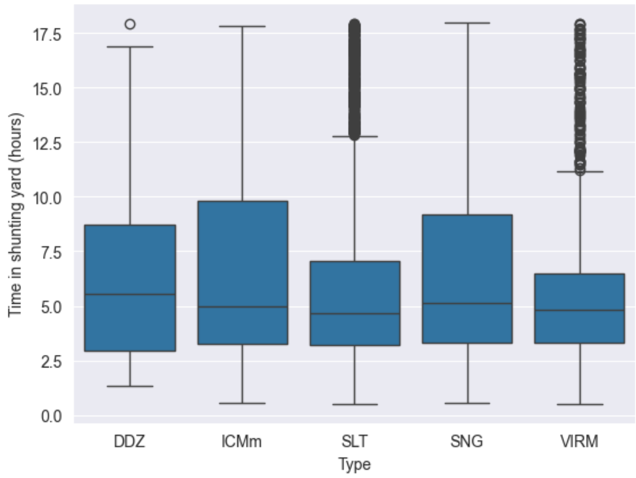

Figure 1: Box plot of parking durations per type in Utrecht Carthesiusweg shunting yard. There are no large differences between the types.

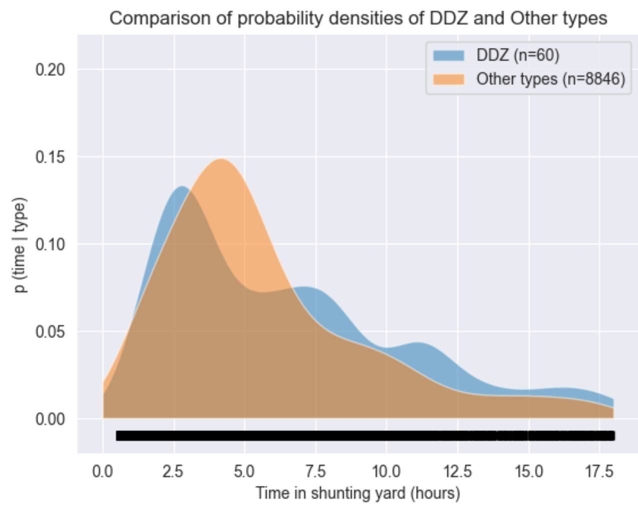

Figure 2: Class conditional probability density functions parking durations of type DDZ and all other types in Utrecht Carthesiusweg shunting yard. There are small differences, but the general distributions are very similar.

5.2 Results

The results for the experiment are computed by running it using the setup described in section 3.2. Only the most important results for the Utrecht Carthesiusweg shunting yard are presented here. Firstly, figure 1 shows the box plot of time in the shunting yard per train type. The figure shows the five boxes for the five types all laying on roughly the same interval. All boxes have a lot of overlap with the others and no box immediately stands out. This indicates that the parking durations are mostly similar between the types.

To analyse the differences in greater detail, the class condition probability density functions (PDF) are plotted. Here, two are analysed. Firstly, figure 2 displays the plot of the PDF of the DDZ train type and the PDF of all other types in this shunting yard.

Interestingly, while the box plot shows the DDZ having a median that is slightly higher than the rest, the PDF shows its peak being before the peak of the combined other types. This can be explained by the DDZ having more of its mass in the PDF distributed over the longer durations. Also, DDZ has two minor peaks at 7.5h and 11h. In contrast, the general trend of the units is to have one clear peak. While this plot shows small differences between the probabilities of the types, it does not show a clear or notable distinction in shunting time.

Secondly, the PDFs of the ICMm type and all other types are analysed using figure 3. The box plot shows this type having the most variance. The PDF shows the cause of this being a small peak for the ICMm between 12 and 17 hours. While the probability of the other types being parked for a duration longer than 12 hours is relatively small, the plot shows ICMm units are more than twice as likely to be parked for

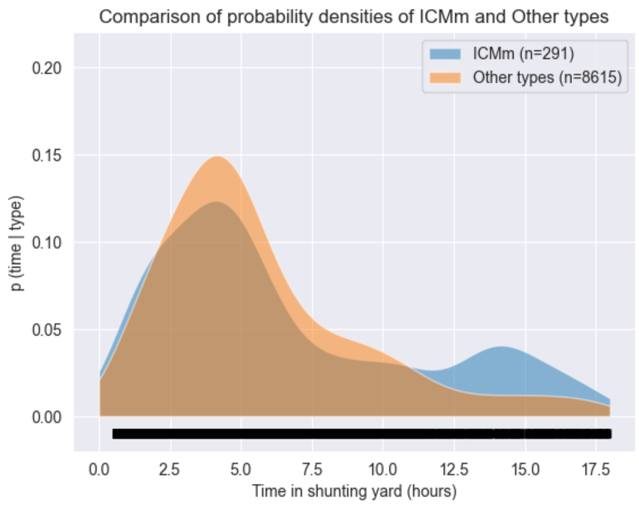

Figure 3: Class conditional probability density functions parking durations of type ICMm and all other types in Utrecht Carthesiusweg shunting yard. There is a notable difference in probability for times between 12 and 17 hours.

this duration. Despite this notable difference for ICMm, the distributions for other types are very similar.

Finally, similar results were obtained for the other examined LIFO shunting yards. These are Amersfoort Bokkeduinen and Arnhem West. Generally, the trend is that the box plots show most types having a median of around five hours. The boxes align, but there are differences in variance. When looking at the probability densities, some types show peaks that do not match the other types in the yard. One noteworthy example is the SNG type in Arnhem, which clearly shows two individual peaks compared to one broad peak for the other types.

6 Comparison with subtype

Train type is the main variable in this research and is key for answering the first two subquestions. An important distinction can be made in defining train type: each train unit has a main type indicating the name of the series (e.g. ICMm), but can also be listed as the series combined with the number of carriages that the unit consists of (e.g. ICMm-IV). In this research the first is referred to as the 'type' and the latter as the 'subtype'.

When looking for patterns, the choice can thus be made to look for patterns in type or subtype. It is hypothesised that there is a difference in patterns when looking at subtype compared to type. Therefore, the objective of this research's third subquestion is to investigate the difference in patterns when considering subtype. To accomplish this, the two experiments described in sections 4.1 and 5.1 are executed twice. Once with the main train type as the label for each unit and again with the subtype as the label. Sections 4.2 and 5.2 present the experiments' results for the main train type. This chapter presents the same results for train subtype and analyses the difference between the results.

Firstly, the results of the track-subtype experiment show some interesting patterns. The main observation is that the same biases between track and type are also present for the same track and the specific subtypes. That is, if a type is biased towards one track, so are its subtypes. However, some subtype biases are not significant because there are fewer units of the individual subtypes compared to the type total.

Track	SLT			SNG		
	SLT	SLT-IV	SLT-VI	SNG	SNG-III	SNG-IV
379L	0.44	0.56	0.36	1.85	1.60	2.15
381R	0.66	0.78	0.57	1.82	1.96	1.66
383L	2.46	1.56	3.10	1.23	1.10	1.40

Table 2: Excerpt of the biases between train subtype and parking track at Amersfoort Bokkeduinen shunting yard. The subtypes of SNG are each biased more for different tracks, while the main type is equally biased for both.

The most interesting observation is for some specific type-track combinations, the bias is much stronger for one subtype than for the other. To illustrate this, table 2 shows an excerpt from the results for Amersfoort. These results show that train type SLT is biased for track 383L with an estimated factor of 2.46. However, when looking at the subtypes, the bias is much stronger for subtype SLT-VI (3.10) than for subtype SLT-IV (1.56). Thus, where the initial experiment showed that this track is biased for type SLT, these results show that the bias is even more severe for one of its subtypes.

Furthermore, a different phenomenon can be observed. When looking at tracks 379L and 381R, it is clear that type SNG is biased for both with an almost equal factor (1.85 and 1.82). Again, it can be seen that the bias is stronger for one subtype and weaker for the other. However, which subtype has the stronger bias is different for the two tracks: SNG-III is more biased towards 381R and SNG-IV is more biased towards 379L. To summarise: both tracks are biased for SNG, but each is individually biased for a specific SNG subtype.

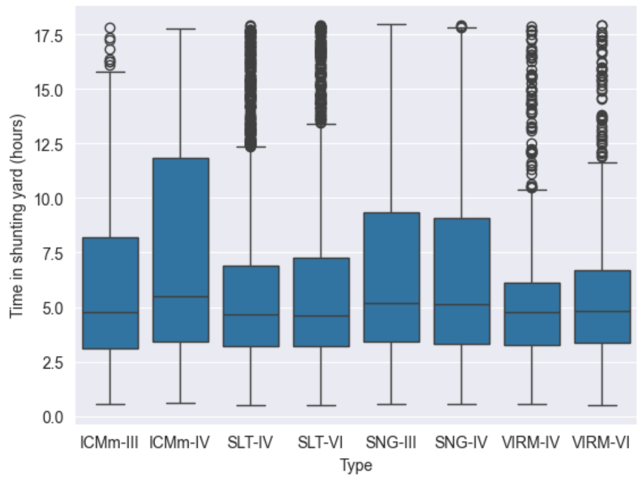

Figure 4: Box plot of parking durations per subtype in Utrecht Carthesiusweg shunting yard, without DDZ. Comparing subtype instead of type does not introduce any large differences in parking time.

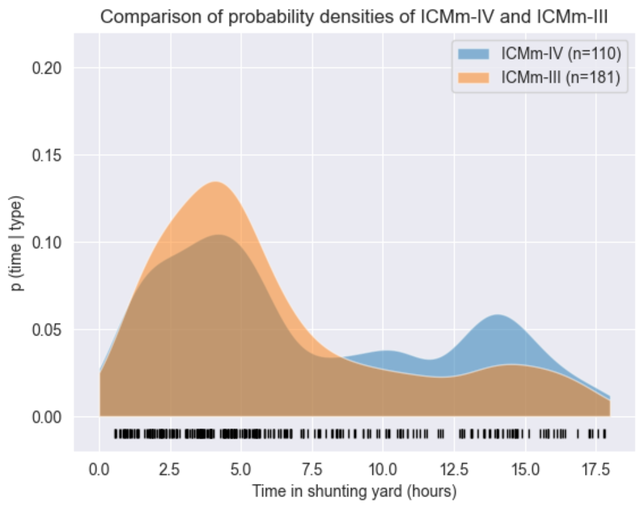

Figure 5: Class conditional probability density functions parking durations of the ICMm subtypes in Utrecht Carthesiusweg shunting yard. The spike between 12 and 17 hours is more prevalent for the ICMm-IV.

Secondly, repeating the duration-type experiment with subtypes shows similar results. The box plot of the durations and the subtypes is pictured in figure 4. Again, the general trend is that there are no large differences in the average parking time between the subtypes. It does show, however, a notable difference between the two ICMm subtypes. To further analyse this, figure 5 shows the class conditional probability density functions for the subtypes of ICMm.

The PDFs show again the small peak of the ICMm main type as discussed in section's 5.2 figure 3. However, the figure makes it clear that this higher probability for a duration between 12 and 17 hours is almost twice as high for ICMm-IV than for ICMm-III. It can also be observed that the peak for ICMm-III lies slightly later. Therefore, there are notable differences when looking at subtype, but no significant differences in pattern.

7 Responsible research

This chapter discusses two important aspects of doing research responsibly. Firstly, section 7.1 explains the consequences of the confidentiality of the dataset. Secondly, section 7.2 goes into how care has been taken to avoid confirmation bias.

7.1 Dataset confidentiality

The dataset that is used in this research consists of train movement data of NS train units on the railway network operated by ProRail. While for finding the patterns no aspects of the data are used that are unique to NS trains or the Dutch shunting yards, data for different shunting yards around the world managed by other train operators can still yield different patterns. Therefore, it is important that the methods can be reproduced on other datasets. Also, in general, it is important that this research can be reproduced on the same dataset.

This reproducibility is limited by the confidentiality of the dataset. The data is owned by ProRail and NS, and can only be accessed by authorised individuals. This limits the reproducibility of the exact research on the same data to only those within or working with NS or ProRail. Furthermore, the characteristics, such as exact fields and possible values, are also confidential. This makes recreating an identical dataset for another train network difficult. However, it is believed that the general description of the dataset given in section 2.1 is sufficient to collect and structure similar data on which the experiments described in this research can be conducted.

7.2 Avoiding confirmation bias

Humans tend to see patterns in information, even when these are not there. When actively looking for train type patterns, the mistake can easily be made to interpret the results in such a way that a desired pattern is there or not. To avoid confirmation bias influencing the conclusions of this research, great care has been taken in formulating the experiments and interpreting the results.

In designing the experiments, emphasis has been put on the significance of the results. For the first subquestion (chapter 4) this significance is made a central part of the experiment by statistically evaluating it. By setting a significance level α before the results were explored, it has been ensured that counting a bias as significant is not influenced by subjective actions. For the second subquestion (chapter 5) a pattern is specifically defined as a large difference in the distribution of durations between the types. When no clear distinction in these distributions can be made, no pattern is reported. Instead, the term 'notable difference' is used.

8 Discussion

The results show strong biases between train type and parking track, but no clear differences between the types for parking times. Also, looking at subtype makes the differences between the types more specific.

Because the used data does not contain optimal solutions but realised solutions made by human planners, the results do not necessarily indicate that the patterns are characteristic of perfect solutions. Rather, it is hypothesised that the patterns are the results of strategies used by the planners to make the problem more feasible. Nonetheless, the found patterns could be used to construct intelligent algorithms that can efficiently solve the TUSP with similar strategies as humans.

It is suspected that the absence of a pattern in parking duration is the result of the effect of the passenger service timetable for the trains. It is hypothesised that for the area of the investigated shunting yards, no one type used in passenger service is significantly less frequent.

The research is somewhat limited by the dataset. It contains only seven shunting yards of which only five were useful. This means that formally the results only apply to the investigated yards. However, it is strongly suspected that similar results would have been found for other shunting yards in the Netherlands. Different results could be found for yards in other countries because the planners might employ different planning strategies.

Furthermore, the research is also limited by a set timeframe. In defining the experiments often multiple approaches are possible. For example, for parking track, one could include parking track reassignment after servicing. Exploring and comparing all approaches would lead to more individual experiments, which would not have been possible to all complete within the set timeframe.

9 Conclusions and recommendations

The main objective of this research is to find train type patterns in solutions to the train unit shunting problem (TUSP). Specifically, subquestions include looking for patterns in parking track and parking duration. Also, the difference in patterns between train type and subtype is investigated.

Patterns between train type and parking track have been found. This confirms the hypothesis. The results show significant biases in how the train types are distributed over the available parking tracks. Almost every investigated parking track shows a bias for or against at least one train type. It can be concluded that specific tracks in shunting yards are repeatedly used to park one or more particular train types. Furthermore, the results for the shunting yard in Amersfoort indicate a geographic split in the shunting yards, where one part of the tracks is mainly used to park intercity trains and another part is mainly used to park sprinter trains.

In contrast, no clear pattern has been found between train type and parking duration. This contradicts the hypothesis. The results show that all train types are mostly likely to stay parked in the shunting yard for roughly the same amount of time. No train type clearly deviates from this. However, differences in variance have been observed. Analysis of the probability distributions shows that some types are more likely to be parked for longer durations than other types.

Analysis of the same results for train subtypes shows similar patterns and differences. It is observed that the patterns are more specific for subtypes, but that no subtype greatly deviates from its main type pattern. The results for parking track patterns show that within some pairs of tracks, one track is more biased for one subtype and the other track for the other subtype, while both tracks are equally biased for the main type. The parking duration experiment shows that some observed small differences in probabilities for longer durations are mostly the cause of one subtype staying longer.

It is recommended that future research further investigates the patterns that have been found. For example, one could look into the effect of parking track reassignment for units that return from servicing. Furthermore, train type patterns regarding the servicing itself could be investigated.

References

[1] R. Freling, R. M. Lentink, L. G. Kroon, and D. Huisman, "Shunting of passenger train units in a railway station," *Transportation Science*, vol. 39, pp. 147–288, 05 2005.

[2] L. van de Gevel, "How human knowledge can support algorithmic decision-making in the train unit shunting problem - an exemplary study," Master's thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2022.

[3] R. van den Broek, H. Hoogeveen, M. van den Akker, and B. Huisman, "A local search algorithm for train unit shunting with service scheduling," *Transportation Science*, vol. 56, pp. 1–264, 11 2021.

[4] I. K. Hanou, S. Dumancic, M. de Weerdt, P. van der Voort, R. van den Broek, and M. van den Akker, "Increasing the capacity of shunting yards within the current infrastructure: A computational perspective." In lecture notes in mobility. Available: http://issahanou.github.io/assets/pdf/tra24.pdf, 2023.

[5] Z. Lin and R. S. K. Kwan, "A two-phase approach for real-world train unit scheduling," *Public Transport*, vol. 6, pp. 35–65, 09 2013.

[6] L. G. Kroon, R. M. Lentink, and A. Schrijver, "Shunting of passenger train units: an integrated approach," *Transportation Science*, vol. 42, pp. 436–449, 12 2006.

[7] E. Beerthuizen, "Optimizing train parking and shunting at ns service yards," Master's thesis, Technische Universiteit Delft, Delft, The Netherlands, 02 2018.

[8] R. Dekker, P. Voogd, and E. van Asperen, "Advanced methods for container stacking," *OR Spectrum*, vol. 28, no. 4, p. 563–586, 2006.

[9] F. Emmert-Streib and M. Dehmer, "Understanding statistical hypothesis testing: The logic of statistical inference," *Machine Learning and Knowledge Extraction*, vol. 1, pp. 945–961, 08 2019.

[10] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures, second edition*. Chapman and Hall/CRC, 2000.

[11] S. Weglarczyk, "Kernel density estimation and its application," in *ITM Web Conf.*, vol. 23, 2018.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer New York, 2006.

[13] J. VanderPlas, *Python Data Science Handbook*. O'Reilly Media, 2016.