Computing Joint Delay Distributions for Trains in Railway Networks

Bachelor Thesis in Computer Science
by Kai Schwierczek

Supervisors: Dr. Mathias Schnee
Prof. Dr. Karsten Weihe
Advisor: Mohammad Keyhani

Fachgebiet Algorithmik
Technische Universität Darmstadt

May 22, 2013

Erklärung

Hiermit versichere ich gemäß der Allgemeinen Prüfungsbestimmungen der Technischen Universität Darmstadt (APB) §23 (7), die vorliegende Bachelorarbeit ohne Hilfe Dritter und nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus den Quellen entnommen wurden, sind als solche kenntlich gemacht worden. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 22. Mai, 2013

(Kai Schwierczek)

Abstract

Stochastic models for delay propagation in railway networks lead to promising results in the prediction of train delays. To respect waiting time policies, while preserving efficiency, these solutions assume that the departure and arrival times of trains are stochastically independent. We show, that the inherent error can become a significant problem. To tackle this problem, we first present formulas to calculate joint delay distributions for dependent trains in a basic structure. Using these distributions we can then calculate exact distributions for connecting trains. We then present a computational study comparing our calculation to a calculation, which uses the independence assumption, on a real world timetable of the German railway network (Deutsche Bahn AG). Our results show that, in the real world timetable, the error is negligible, but we still discuss how different structures could influence the result.

Contents

1 Introduction **1**

2 Structure and Examples **3**
2.1 Structure Introduction 3
2.2 Calculation Steps 3
2.3 Examples 4

3 Probability Distributions **6**
3.1 Definitions 6
3.2 Calculation of Distributions 7
3.2.1 Calculation of the Two Dimensional Distribution 7
3.2.2 Calculation of the Final Distribution 13

4 Computational Study **15**
4.1 Setup 15
4.2 Found Structures 15
4.3 Results 16
4.3.1 Two Dimensional Distributions 16
4.3.2 Departure Distribution 18
4.4 Evaluation 19
4.4.1 Discussion 19

5 Conclusion and Future Work **22**

A Arrival Distribution Plots **23**

Chapter 1

Introduction

Finding reliable connections in railway networks is an important task, since so many people use public transportation each day. Timetable information systems mostly only solve a part of this task: they find, using multiple criteria, attractive connections which are feasible according to the timetable. Modern systems factor real time delay data into their search, but they do not attempt to rate the reliability of the connections found.

Probabilistic approaches have been developed, which calculate a probability for every possible delay, resulting in probability distributions. These distributions are then propagated through the railway network and are used in the calculation of the distributions of connecting trains respecting waiting policies.

A problem these approaches have in common is that they use an assumption, that the arrival of each pair of trains at each station is independent. This assumption allows a much simpler and more efficient implementation of the probability distributions. In spite of this assumption, these approaches lead to promising results in the rating of connections.

As we will see in an example, the error, which is introduced by the independence assumption can become quite large. In the case that a train depends on several arrivals which aren't independent, the error can lead to an underestimation of the probabilities of a departure with no delay or small delays. If then another train depends on this train, the error can also lead to probabilities for those delays which are higher than they should be. This can happen because of maximum waiting times: if the connecting train only has to wait for the small delays of the previous train, and those delays are underestimated it leads to an overestimation of a punctual departure of the connecting train. So having a look at the exact calculation for dependent arrivals is worthwhile. Simple probability distributions are insufficient for this purpose, which is why joint probability distributions are introduced.

Our Contribution We present formulas for the calculation of a two dimensional joint delay distribution to support two trains depending on a common feeder train. We then use these joint delay distributions to calculate the distribution of a connecting train of these two dependent trains. Afterwards, we analyze the results of the calculation with timetables of the German railway network. We compare the result on a real world timetable to a calculation, which uses an independence assumption, and evaluate the error.

Related work Berger et al. [BGMHO11] and Keyhani et al. [KSWZ12] both use the assumption that arrivals of trains are independent. This work uses the model and formulas of [KSWZ12] for comparison. The dependency of two trains sharing a single track, where one train has a specific headway has been analyzed by Carey and Kwieciński in [CK94]. Meester and Muns studied models for stochastic delay propagation in [MM07].

Overview This thesis is organized as follows. In Chapter 2, the structure with two dependent trains, which we use throughout this thesis, is introduced, along with an example of calculating departure and arrival probabilities of the involved trains. The model we used along with the formulas for all steps of the dependent calculation are then presented in Chapter 3. The steps are the calculation of two dimensional departure and arrival distributions for the dependent trains and the calculation of the departure distribution for the connecting train. A computational study follows in Chapter 4. We search for structures in the German railway network, compare our dependent calculation against an independent calculation and discuss the result. Finally, Chapter 5 concludes this thesis and presents ideas for future work to further investigate dependency in railway networks.

Chapter 2

Structure and Examples

2.1 Structure Introduction

The structure with dependent trains we analyze can be seen in Figure 2.1. tr_2 is the connecting train departing at station s_2. It has has two feeders f_1 and f_2. Feeders are trains whose passengers have the chance to change to the connecting train. According to waiting time policies, the connecting train might have to wait for feeders if they arrive delayed. The two feeders themselves depend on a common feeder tr_1 at their previous station s_1. This results in a dependency between the arrivals of f_1 and f_2. There may be additional independent feeders for the departures of f_1 and f_2 at station s_1 and for the departure of tr_2 at station s_2. In the formulas we present, all four trains have to be pairwise different. In general it could be possible that tr_1 is the same as f_1 for example.

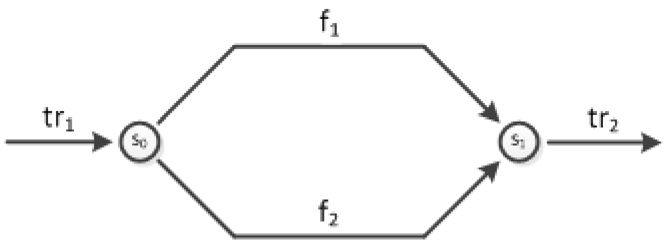


Figure 2.1: Basic structure with a dependency.

This is probably the simplest structure with a dependency, though the calculation of the joint delay distribution is already complicated. Also, in this simple structure we do not have too many influences, which could diminish the dependency. Imagine f_1 and f_2 would part ways after their departure at s_1, stop several times and then get back together at s_2. At each stop they can be influenced independently, till their common feeder is not important anymore.

2.2 Calculation Steps

First we calculate a two dimensional joint probability distribution for the departure of f_1 and f_2 at station s_1. We then apply travel time distributions, which model the possibility to introduce further delays or catch up delays during the travel from one station to another. The travel time distributions are applied independently, since we have no data available to determine dependent distributions for this step. The result is a two dimensional arrival distribution. The formulas we use for these two tasks are described in Section 3.2.1.

The final step is the calculation of the departure distribution of tr_2 using the calculated two dimensional arrival distribution (see Section 3.2.2). At this point the dependent calculation is finished. In the following there are two examples of the calculation to illustrate the difference to a calculation using the independence assumption.

2.3 Examples

For the sake of simplicity, let's assume, that all trains except tr_1 arrive at their scheduled time, f_1 and f_2 travel simultaneously, needing their scheduled travel time, and any delay results in f_1 and f_2 and in turn tr_2 having to wait. There are no other independent feeders for any of the trains.

Example One. If train tr_1 has a 50% probability to arrive on time and another 50% to arrive one minute late, f_1 and f_2 will both have a 50% probability to depart one minute late, too. With the independence assumption one would falsely conclude that every combination of delays of f_1 and f_2 has a probability of 25%, which would result in tr_2 departing one minute late with a chance of 75%. Actually f_1 and f_2 can only be delayed both at the same time, so tr_2 in turn only has a 50% chance of departing late, too. This example is illustrated in Figure 2.2.

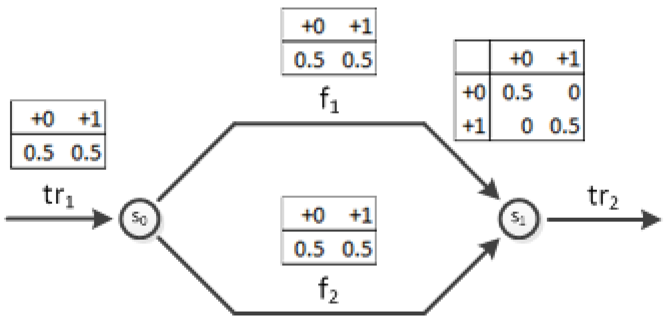


Figure 2.2: Example One.

This shows, the immediate case of underestimating a punctual departure. Now for another example, which is a bit more complex and shows the case where a punctual departure of a connecting train of tr_1 is overestimated with the independence assumption.

Example Two. For this example we introduce a connecting train of tr_2 called tr_3. In this example train tr_3 does only have to wait for one minute for tr_2. Train tr_0 arrives punctual with a 10% probability and delayed by one or two minutes with chance of 50% and 40% respectively. Figure 2.3 shows the result

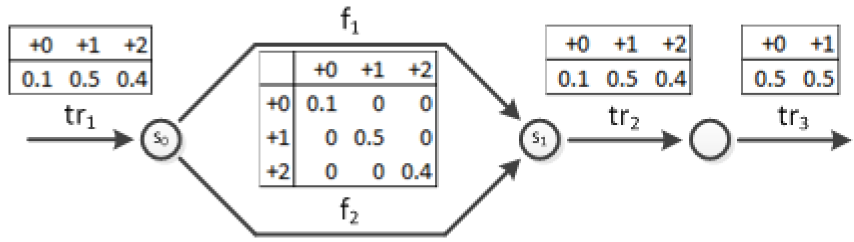


Figure 2.3: Example Two, using dependent calculation.

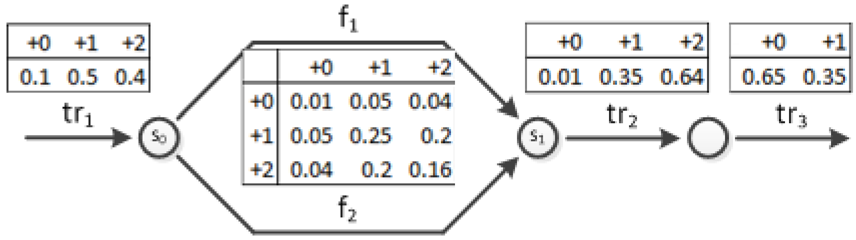


Figure 2.4: Example Two, using independent calculation.

of a dependent calculation, while Figure 2.4 shows the result of an independent calculation.

As we can see in this case the underestimation of a punctual departure of tr_2 leads to an overestimation of a punctual departure of tr_3.

Chapter 3

Probability Distributions

3.1 Definitions

For comparison reasons we use the same model as Keyhani et al. in [KSWZ12] which is based on the model introduced by Berger et al. in [BGMHO11]. A railway timetable $TT = (TR, S, EC)$ includes trains, stations and connections.

Each train $tr \in TR$ starts at one station s_0 and is connected to its destination s_{end} by a chain of elementary connections $ec_i \in EC; i \in [0, ..., end - 1]$. Each elementary connection ec_i consists of a tuple of events $ec_i = (dep_{tr,s_i}, arr_{tr,s_{i+1}})$. The first event of the tuple is the departure event at one station, the second event is the arrival event at the next station. Let $EVENTS$ be the set of all events. Each $event \in EVENTS$ in the timetable has a scheduled time, which is denoted by $sched(event)$.

There are several constraints for departures. A train may never depart before its scheduled departure time. After arriving at any station s each train tr has a minimal standing time before it may depart again. This minimal standing time is defined by $stand(tr, s)$. Furthermore, each train tr has a set of feeders at each station s. Feeders are trains which arrive before tr at the station s with enough time for the passengers to change from the feeder to tr. This change time is defined for each feeder f as $transfer(f, tr)$. Also there is the waiting time $wait(tr, f)$, which is the amount of time tr waits for its feeder f after its scheduled departure time. This is the waiting time policy used by Deutsche Bahn AG. The latest feasible arrival time of a feeder f is then the latest time f has to arrive so that tr still has to wait. It can be calculated as $lfa(tr, f, s) = sched(dep_{tr,s}) + wait(tr, f) - transfer(f, tr)$[1]. Trains which arrive earlier than $sched(dep_{tr,s}) - \gamma$ are not considered as potential feeders for tr. In this work, we set $\gamma = 30$ minutes. The resulting set of feeders for tr at a station s is denoted by $FD(tr, s) \subset TR$.

During daily operation delays can occur. A departure of a train can be delayed due to the train arriving late itself, because of previous delays and the minimum standing time, but also due to the waiting time policies, if a feeder arrives too late and the train has to wait for it. Because no train may depart before its scheduled departure time, departure delays can only be positive, while arrivals can also happen before their scheduled time.

[1]In [KSWZ12] this is called *latestFeasibleArr*

The probabilistic approach uses discrete random variables, which map the sample space Ω of the underlying discrete probability space (Ω, A, P) with σ-Algebra A and probability measure P to time-stamps in minutes. We define for each $event \in EVENTS$ the discrete random variable $X_{event} : \Omega \mapsto \{..., sched(event) - 1, sched(event), sched(event) + 1, ...\}$. The calculation of the probability distributions with the assumption, that all arrival times of feeder trains are stochastically independent is explained in [KSWZ12] with the addition of probabilities for whole connections, which we do not look at in this work.

For travel times between two stations, there is a set of travel distributions $X_{travel} = \{X^d_{travel} | d \in \mathbb{N}\}$. The distribution X^d_{travel} is to be used for a departure delay of $d \in \mathbb{N}_0$. They are used to model the possibility of making up delays on the track, or introducing further delays. We have two sets of travel time distributions. We have generated distributions, which depend on the scheduled travel times and distributions from Deutsche Bahn AG.

3.2 Calculation of Distributions

3.2.1 Calculation of the Two Dimensional Distribution

The following formulas are used to calculate the two dimensional distribution of the dependent trains f_1 and f_2 at station s. Dependent means, that they have got a common feeder tr_1 which is element of $FD(f_1, s)$ and $FD(f_2, s)$. Also there has to be a time interval in which f_1 and f_2 both have to wait for tr_1, otherwise there is no dependency. So $[sched(dep_{f_1,s}) - transfer(tr_1, f_1) + 1, lfa(f_1, tr_1, s)] \cap [sched(dep_{f_2,s}) - transfer(tr_1, f_2) + 1, lfa(f_2, tr_1, s)] \neq \emptyset$. The formula is only for exactly one common feeder.

Departure. First, the main formulas to calculate the departure distribution for each possible case are explained. The explanation is divided into a formula for the departure after the scheduled time and a departure at the scheduled time. Afterwards, an explanation of the referenced formulas follows.

The two dimensional probability distribution for the departure events $dep_{f_1,s}$ and $dep_{f_2,s}$ can then be determined by using the formulas to get the probabilities $P\big(X_{dep_{f_1,s}} = t_1 \cap X_{dep_{f_2,s}} = t_2\big)$ for all $t_1 \in [sched(dep_{f_1,s}),\ t_{end_1}], t_2 \in [sched(dep_{f_2,s}), t_{end_2}]$, where t_{end_1} and t_{end_2} are chosen so that there are no times $t_1 > t_{end_1}, t_2$ and no times $t_1, t_2 > t_{end_2}$ with $P\big(X_{dep_{f_1,s}} = t_1 \cap X_{dep_{f_2,s}} = t_2\big) > 0$. The lower bound of the distribution is the scheduled departure time, since early departures are not allowed.

Arrival. After the departure formulas the formula to get the two dimensional arrival distribution for f_1 and f_2 at the next station where both are feeders for a connecting train is presented. It uses the calculated departure distribution and travel time distributions is presented.

The two dimensional probability distribution for the arrival events $arr_{f_1,s}$ and $arr_{f_2,s}$ can then be determined by using the formula to get the probabilities $P\big(X_{arr_{f_1,s}} = t_1 \cap X_{arr_{f_2,s}} = t_2\big)$ for all $t_1 \in [t_{start_1}, t_{end_1}]$, $t_2 \in [t_{start_2}, t_{end_2}]$, where t_{start_1}, t_{end_1}, t_{start_2} and t_{end_2} are chosen so that there are no times

$t_1 \notin [t_{start_1}, t_{end_1}], t_2$ and no times $t_1, t_2 \notin [t_{start_2}, t_{end_2}]$ with $P\big(X_{dep_{f_1,s}} = t_1 \cap X_{dep_{f_2,s}} = t_2\big) > 0$.

The calculation makes use of the Boolean function $B(\mathit{condition})$, which results in 1 if *condition* is true, and in 0 if *condition* is false.

3.2.1.1 Departing After the Scheduled Time

This is the main formula used to calculate the two dimensional departure distribution of f_1 and f_2 in case both trains depart at a delayed time. Train f_1 departs delayed at time $t_1 > \mathit{sched}(dep_{f_1,s})$ and train f_2 departs delayed at time $t_2 > \mathit{sched}(dep_{f_2,s})$. This happens in several cases. The first case is that they do not have to wait for other feeders at all but are delayed themselves. The probabilities of not waiting for feeders are $P_{no\,WaitingForIndepFeeders}$ and $P_{no\,WaitingForDepFeeder}$. The calculation of these probabilities is explained in chapters 3.2.1.3 and 3.2.1.4. Then there are cases, where the common feeder tr_1 is involved. Here tr_1 can delay both trains, or only one of them, while the other train could have arrived late, or has to wait for an independent feeder. The probability of waiting for an independent feeder is $P_{waitingForIndepFeeder}$, which is explained in 3.2.1.5. Finally there are cases, where both do not have to wait for tr_1, but either one, or both of f_1 and f_2 have to wait for independent feeders. In the following, we explain how the different cases are handled.

1. The delayed departures of f_1 and f_2 are because of f_1 and f_2 arriving late themselves:

   - f_1 has a delay and arrives at time $t_{1,arr} = t_1 - \mathit{stand}(f_1, s)$,
   - f_2 has a delay and arrives at time $t_{2,arr} = t_2 - \mathit{stand}(f_2, s)$,
   - f_1 and f_2 do not have to wait for any feeders.

$$
\begin{aligned}
&P_{selfDelay}\big(f_1, t_1, f_2, t_2\big) = \\
&\quad P\big(X_{arr_{f_1,s}} = t_1 - \mathit{stand}(f_1, s)\big) \cdot P\big(X_{arr_{f_2,s}} = t_2 - \mathit{stand}(f_2, s)\big) \\
&\qquad \cdot P_{no\,WaitingForDepFeeder}(tr_1, s, t_1 - \mathit{transfer}(tr_1, f_1), \\
&\qquad\qquad t_2 - \mathit{transfer}(tr_1, f_2), \mathit{lfa}(f_1, tr_1, s), \mathit{lfa}(f_2, tr_1, s)) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_1, s, t_1) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_2, s, t_2)
\end{aligned}
$$

2. Train tr_1 is responsible for at least one delayed departure:

   (a) Both f_1 and f_2 have to wait for tr_1. This happens if

   - f_1 arrives at time $t_{1,arr} < t_1 - \mathit{stand}(f_1, s)$,
   - f_2 arrives at time $t_{2,arr} < t_2 - \mathit{stand}(f_2, s)$,
   - tr_1 arrives at time $t_1 - \mathit{transfer}(tr_1, f_1)$,
   - $t_1 - \mathit{transfer}(tr_1, f_1) = t_2 - \mathit{transfer}(tr_1, f_2)$,
   - $t_1 - \mathit{transfer}(tr_1, f_1) \leq \mathit{lfa}(f_1, tr_1, s)$,
   - $t_2 - \mathit{transfer}(tr_1, f_2) \leq \mathit{lfa}(f_2, tr_1, s)$,
   - f_1 and f_2 do not have to wait for any independent feeders.

$$
\begin{aligned}
&P_{bothDepDelay}\big(f_1, t_1, f_2, t_2\big) = \\
&\quad P\big(X_{arr_{f_1,s}} < t_1 - \mathit{stand}(f_1, s)\big) \cdot P\big(X_{arr_{f_2,s}} < t_2 - \mathit{stand}(f_2, s)\big) \\
&\qquad \cdot P\big(X_{arr_{tr_1,s}} = t_1 - \mathit{transfer}(tr_1, f_1)\big) \\
&\qquad \cdot B\big(t_1 - \mathit{transfer}(tr_1, f_1) = t_2 - \mathit{transfer}(tr_1, f_2)\big) \\
&\qquad \cdot B\big(t_1 - \mathit{transfer}(tr_1, f_1) \leq \mathit{lfa}(f_1, tr_1, s)\big) \\
&\qquad \cdot B\big(t_2 - \mathit{transfer}(tr_1, f_2) \leq \mathit{lfa}(f_2, tr_1, s)\big) \\
&\qquad \cdot P_{noWaitingForIndepFeeders}(f_1, s, t_1) \\
&\qquad \cdot P_{noWaitingForIndepFeeders}(f_2, s, t_2)
\end{aligned}
$$

(b) Only f_1 or f_2 has to wait for tr_1. This happens if (exemplary for f_1, analogous for f_2)

- f_1 arrives at time $t_{1,arr} < t_1 - \mathit{stand}(f_1, s)$
- tr_1 arrives at time $t_1 - \mathit{transfer}(tr_1, f_1)$,
- $t_1 - \mathit{transfer}(tr_1, f_1) \leq \mathit{lfa}(f_1, tr_1, s)$,
- f_1 does not have to wait for any independent feeders,
- f_2 does not have to wait for tr_1 and f_2 either
  i. arrives at time $t_{2,arr} = t_2 - \mathit{stand}(f_2, s)$ and does not have to wait for any independent feeder, too or
  ii. arrives at time $t_{2,arr} < t_2 - \mathit{stand}(f_2, s)$ and does have to wait for at least one independent feeder.

$$
\begin{aligned}
&P_{oneDepDelay}\big(f_1, t_1, f_2, t_2\big) = \\
&\quad P\big(X_{arr_{f_1,s}} < t_1 - \mathit{stand}(f_1, s)\big) \cdot P\big(X_{arr_{tr_1,s}} = t_1 - \mathit{transfer}(tr_1, f_1)\big) \\
&\qquad \cdot P_{noWaitingForIndepFeeders}(f_1, s, t_1) \\
&\qquad \cdot B\big(t_1 - \mathit{transfer}(tr_1, f_1) \leq \mathit{lfa}(f_1, tr_1, s)\big) \\
&\qquad \cdot B\big(t_2 - \mathit{transfer}(tr_1, f_2) \leq t_1 - \mathit{transfer}(tr_1, f_1) \\
&\qquad\qquad \vee\, t_2 - \mathit{transfer}(tr_1, f_2) > \mathit{lfa}(f_2, tr_1, s)\big) \\
&\qquad \cdot \big[P\big(X_{arr_{f_2,s}} = t_2 - \mathit{stand}(f_2, s)\big) \cdot P_{noWaitingForIndepFeeders}(f_2, s, t_2) \\
&\qquad\quad + P\big(X_{arr_{f_2,s}} < t_2 - \mathit{stand}(f_2, s)\big) \cdot P_{waitingForIndepFeeder}(f_2, s, t_2) \\
&\qquad\qquad \cdot B\big(t_2 - \mathit{transfer}(tr_1, f_2) < t_1 - \mathit{transfer}(tr_1, f_1) \\
&\qquad\qquad\qquad \vee\, t_2 - \mathit{transfer}(tr_1, f_2) > \mathit{lfa}(f_2, tr_1, s)\big)\big]
\end{aligned}
$$

3. The delayed departure is because at least one of the trains has to wait for at least one independent feeder, while both trains do not have to wait for the dependent feeder:

(a) Only f_1 or f_2 has to wait for at least one independent feeder, the other one is delayed itself. This happens if (exemplary for f_1, analogous for f_2)

- f_2 arrives at time $t_{2,arr} = t_2 - \mathit{stand}(f_2, s)$,
- f_2 does not have to wait for any feeder,
- f_1 arrives at time $t_{1,arr} < t_1 - \mathit{stand}(f_1, s)$,
- f_1 also does not have to wait for tr_1,

- f_1 does have to wait for at least one independent feeder.

$$P_{oneIndepDelay}\big(f_1, t_1, f_2, t_2\big) =$$
$$P\big(X_{arr_{f_1,s}} < t_1 - \mathit{stand}(f_1, s)\big) \cdot P\big(X_{arr_{f_2,s}} = t_2 - \mathit{stand}(f_2, s)\big)$$
$$\cdot P_{waitingForIndepFeeder}(f_1, s, t_1) \cdot P_{noWaitingForIndepFeeder}(f_2, s, t_2)$$
$$\cdot P_{noWaitingForDepFeeder}(tr_1, s, t_1 - \mathit{transfer}(tr_1, f_1) - 1,$$
$$t_2 - \mathit{transfer}(tr_1, f_2), \mathit{lfa}(f_1, tr_1, s), \mathit{lfa}(f_2, tr_1, s))$$

(b) Both f_1 and f_2 have to wait for independent feeders. This happens if
- f_1 arrives at time $t_{1,arr} < t_1 - \mathit{stand}(f_1, s)$,
- f_2 arrives at time $t_{2,arr} < t_2 - \mathit{stand}(f_2, s)$,
- f_1 and f_2 do not have to wait for tr_1,
- f_1 and f_2 both have to wait for at least one independent feeder.

$$P_{bothIndepDelay}\big(f_1, t_1, f_2, t_2\big) =$$
$$P\big(X_{arr_{f_1,s}} < t_1 - \mathit{stand}(f_1, s)\big) \cdot P\big(X_{arr_{f_2,s}} < t_2 - \mathit{stand}(f_2, s)\big)$$
$$\cdot P_{waitingForIndepFeeder}(f_1, s, t_1) \cdot P_{waitingForIndepFeeder}(f_2, s, t_2)$$
$$\cdot P_{noWaitingForDepFeeder}(tr_1, s, t_1 - \mathit{transfer}(tr_1, f_1) - 1,$$
$$t_2 - \mathit{transfer}(tr_1, f_2) - 1, \mathit{lfa}(f_1, tr_1, s), \mathit{lfa}(f_2, tr_1, s))$$

Since the cases are chosen to not overlap, the total probability of f_1 departing at t_1 and f_2 departing at t_2 is the sum of the cases above:

$$P\big(X_{dep_{f_1,s}} = t_1 \cap X_{dep_{f_2,s}} = t_2\big) =$$
$$P_{selfDelay}\big(f_1, t_1, f_2, t_2\big)$$
$$+ P_{bothDepDelay}\big(f_1, t_1, f_2, t_2\big)$$
$$+ P_{oneDepDelay}\big(f_1, t_1, f_2, t_2\big)$$
$$+ P_{oneDepDelay}\big(f_2, t_2, f_1, t_1\big)$$
$$+ P_{oneIndepDelay}\big(f_1, t_1, f_2, t_2\big)$$
$$+ P_{oneIndepDelay}\big(f_2, t_2, f_1, t_1\big)$$
$$+ P_{bothIndepDelay}\big(f_1, t_1, f_2, t_2\big)$$

3.2.1.2 Departing at the Scheduled Time

If t_1 is the scheduled departure time of f_1 or t_2 is the scheduled departure time of f_2 several cases change. The main difference is, that for a departure at the scheduled time, no feeder may delay the departure of the train and the train has to arrive early enough for the departure (w.r.t. the standing time). So, for a departure at the scheduled time, parts of the formula above which are about a train waiting ($X_{arr_{f,s}} < t$) are omitted. Also, parts which are about the train itself having a delay change ($X_{arr_{f,s}} = t$ becomes $X_{arr_{f,s}} \leq t$) because a arrival, which is before the scheduled departure time minus the standing, does not have a different result than an arrival at this time: the train can depart on time in both cases.

Thus, the case that train f_1 departs at time $t_1 = \mathit{sched}(dep_{f_1,s})$ and train f_2 departs at time $t_2 = \mathit{sched}(dep_{f_2,s})$ leaves only one part. It only happens if they arrive in time to depart at the given time and do not have to wait for any feeder at all:

$$
\begin{aligned}
&P\big(X_{dep_{f_1,s}} = t_1 \cap X_{dep_{f_2,s}} = t_2\big) = \\
&\quad P\big(X_{arr_{f_1,s}} \le t_1 - \mathit{stand}(f_1,s)\big) \cdot P\big(X_{arr_{f_2,s}} \le t_2 - \mathit{stand}(f_2,s)\big) \\
&\qquad \cdot P_{no\,WaitingForDepFeeder}(tr_1, s, t_1 - \mathit{transfer}(tr_1, f_1), \\
&\qquad\qquad t_2 - \mathit{transfer}(tr_1, f_2), \mathit{lfa}(f_1, tr_1, s), \mathit{lfa}(f_2, tr_1, s)) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_1, s, t_1) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_2, s, t_2)
\end{aligned}
$$

In the case that one train departs at its scheduled time all parts where that train does not have to wait remain. There are changed according to the explanation above. This results in the formula still becoming a lot shorter than the original one.

The probability of f_1 departing at the scheduled time $t_1 = \mathit{sched}(dep_{f_1,s})$ and f_2 departing delayed at time $t_2 > \mathit{sched}(dep_{f_2,s})$ can then be calculated with the following formula. To get the formula for f_2 departing at its scheduled time $t_2 = \mathit{sched}(dep_{f_2,s})$ and f_1 departing delayed at time $t_1 > \mathit{sched}(dep_{f_1,s})$ simply exchange all parts regarding f_1 and f_2.

$$
\begin{aligned}
&P\big(X_{dep_{f_1,s}} = t_1 \cap X_{dep_{f_2,s}} = t_2\big) = \\
&\quad P\big(X_{arr_{f_1,s}} \le t_1 - \mathit{stand}(f_1,s)\big) \cdot P\big(X_{arr_{f_2,s}} = t_2 - \mathit{stand}(f_2,s)\big) \\
&\qquad \cdot P_{no\,WaitingForDepFeeder}(tr_1, s, t_1 - \mathit{transfer}(tr_1, f_1), \\
&\qquad\qquad t_2 - \mathit{transfer}(tr_1, f_2), \mathit{lfa}(f_1, tr_1, s), \mathit{lfa}(f_2, tr_1, s)) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_1, s, t_1) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_2, s, t_2) \\
&\quad + P\big(X_{arr_{f_2,s}} < t_2 - \mathit{stand}(f_2,s)\big) \cdot P\big(X_{arr_{tr_1,s}} = t_2 - \mathit{transfer}(tr_1, f_2)\big) \\
&\qquad \cdot P_{no\,WaitingForIndepFeeders}(f_2, s, t_2) \\
&\qquad \cdot B\big(t_2 - \mathit{transfer}(tr_1, f_2) \le \mathit{lfa}(f_2, tr_1, s)\big) \\
&\qquad \cdot B\big(t_1 - \mathit{transfer}(tr_1, f_1) \le t_2 - \mathit{transfer}(tr_1, f_2) \\
&\qquad\qquad \vee\, t_1 - \mathit{transfer}(tr_1, f_1) > \mathit{lfa}(f_1, tr_1, s)\big) \\
&\qquad \cdot P\big(X_{arr_{f_1,s}} \le t_1 - \mathit{stand}(f_1,s)\big) \cdot P_{no\,WaitingForIndepFeeders}(f_1, s, t_1) \\
&\quad + P\big(X_{arr_{f_1,s}} \le t_1 - \mathit{stand}(f_1,s)\big) \cdot P\big(X_{arr_{f_2,s}} < t_2 - \mathit{stand}(f_2,s)\big) \\
&\qquad \cdot P_{waitingForIndepFeeder}(f_2, s, t_2) \cdot P_{no\,WaitingForIndepFeeder}(f_1, s, t_1) \\
&\qquad \cdot P_{no\,WaitingForDepFeeder}(tr_1, s, t_1 - \mathit{transfer}(tr_1, f_1), \\
&\qquad\qquad t_2 - \mathit{transfer}(tr_1, f_2) - 1, \mathit{lfa}(f_1, tr_1, s), \mathit{lfa}(f_2, tr_1, s))
\end{aligned}
$$

3.2.1.3 Probability of Not Waiting for Independent Feeders

This formula is the same as the formula explained in [KSWZ12, A.1], with the exception, that tr_1 is the train, which is excluded in the calculation of the departure distribution of f_1 and f_2.

$$P_{noWaitingForIndepFeeders}\big(tr, s, t)\big) = \prod_{f \in FD(tr,s)\setminus\{tr_1\}} \big[P\big(X_{arr_{tr,s}} > lfa(tr, f, s)\big) + P\big(X_{arr_{tr,s}} \leq min\{t - transfer(f, tr), lfa(tr, f, s)\}\big)\big]$$

3.2.1.4 Probability of Not Waiting for the Dependent Feeder

The probability of f_1 and f_2 both not having to wait for the dependent feeder tr_1 depends on whether f_1, f_2, or both arrive at time $t_{arr} = t - stand(f, s)$ or at time $t_{arr} < t - stand(f, s)$ because the dependent feeder has priority in the formula. In the first case tr_1 may arrive before or at time $t - transfer(tr_1, f)$ or after the latest feasible arrival time. In the second case tr_1 must arrive before time $t - transfer(tr_1, f)$ or after the latest feasible arrival time. Otherwise the case for calculating the probability of only having to wait for independent feeders could overlap with the case of waiting for the dependent feeder.

So we define $P_{noWaitingForDepFeeder}(tr, s, t_{arrLimit_1}, t_{arrLimit_2}, t_{lfa_1}, t_{lfa_2})$ depending on the specified input times, where $t_{arrLimit}$ is either $t - transfer(tr_1, f)$ or $t - transfer(tr_1, f) - 1$ and t_{lfa} are the latest feasible arrival times.

The probability of both trains not having to wait for the dependent feeder needs, in contrast to the probability of not waiting for independent feeders a distinction of cases. In the formula for independent feeders this distinction is not necessary because it is covered by the use of min in $P\big(X_{arr_{tr,s}} \leq min\{t - transfer(f, tr), lfa(tr, f, s)\}\big)$.

The trains do not have to wait for the dependent feeder at the same time in the following cases:

1. Both trains do not have to wait under any circumstances. This happens if
   - $t_{arrLimit_1} \geq t_{lfa_1}$ and
   - $t_{arrLimit_2} \geq t_{lfa_2}$.
2. Only one train has to wait for certain arrival times of tr. They do not have to wait if (exemplary for the first one)
   - $t_{arrLimit_1} < t_{lfa_1}$,
   - $t_{arrLimit_2} \geq t_{lfa_2}$ and
   - tr arrives at $t \leq t_{arrLimit_1}$ or at $t > t_{lfa_1}$.
3. Both trains could have to wait for certain arrival times of tr. They do not have to wait if
   - $t_{arrLimit_1} < t_{lfa_1}$,
   - $t_{arrLimit_2} < t_{lfa_2}$,
   - (a) tr arrives at $t \leq min\{t_{arrLimit_1}, t_{arrLimit_2}\}$ or
     (b) tr arrives at $t > max\{t_{lfa_1}, t_{lfa_2}\}$ or
     (c) tr arrives at $min\{t_{lfa_1}, t_{lfa_2}\} < t \leq max\{t_{arrLimit_1}, t_{arrLimit_2}\}$.

$$
\begin{aligned}
&P_{noWaitingForDepFeeder}\big(tr, s, t_{arrLimit_1}, t_{arrLimit_2}, t_{lfa_1}, t_{lfa_2}\big) = \\
&\quad B\big(t_{arrLimit_1} \geq t_{lfa_1}\big) \cdot B\big(t_{arrLimit_2} \geq t_{lfa_2}\big) \\
&\quad + B\big(t_{arrLimit_1} < t_{lfa_1}\big) \cdot B\big(t_{arrLimit_2} \geq t_{lfa_2}\big) \\
&\qquad \cdot \big[P\big(X_{arr_{tr,s}} \leq t_{arrLimit_1}\big) + P\big(X_{arr_{tr,s}} > t_{lfa_1}\big)\big] \\
&\quad + B\big(t_{arrLimit_1} \geq t_{lfa_1}\big) \cdot B\big(t_{arrLimit_2} < t_{lfa_2}\big) \\
&\qquad \cdot \big[P\big(X_{arr_{tr,s}} \leq t_{arrLimit_2}\big) + P\big(X_{arr_{tr,s}} > t_{lfa_2}\big)\big] \\
&\quad + B\big(t_{arrLimit_1} < t_{lfa_1}\big) \cdot B\big(t_{arrLimit_2} < t_{lfa_2}\big) \\
&\qquad \cdot \big[P\big(X_{arr_{tr,s}} \leq min\{t_{arrLimit_1}, t_{arrLimit_2}\}\big) \\
&\qquad\quad + P\big(X_{arr_{tr,s}} > max\{t_{lfa_1}, t_{lfa_2}\}\big) \\
&\qquad\quad + P\big(min\{t_{lfa_1}, t_{lfa_2}\} < X_{arr_{tr,s}} \leq max\{t_{arrLimit_1}, t_{arrLimit_2}\}\big)\big]
\end{aligned}
$$

3.2.1.5 Probability of Waiting for Independent Feeders

This formula is the same as the formula $P_{waitingForFeeders}$ in [KSWZ12, A.2], with the exception, that tr_1 is the train, which is excluded in the calculation:

$$
\begin{aligned}
P_{waitingForIndepFeeders}\big(tr, s, t)\big) = &\prod_{f \in FD(tr,s) \setminus \{tr_1\}} \big[1 - P\big(Y_{arr_{f,s}} > t\big)\big] \\
&- \prod_{f \in FD(tr,s) \setminus \{tr_1\}} \big[1 - P\big(Y_{arr_{f,s}} > t - 1\big)\big].
\end{aligned}
$$

3.2.1.6 Probability of Arriving at a Given Time

The probability of f_1 and f_2 arriving at specific times at station s_2 after departing at station s_1 depends on their two dimensional departure distribution $X_{dep_{f_1,s_1}}$, $X_{dep_{f_2,s_1}}$ and their respective travel time distributions $X^d_{travel_1}$ and $X^d_{travel_2}$ which are, as explained, assumed to be independent:

$$
\begin{aligned}
&P\big(X_{arr_{f_1,s_2}} = a_1 \cap X_{arr_{f_2,s_2}} = a_2\big) = \\
&\quad \sum_{d_1=0}^{a_1} \sum_{d_2=0}^{a_2} \big[P\big(X_{dep_{f_1,s_1}} = d_1 \cap X_{dep_{f_2,s_1}} = d_1\big) \\
&\qquad\qquad \cdot P\big(X^{d_1}_{travel_1} = a_1 - d_1\big) \cdot P\big(X^{d_2}_{travel_2} = a_2 - d_2\big)\big].
\end{aligned}
$$

3.2.2 Calculation of the Final Distribution

The departure distribution of tr_2 is the last step of the dependent feeder calculation. It is calculated as described in [KSWZ12, ch. 3.2.1] with a few exceptions. All parts regarding a specific connection are left out. Also special handling of f_1 and f_2 is necessary to actually use the calculated two dimensional arrival distribution.

In the following the changes for f_1 and f_2 are explained.

3.2.2.1 Probability of Not Waiting for Feeders

This formula is the same as the formula explained in [KSWZ12, ch. A.1], with the exception, that f_1 and f_2 require a special handling, if both are relevant for the formula.

They are relevant, if a dependency is actually present. If there is no possibility for tr_2 to have to wait for either f_1 or f_2 or even both, the calculation uses the original formula. So if

- $t - \mathit{transfer}(f_1, tr) < \mathit{lfa}(tr, f_1, s)$ and
- $t - \mathit{transfer}(f_2, tr) < \mathit{lfa}(tr, f_2, s)$

hold true for the minute t the extended formula

$$
\begin{aligned}
&P_{noWaitingForFeeders}\big(tr, s, t\big) = \\
&\quad \prod_{f \in FD(tr,s) \setminus \{f_1, f_2\}} \big[P\big(X_{arr_{tr,s}} > \mathit{lfa}(tr, f, s)\big) \\
&\qquad\qquad + P\big(X_{arr_{tr,s}} \le \mathit{min}\{t - \mathit{transfer}(f, tr), \mathit{lfa}(tr, f, s)\}\big)\big] \\
&\quad \cdot \big[P\big(X_{arr_{f_1,s}} \le t - \mathit{transfer}(f_1, tr) \cap X_{arr_{f_2,s}} \le t - \mathit{transfer}(f_2, tr)\big) \\
&\qquad + P\big(X_{arr_{f_1,s}} > \mathit{lfa}(tr, f_1, s) \cap X_{arr_{f_2,s}} \le t - \mathit{transfer}(f_2, tr)\big) \\
&\qquad + P\big(X_{arr_{f_1,s}} \le t - \mathit{transfer}(f_1, tr) \cap X_{arr_{f_2,s}} > \mathit{lfa}(tr, f_2, s)\big) \\
&\qquad + P\big(X_{arr_{f_1,s}} > \mathit{lfa}(tr, f_1, s) \cap X_{arr_{f_2,s}} > \mathit{lfa}(tr, f_2, s)\big)\big]
\end{aligned}
$$

is used.

3.2.2.2 Probability of Waiting for Feeders

The formula for $P_{waitingForFeeders}$ in [KSWZ12, A.2] again needs a special treatment of f_1 and f_2.

Note that the introduction of $Y_{arr_{f,s}}$ is an optimization and with the following equation it is not needed to use it:

$$P\big(Y_{arr_{f,s}} > t\big) = P\big(t - \mathit{transfer}(f, tr) < X_{arr_{f,s}} \le \mathit{lfa}(tr, f, s)\big).$$

While it is useful to introduce $Y_{arr_{f,s}}$ for the one dimensional case it is not possible to use it for the two dimensional calculation, because some regions of the distribution where one time is after the latest feasible arrival, while the other is not are needed.

$$
\begin{aligned}
&P_{waitingForFeeders}\big(tr, s, t)\big) = \\
&\quad \prod_{f \in FD(tr,s) \setminus \{f_1, f_2\}} \big[1 - P\big(Y_{arr_{f,s}} > t\big)\big] \\
&\qquad \cdot \big[1 - P\big(t - \mathit{transfer}(f_1, tr) < X_{arr_{f_1,s}} \le \mathit{lfa}(tr, f_1, s) \\
&\qquad\qquad \cup\, t - \mathit{transfer}(f_2, tr) < X_{arr_{f_2,s}} \le \mathit{lfa}(tr, f_2, s)\big)\big] \\
&\quad - \prod_{f \in FD(tr,s) \setminus \{f_1, f_2\}} \big[1 - P\big(Y_{arr_{f,s}} > t - 1\big)\big] \\
&\qquad \cdot \big[1 - P\big(t - \mathit{transfer}(f_1, tr) - 1 < X_{arr_{f_1,s}} \le \mathit{lfa}(tr, f_1, s) \\
&\qquad\qquad \cup\, t - \mathit{transfer}(f_2, tr) - 1 < X_{arr_{f_2,s}} \le \mathit{lfa}(tr, f_2, s)\big)\big]
\end{aligned}
$$

Chapter 4

Computational Study

4.1 Setup

To analyze the difference of the dependent calculation against an independent calculation we used the multi-criteria timetable-information system MOTIS. The time interval γ, which is used to determine the set of feeders of a train is set to $\gamma = 30$ minutes.

We used time-expanded graphs of several days as used by MOTIS, which were introduced in the work of Schulz, Wagner and Weihe in [SWW00]. The graphs are always for two days in order to support overnight connections. The data is from the German railway network. We used data from a week in November 2012 (20.-26.), a week in February 2013 (11.-17.) and a week in April 2013 (10.-16.). The data sets of February and April are mostly the same, with graph sizes between 2M arrival and departure nodes, 0.9M train edges and 66K trains on the Saturday/Sunday data set, 2.2M nodes, 1M edges and 76K trains on the Friday/Saturday and Sunday/Monday data set and 2.6M nodes, 1.3M edges and 86K on other weekday graphs. The November data set is a little bit smaller since mid-December there was an update of the timetable. There are between 250K and 400K less event nodes, between 50K and 200K less train edges and about 5K less trains in each two-day graph.

4.2 Found Structures

On each two-day graph we searched for structures, where the departure of f_1 and f_2 is on the first day of the two-day graph. The average amount of structures found, along with the amount of connecting trains (in several cases there were multiple trains tr_2 depending on the two dependent feeders) and the amount of potentially affected nodes is given in Table 4.1. The table also gives a percentage of the potentially affected nodes of the whole two-day graph. The actual percentage of affected nodes can be estimated to be about twice as much, since we only search for structures on the first day (which may affect nodes of the second day, though). The minimum amount of structures found on a single weekday is 126 (40 on a weekend day), the maximum is 179 (74 on a weekend day). There are between 285 and 451 connecting trains (between 76 and 184 on weekends). More than 50% of the found structures are at only 5 different

Days	Number of structures	Number of connecting trains	Number of potentially affected nodes
Nov. Mo-Fr	135	307	43198 (1.96%)
Nov. Sa-So	47	93	24293 (1.43%)
Feb. Mo-Fr	156	402	48108 (1.85%)
Feb. Sa-So	47	104	25492 (1.27%)
Apr. Mo-Fr	163	413	49112 (1.89%)
Apr. Sa-So	66	158	28376 (1.42%)
Overall Avg.	123	301	40877 (1.75%)

Table 4.1: Average number of found structures, connecting trains and potentially affected nodes per two day graph.

station-pairs (departure and arrival station of f_1 and f_2).

Potentially affected nodes are nodes, which are connected to a departure node of any of the connecting trains by a chain of train and feeder edges. As time passes, the delay of a train at one station is not important anymore. One reason for this decay of influence are buffer times, which can be used to catch up delays and another reason are other feeders, the trains afterwards have to wait for at following stations. So we chose a time limit of 240 minutes after which with a very high probability any departure will not affect other nodes anymore.

While searching, we allowed structures which do not only depend on one common feeder, but may also depend on multiple feeders. We did this to find more structures to analyze the difference of the dependent feeder calculation on. While the calculation then may not be entirely correct, in these cases we based the dependent feeder calculation on the feeder with the latest arrival time. The calculation is then still more exact than an independent calculation and allows a comparison. As explained in Section 2.1 we only include cases where tr_1, f_1, f_2 and tr_2 are pair-wise different. There are 10 to 30 additional structures per day where tr_1 or tr_2 is the same as f_1 or f_2.

4.3 Results

4.3.1 Two Dimensional Distributions

First, we checked whether the specific arrival distributions of the common feeders actually have an effect on the departure probability or not. To make a difference the probability that both feeders have to wait for the common feeder has to be greater than zero. So, in turn the probability that the common feeder arrives with a delay, which is large enough, has to be greater than zero.

$$P\big(X_{arr_{tr_1,s}} > max(sched(dep_{f_1,s}) - transfer(tr_1, f_1), sched(dep_{f_2,s}) - transfer(tr_1, f_2))\big) > 0$$

Note that a delay which is large enough to make the later train wait is still in the waiting interval of the earlier train because otherwise we would not have found this structure in the first place.

Using this, we could already say that for 24% to 55% of the structures the dependent calculation differ from the independent calculation. The amount of

Days	Affected structures	Avg. Dep. diff.	Max. Dep. diff.	Avg. Arr. diff.	Max. Arr. diff.
Nov. Mo-Fr	89	3.11%	27.52%	1.87%	17.38%
Nov. Sa-So	22	2.58%	11.28%	1.44%	8.60%
Feb. Mo-Fr	112	6.74%	52.49%	5.10%	44.49%
Feb. Sa-So	30	3.06%	12.78%	1.52%	8.89%
Apr. Mo-Fr	117	7.10%	54.11%	5.25%	44.48%
Apr. Sa-So	50	4.01%	23.68%	1.66%	17.18%
Overall Avg.	85	4.95%		3.35%	

Table 4.2: Number of structures with differences and the average and maximum of the sums of the absolute values of differences of the two dimensional distributions.

structures left each day (on average) can be seen in Table 4.2. The remaining columns are about the differences in the two dimensional departure and arrival distributions compared to the independent distributions. As metric we used the sum of the absolute value of the differences in the probability for each pair of delay minutes:

$$diff := \sum_{t_1=lb_1}^{ub_1} \sum_{t_2=lb_2}^{ub_2} \Big| P_{dependent}\big(X_{event_{f_1,s}} = t_1 \cap X_{event_{f_2,s}} = t_2\big) - P_{independent}\big(X_{event_{f_1,s}} = t_1\big) \cdot P_{independent}\big(X_{event_{f_2,s}} = t_2\big) \Big|$$

where lb and ub are the lower and upper bounds of the respective distributions of $X_{event_{f_1,s}}$ and $X_{event_{f_2,s}}$.

Some of the calculated two dimensional distribution have notable differences to the one dimensional distributions. Departure distributions have a difference of up to 54.11%, while the average difference is only between 2.58% and 7.10%. So we have some cases where the common feeder arrives shortly before the dependent trains and influences their departure by a great amount. On the other hand we have lots of occurrences with the feeder only influencing the trains with its maximum delay minutes by minimal amounts. We can see a difference on weekdays between the November week and the February/April weeks, which is due to a timetable update mentioned before.

Calculating the dependent arrival distribution using the independent travel time distributions reduced the difference. For the evaluation we mainly used the generated travel time distributions. They aren't as wide as the distributions from Deutsche Bahn AG, due to which more structures and connecting trains are labeled as not affected by us. This is no problem, because these wide distributions got only very small probabilities for the later minutes. In fact, the distributions from Deutsche Bahn AG lead to even smaller average and maximum differences. The maximum difference here is 44.49% with averages between 1.44% and 5.25%. The average differences are 25% to 58% smaller then the average differences of the departure distributions.

Figures A.1, A.2, A.3 and A.4 in the Appendix show exemplary comparisons of arrival distributions with rather huge differences (the maximum differences in the weeks of November and April and some other ones of November and

Days	Affected departures	Avg. relevant Arr. diff.	Max. relevant Arr. diff.	Avg. Dep. diff.	Max. Dep. diff.
Nov. Mo-Fr	76	0.13%	6.37%	< 0.01%	0.07%
Nov. Sa-So	28	0.11%	0.30%	< 0.01%	0.13%
Feb. Mo-Fr	101	0.12%	2.72%	< 0.01%	0.03%
Feb. Sa-So	37	0.09%	1.67%	< 0.01%	< 0.01%
Apr. Mo-Fr	122	0.20%	3.71%	0.02%	1.30%
Apr. Sa-So	85	0.09%	2.32%	0.02%	0.79%
Overall Avg.	85	0.13%		< 0.01%	

Table 4.3: Number of structures with differences and the average and maximum of the sums of the absolute values of differences of the two dimensional distributions.

February). Most of the differences are in the first delay minutes of the trains, and the later delay minutes do not have any noticeable difference. Also, the differences are mostly split up to several of the small minutes with no huge difference at one point. The dependent plots usually have a noticeable diagonal, which is the result of the common feeder. These observations hold true for all other arrival distributions.

4.3.2 Departure Distribution

As with the calculation of the two dimensional departure distribution, we first checked whether the specific arrival distributions of the dependent feeders make a difference whether we use our dependent calculation. To make a difference the probability that the connecting train has to wait for both of them at the same time has to be greater than zero.

$$P\big(X_{arr_{f_1},s} > sched(dep_{tr_2,s}) - transfer(f_1, tr_2)\big) > 0$$
$$P\big(X_{arr_{f_2},s} > sched(dep_{tr_2,s}) - transfer(f_2, tr_2)\big) > 0$$

After this check, and the previous filtering of the structures, only 24% to 53% of the departures of the connecting trains remain. Those are actually affected by the dependent feeder calculation.

Table 4.3 shows the average and maximum sums of the absolute values of differences in the departure distribution. The average difference is lower than 0.02% with a maximum of 1.30%. The reason for this small difference in spite the fact that some of the arrival distributions have large differences is, that in all found cases the portion of the difference sum in the arrival distribution of the dependent feeders, which actually has an effect on the departure or tr_2 is a lot smaller then the total difference. The parts of the distribution, which have an influence on the departure distribution of tr_2 are those, where tr_2 has to wait for one of the two dependent feeders:

$$relevant\_diff := \sum_{t_1=lb_1}^{ub_1} \sum_{t_2=sched(dep_{tr_2,s})-transfer(f_2,tr_2)+1}^{ub_2} \Big| P_{dependent}\big(X_{arr_{f_1,s}} = t_1 \cap X_{arr_{f_2,s}} = t_2\big) - P_{independent}\big(X_{arr_{f_1,s}} = t_1\big) \cdot P_{independent}\big(X_{arr_{f_2,s}} = t_2\big)\Big| + \sum_{t_1=sched(dep_{tr_2,s})-transfer(f_1,tr_2)+1}^{ub_1} \sum_{lb_2}^{sched(dep_{tr_2,s})-transfer(f_2,tr_2)} \Big| P_{dependent}\big(X_{arr_{f_1,s}} = t_1 \cap X_{arr_{f_2,s}} = t_2\big) - P_{independent}\big(X_{arr_{f_1,s}} = t_1\big) \cdot P_{independent}\big(X_{arr_{f_2,s}} = t_2\big)\Big|$$

The average and maximum differences in the relevant parts of the arrival distributions can be found in Table 4.3, too. These differences, influence the result of the early delay minutes of the connecting train by small amounts.

4.4 Evaluation

The results show that in real world timetables the error that is introduced by the independence assumption for all detected dependent structures is actually really small. The reason for this is, that often the common feeder tr_1 does not influence one or both of the dependent feeders f_1 and f_2 all that much, because it arrives a lot earlier and only influences the later delay minutes of them by minimal amounts. In the cases where there is a notable difference in the departure distribution, and afterwards in the arrival distribution, in no detected structure f_1 and f_2 had then a huge influence on the connecting train tr_2.

4.4.1 Discussion

A important issue to mention are rounding errors. The implementation uses 32-bit floating point numbers. Since the calculation becomes rather complex, and the application of the travel time distributions is a huge sum over the two dimensional distribution which consists mostly of rather small probabilities, except some larger ones near the small delay minutes, rounding errors become an issue. For the probabilities with small differences we actually have the problem, that we cannot be sure, how much of this difference is due to the dependent calculation and how much is due to rounding errors. The same applies than to the resulting departure distribution of the connecting train tr_2. The differences in the independent and the dependent departure distributions are all so small, that we cannot say how huge the exact difference is. Here it is no problem though, since the sum of the absolute values of the differences could be two or three times as high and our result, that in our cases there is no significant difference would not change.

While it is true, that in the timetable we used we could not find a single case with a notable difference, the timetable of other organizations could include some. It is unlikely though, that there will be more than a handful. Also this

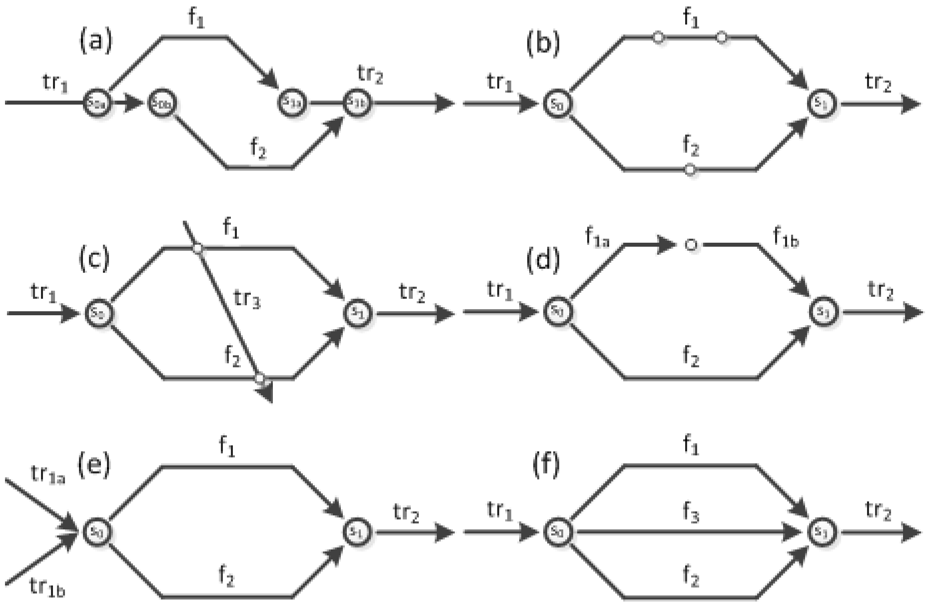


Figure 4.1: Variations of the basic structure.

knowledge could be used to watch out for these cases during creation of the timetable.

Another problem with the analysis is that we use a very specific case, but that does not matter in the overall picture. One reason for this is that railway networks are mostly homogeneous. So in all likelihood different structures have the same properties with regards to the differences between a dependent and an independent calculation as the basic structure we researched. Though in the following we will still have a look at different structures.

First, there could be additional stops at any point. The dependent feeders f_1 and f_2 could meet the common feeder tr_1 at different stops, they could have several stops on the way to meet the following train tr_2 and they do not have to meet tr_2 at the same station (Figure 4.1 a and b). Now the question is, what the additional stops mean for the faulty independent calculation. Most stops have buffer times used to catch up delays. Also the trains may have to wait for other independent trains at any stop. All of these things are in favor of the independent calculation, since the probability distribution is always modified for only one of the dependent feeders and not both of them in common. Beware that here something might happen, that increases the dependency. Another train tr_3 could strengthen the dependency if it is a feeder to both of the dependent feeders at one of their stops (Figure 4.1 c). Though, the loss of the original dependency due to the extra stop should mostly outweigh the new dependency.

Secondly, the dependency could transfer over several trains. Let for example one of the dependent feeders f_1 be not one train, but two trains f_{1a} and f_{1b}. f_{1a} depends together with f_2 on the common feeder tr_1 and f_{1b} depends on f_{1a}. This case is similar to adding a stop in the way of f_1, just that the train is changed, too (Figure 4.1 d). It is even less important here, because if f_{1b} does not have to wait for f_{1a}, either because it is on time, or because it is too late according to the waiting time policy, the dependency is lost completely in one stop. Dependency structures can get far more complex with this method, but the dependency will not become stronger because of this.

Then, there could be several common feeders tr_1, which the dependent feeders f_1 and f_2 depend on (Figure 4.1 e). Also there could be more than two dependent feeders f_1 and f_2 (Figure 4.1 f).

The former case usually won't be a problem. If there are a lot of common feeders, our results show, that mostly the latest one matters. It could only become a problem if there are several common feeders arriving at nearly the same time, all in a time frame which actually influences the departure of the dependent trains by a lot. In the structures we found in the schedule of the German railway network, the case that there are multiple common feeders which influence the dependent trains at all happens only a few times per day. Among these occurrences there are even fewer with several feeders which arrive shortly before the dependent trains, so that more than one could influence the dependent trains by more than a tiny amount. The maximum amount of occurrences is on one week day in April, with eleven structures with more than one feeder which can delay both trains with more than a 25% probability.

The latter in turn would usually appear in complex structures, which are, as already mentioned, no problem. With more than two dependent feeders another issue comes up. For each dependent feeder a dimension is needed in the distributions. So while two dependent feeders result in two dimensional distributions, three for example would result in three dimensional distributions. The calculation of such cases would consume a lot of time and RAM. If a case with multiple dependent feeders would appear in a small structure the trains would actually have to share a track on their way, which leads to the last problem with which we currently cannot cope.

Travel time distributions are still assumed to be independent. This might actually be a huge error, because several trains need to use the same track, so one has to wait for the other if they both want to use it in a short time frame. Also if there is a problem on one track, all trains would suffer common delays because of it. The problem with this is, that currently there is no data available to base a dependent calculation upon.

Chapter 5

Conclusion and Future Work

In this work we looked at the independence assumption which is used in probabilistic approaches for estimating delays in railway networks. We analyzed a simple dependency structure, where the independence assumption can possibly introduce a large error.

Then, we modified the formulas used in the work of Keyhani et al. [KSWZ12], which is about reliability ratings of train connections, to support a dependent calculation for the chosen structure. Our analysis, with the timetable of the German railway network, showed, that for all analyzed structures the departure distribution of connecting trains is not significantly different to the departure distribution calculated without the independence assumption. Though, the arrival distribution of the dependent feeders had noteworthy differences in some cases, but only in the early delay minutes which did not influence the connecting train.

Since the calculation of the dependent feeders is rather complicated and computationally intensive and the error is small, it is not worth to use it for this structure and timetable. We argued, that it is even more unlikely for cases with a huge error to happen in more complex structures, but it is possible. Future work could analyze more complex structures in more detail. Another important issue would be use dependent travel time distributions. Since several feeders come from the same direction or even use the same track, the error of using independent travel time distributions could be important.

From a different perspective future work could try to avoid critical structures during timetable creation. Since usually only small delays have high probabilities it would be possible to identify these critical structures, which would require, that a common feeder arrives shortly before some dependent feeders, which in turn all arrive shortly before the connecting train. Since there aren't too many of those critical structures (at least if our argumentation for more complex structures holds true) the timetable wouldn't require too many changes. If none of these critical structures make it into the final timetable, algorithms, which rate connections with an independence assumption, could achieve the same result as algorithms not using the independence assumption with a high probability.

Chapter A

Arrival Distribution Plots

Figure A.1: A arrival distribution of the November week. Difference 17.38%

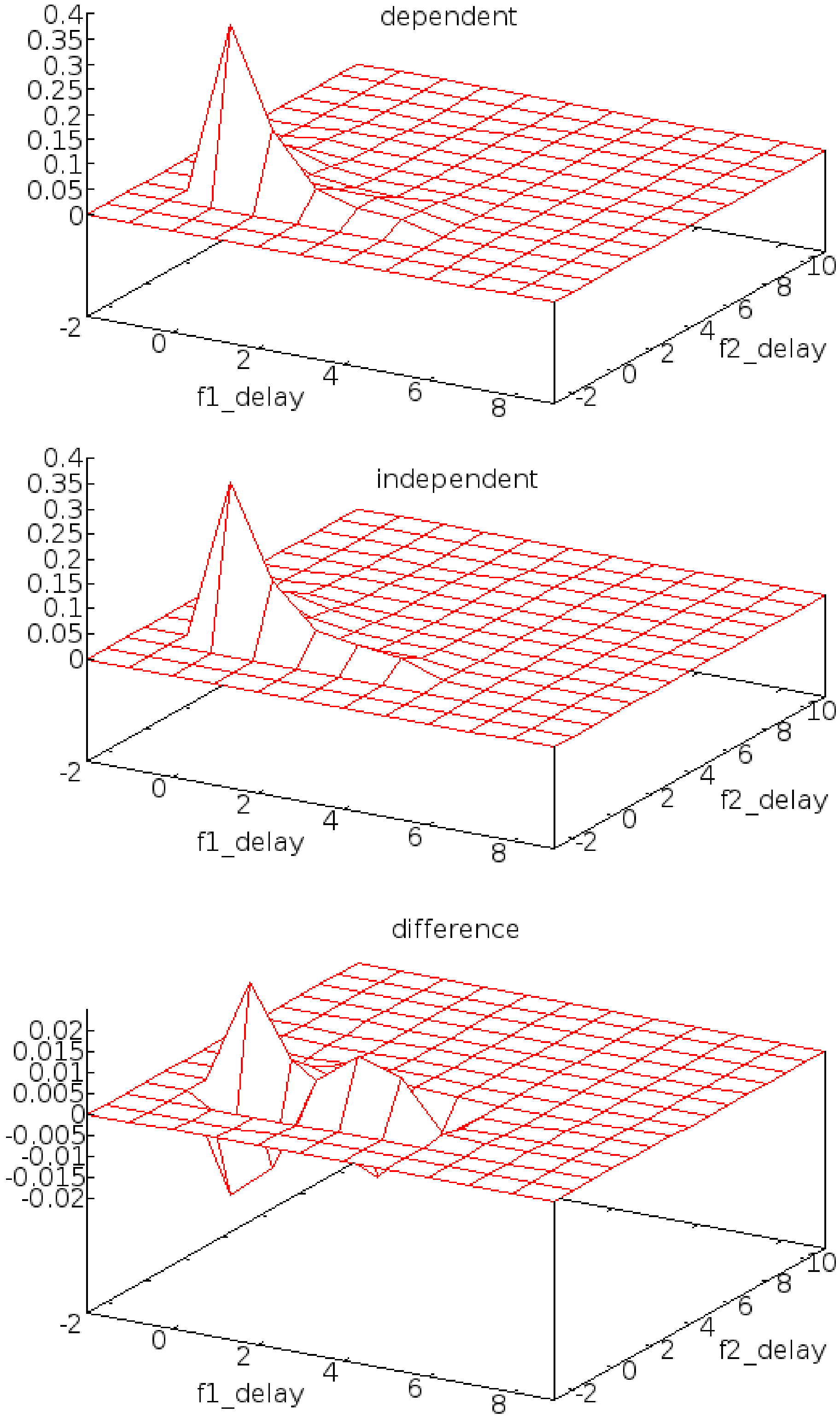


Figure A.2: A arrival distribution of the April week. Difference 44.48%

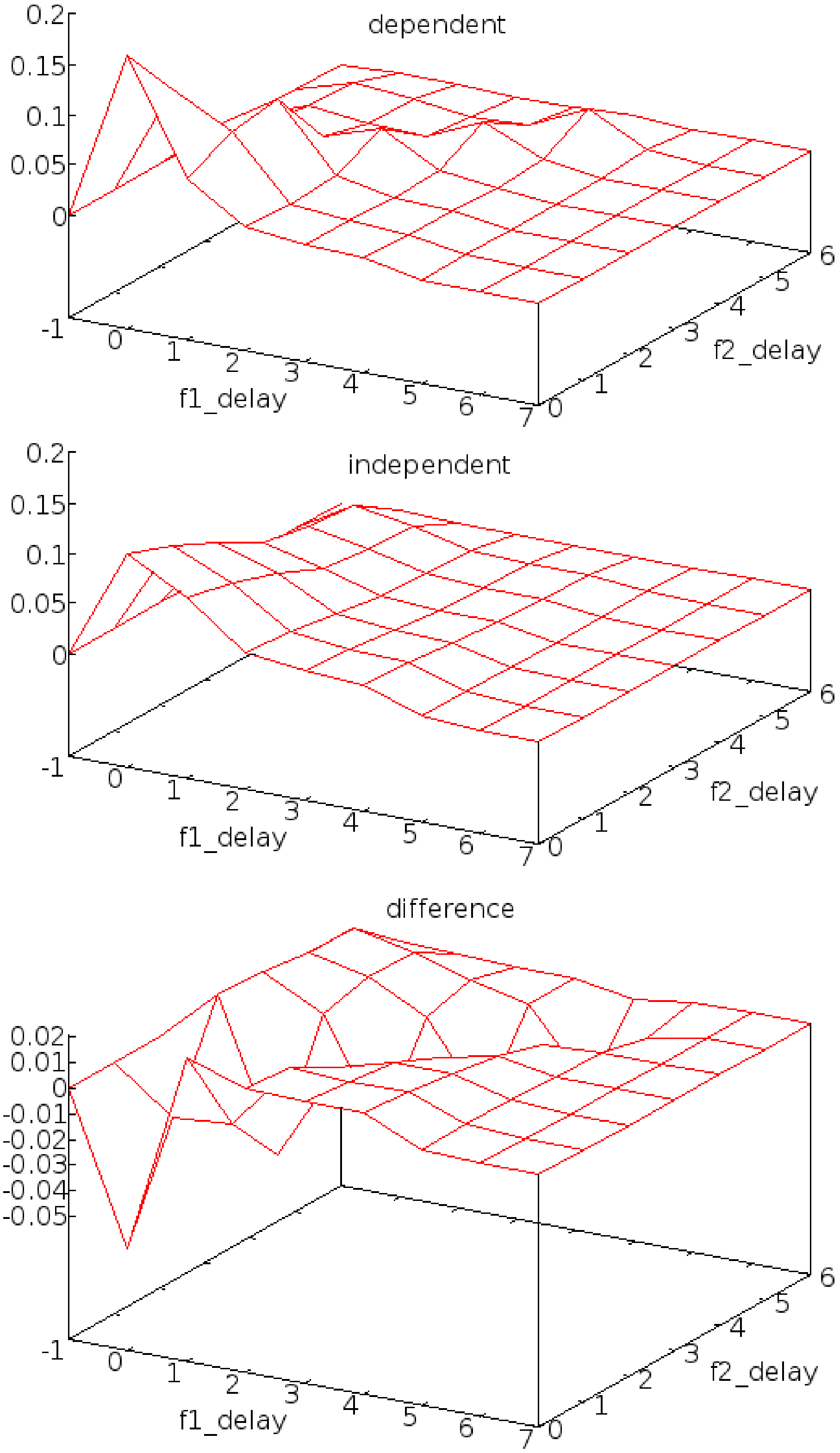


Figure A.3: A arrival distribution of the November week. Difference 13.70%

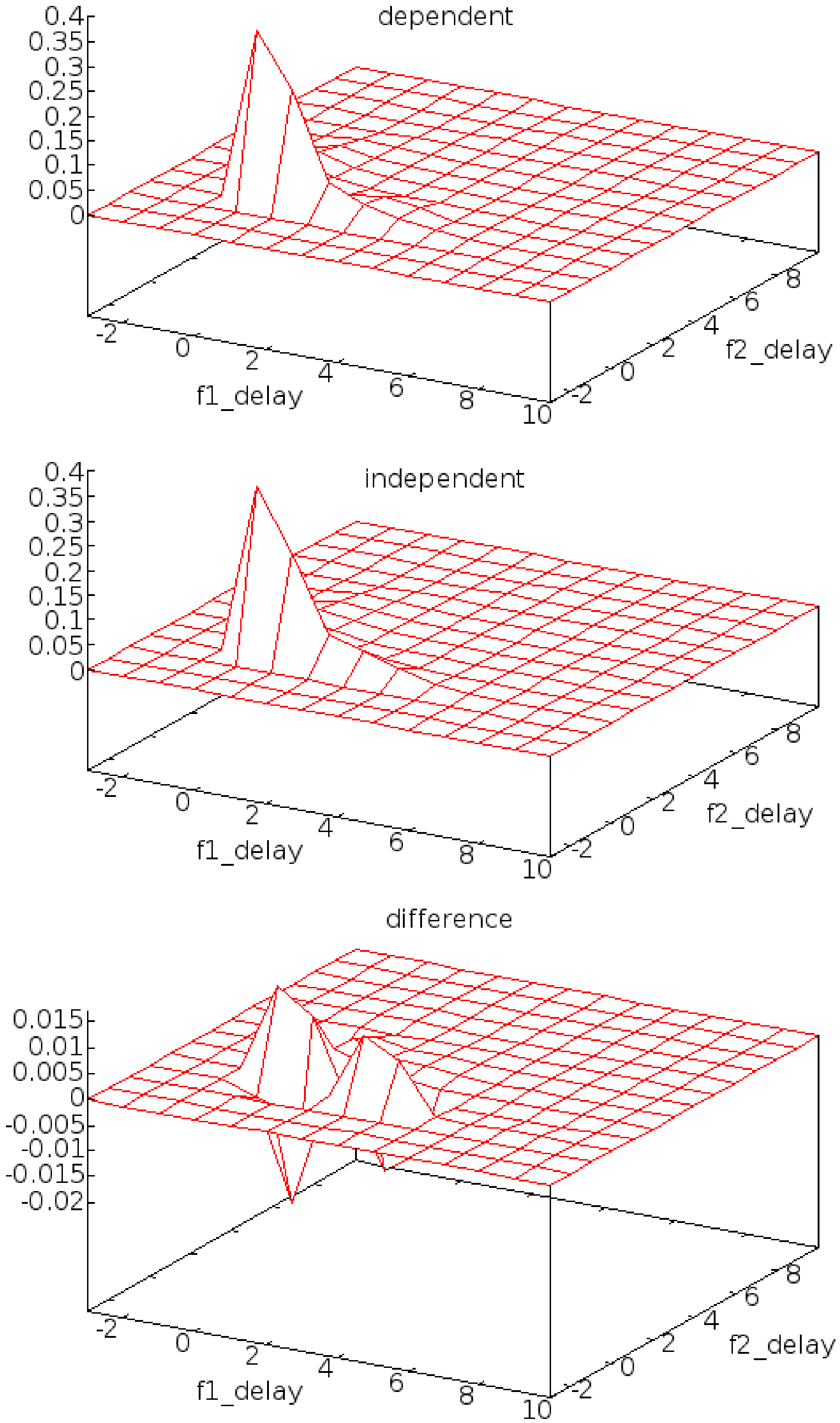
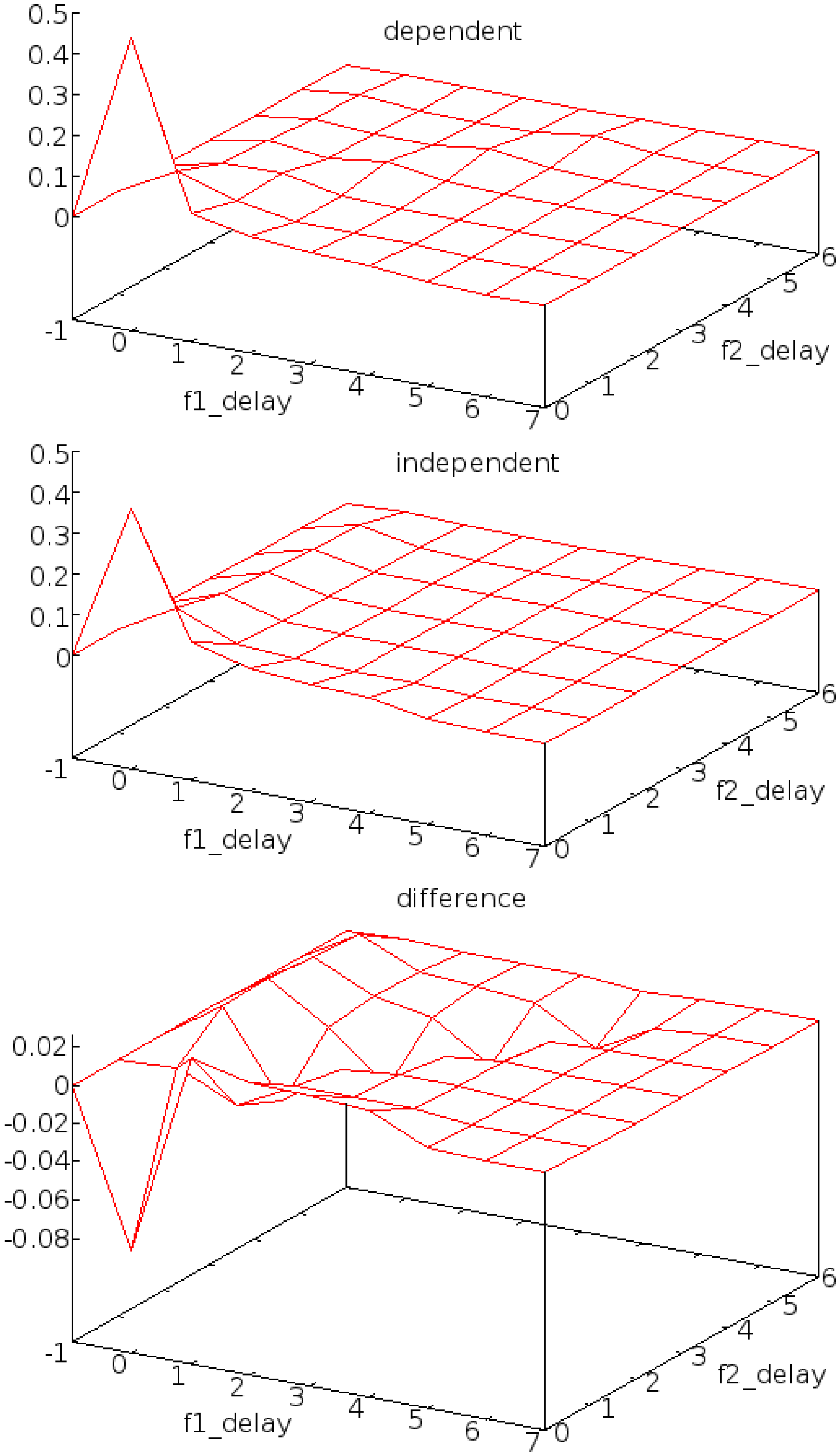


Figure A.4: A arrival distribution of the February week. Difference 37.39%



Bibliography

[BGMHO11] Annabell Berger, Andreas Gebhardt, Matthias Müller-Hannemann, and Martin Ostrowski. “Stochastic Delay Prediction in Large Train Networks”. In: *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*. Ed. by Alberto Caprara and Spyros Kontogiannis. Vol. 20. OASIcs. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2011, pp. 100–111.

[CK94] Malachy Carey and Andrzej Kwieciński. “Stochastic approximation to the effects of headways on knock-on delays of trains”. In: *Transportation Research Part B: Methodological* 28.4 (1994), pp. 251 –267.

[KSWZ12] Mohammad H. Keyhani, Mathias Schnee, Karsten Weihe, and Hans-Peter Zorn. “Reliability and Delay Distributions of Train Connections”. In: *12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*. Ed. by Daniel Delling and Leo Liberti. Vol. 25. OASIcs. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2012, pp. 35–46.

[MM07] Ludolf E. Meester and Sander Muns. “Stochastic delay propagation in railway networks and phase-type distributions”. In: *Transportation Research Part B: Methodological* 41.2 (2007), pp. 218 –230.

[Sch09] Mathias Schnee. “Fully Realistic Multi-Criteria Timetable Information Systems”. PhD thesis. TU Darmstadt, 2009.

[SWW00] Frank Schulz, Dorothea Wagner, and Karsten Weihe. “Dijkstra’s algorithm on-line: an empirical case study from public railroad transport”. In: *J. Exp. Algorithmics* 5 (Dec. 2000).