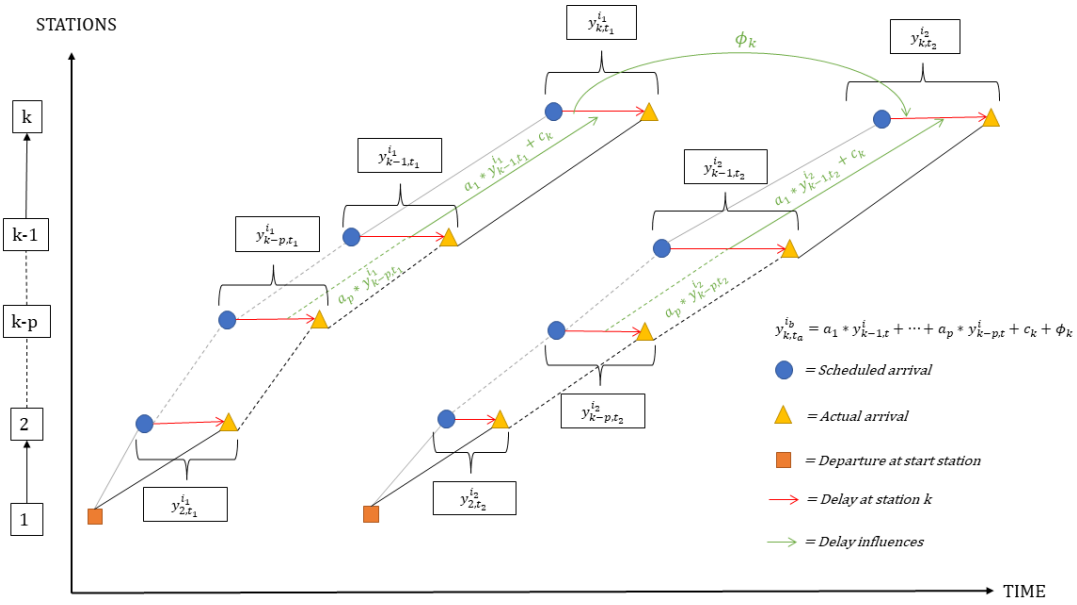


Predicting Train Delays by means of Time Series Analysis

Fiona Arnet
Supervisor: Prof. Dr. Francesco Corman
Supervisor: Thomas Spanninger

Bachelor's Thesis

June 2022

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Predicting Train Delays by means of Time Series Analysis

Fiona Arnet
ETH Zürich
CH-8093 Zurich
arnetfi@student.ethz.ch

Supervisor: Thomas Spanninger
Institute for Transport Planning and Systems
ETH Zürich

Supervisor: Prof. Dr. Francesco Corman
Institute for Transport Planning and Systems
ETH Zürich

June 2022

Abstract

Train delays are a challenging factor in traffic operations. Using time series analysis and more specifically auto-regressive processes to predict train delays and thereby facilitating traffic control is a promising approach. On the section of the S2 service travelling between Ziegelbrücke and Zürich HB, a simple linear regression manages to improve prediction accuracy already significantly by an average of 65% compared to a baseline approach assuming that delay stays constant between two stations. Only for few stations, the simple linear regression approach does not provide convenient results. One of them is the station of Zürich Enge, where the simple linear regression improves prediction accuracy by 21% only. By accounting for knock-on effects between trains in the model, prediction accuracy at this specific could be improved by 55% compared to the baseline approach. While the results are already convenient, the model does not tap its full potential and leaves room for further investigation and improvement, promising even better prediction results.

Keywords

train delay prediction, knock-on delays, time series analysis

Contents

List of Tables

List of Figures

1 Introduction

Public transport networks are crucial for mobility provision in a country. They grant basic mobility services to everybody including people who do not have access to a car and form the backbone of sustainable transport. In Switzerland, rail-traffic is the most important mode of public transport. Between the year 2000 and the year 2018, passenger-kilometres almost doubled (Walker *et al.* (2018)). Attractiveness of rail-traffic in Switzerland is corroborated by drawing a comparison with other European countries: Switzerland is the country with most railway-kilometres per capita (Walker *et al.* (2018)). Enhancing quality of rail-traffic is desirable because rail-traffic is the most environmentally sound mode of transport apart from non-motorised traffic (SBB (2022a)) and therefore an auspicious and forward-looking transport solution. Providing high-quality rail-traffic is determined by two main criteria: punctuality and safety (LITRA (2022)). Ensuring punctuality in the Swiss railway network is a complicated undertaking as the Swiss railway network is one of the most densely used in the whole world (SBB (2022b)). While rail-traffic can be very convenient due to its density, the far side is that already small train delays can have far-reaching effects in the network. Train delays therefore are a crucial problem in public transport operations as they not only reduce passenger comfort but also might propagate through the network and thereby complicate traffic control and reduce service quality. Train delay prediction is an important tool to facilitate traffic control and make passenger information more convenient. Especially real time train delay prediction is an active area of research. Spanninger *et al.* (2020) provide a summary of the most important publications and approaches. Two papers presented by Spanninger *et al.* (2020) predict train delays by means of time series analysis: Pongnumkul *et al.* (2014) using a moving average approach and Wang and Work (2015) using vector auto-regressive processes. This thesis builds on the model developed by Wang and Work (2015) aiming at refining it and applying it to a Swiss data set. Both approaches are briefly described in the following paragraphs.

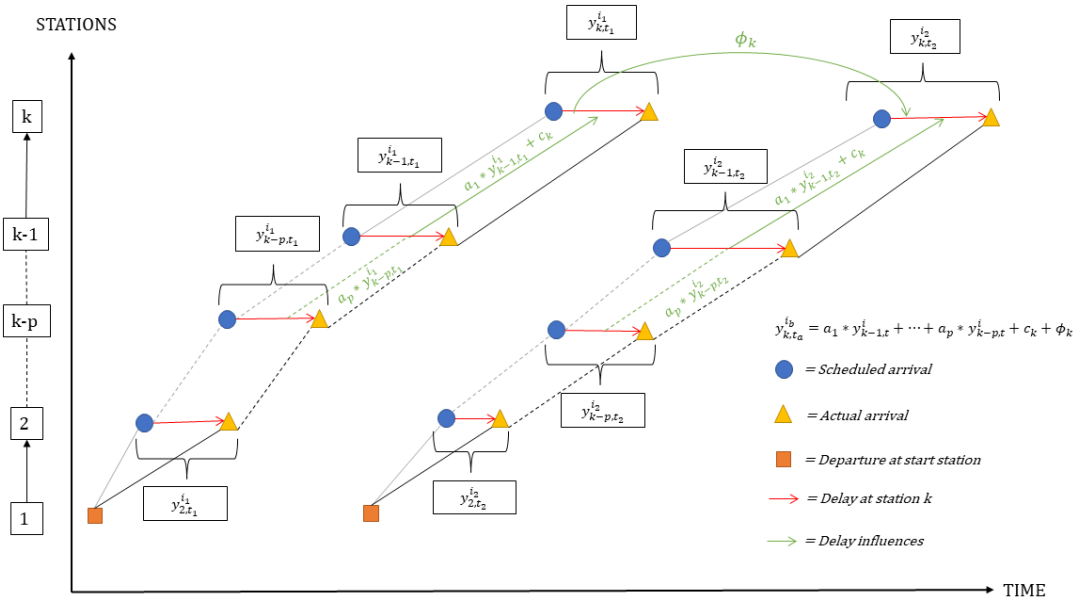
Pongnumkul *et al.* (2014) compare moving average delay predictions with a k-nearest neighbours algorithm whereat the latter is a clustering instead of a time series analysis approach. It will not be discussed here any further. To predict the delay at station k by means of a moving average, Pongnumkul *et al.* (2014) use historical travel times recorded on section s leading to station k. The historical travel times on this section are averaged and added to the known delay at station $k-1$ of train i on its current trip t to predict the train's arrival at station k. The predicted results are compared to a baseline algorithm which assumes that the delay stays constant between station $k-1$ and station k. The moving average prediction increases prediction accuracy by approximately 18.5%. While

prediction accuracy is increased and the modelling approach is easy to implement and understand, it does not leave room for improvement or a more complex modelling of the phenomenon of delay propagation but rather seems to tap the full potential already.

Wang and Work (2015) use auto-regressive processes to predict train delays. They build two different models. The first one, a historical regression model, allows to estimate the delay at each station k of a trip prior to the departure of the train based on historical delay data. The second one is an online auto-regressive model that allows for real time train delay prediction. The online auto-regressive model comes in three complexity levels. Complexity level one corresponds to the baseline approach of Pongnumkul *et al.* (2014). It assumes the delay to stay constant between station $k-1$ and station k. A second level approach models the delay of train i at station k as a linear regression of the known delays at the p previous stations on the current trip t of train i. On a third level, the auto-regressive model further accounts for knock-on delays. A knock-on effect is "something (such as a process, action, or event) that causes other things to happen" (Webster (2022)). In train operations, knock-on effects occur because the delay of one train might influence the delay of another train in case trains share the same infrastructure. A graphical representation of the model in its third complexity level is shown in figure 1. The exact mathematics of all three levels are described in more detail in section 2. The performance of the prediction results presented by Wang and Work (2015) are compared to making no prediction at all, i.e. assuming that the train arrives at station k as scheduled. The historic regression model improves prediction accuracy by 18%, the first level online regression model by 57%, the second level online regression model by 60% and the third level online regression model also by approximately 60%. While the historic regression model seems to tap the full potential, the online regression model promises to deliver better results if developed further. Wang and Work (2015) themselves suggest several possibilities for advancement, including investigation of knock-on delays, separate modelling of factors influencing delays such as track geometry and weather or the use of other data driven approaches than linear regressions.

The focus of this thesis is put on investigating and modelling knock-on delays more precisely. Wang and Work (2015) did not manage to improve prediction accuracy any further by including knock-on delays in their model but their modelling approach was based on rather far-reaching assumptions (section 2). Investigating knock-on delays more thoroughly therefore promises to improve prediction accuracy. Eventually,this thesis aims at answering two questions:

Figure 1: Schematic representation of the online regression model



- How does the model of Wang and Work (2015) perform on a Swiss data set?
- How can knock-on delays be modelled more accurately and how does it affect the prediction accuracy of the model?

Section 2 describes the different models that were implemented and their parameters in more detail. The implementation of the models and the performance of the models on a Swiss data set are presented in section 3 whereupon section 4 discusses them. Finally, section 5 presents the conclusions.

2 Methodology

Generally speaking, all models that were implemented in the scope of this thesis can be classified according to the same input-output relation. The input is given by actual arrival delays measured along the route of a train. These arrival delays are used as input data for an auto-regressive model as it is commonly used in time series analysis. The output is a deterministic prediction of the arrival delay at station k. The model is designed to make real-time train delay predictions with a prediction horizon of one station in advance. Possible model adaptions that could make predictions further into the future are discussed in section 4 but not implemented as part of this thesis.

This thesis implements one baseline approach and three models that deterministically predict the delay of train i on trip t at station k. The absolute error associated with the model is calculated as

$$u_{k,t} = |y^i_{k,t} - \hat{y}^i_{k,t}| \tag{1}$$

$u_{k,t}$ = absolute error associated with the model
$y^i_{k,t}$ = actual delay of train i on trip t at station k
$\hat{y}^i_{k,t}$ = predicted delay of train i on trip t at station k

The baseline approach and the three models that were implemented are discussed in more detail in the following sections. The baseline approach and the model described in section 2.2 are one-to-one taken from the publication of Wang and Work (2015). The mathematics of the interacting model described in section 2.3 are taken from Wang and Work (2015), however, modelling conditions are adapted. The binary model described in the same section is a further development of the approaches described by Wang and Work (2015) and presents a new approach that has not been implemented by Wang and Work (2015).

2.1 The Baseline Approach

The baseline approach is based on the assumption that train delays do neither increase nor decrease between the stops on a train route. Hence, if train i delayed by 5 minutes at station $k-1$, the baseline approach suggests that this delay remains constant resulting in

a 5 minutes delay at station k as well. The estimated delay at station k is consequently given as:

$$\hat{y}^i_{k,t} = y^i_{k-1,t} \tag{2}$$

$\hat{y}^i_{k,t}$: predicted delay of train i on trip t at station k
$y^i_{k-1,t}$: delay of train i on trip t at station $k-1$

No parameters need to be chosen for this approach. All models implemented will be compared to the benchmark of this baseline approach.

2.2 The Standard Model

The standard model predicts the delay of train i at station k by means of an auto-regressive process based on the actual delay of train i at the p previous stations $k-1, ..., k-p$ on its current trip t. The model was developed by Wang and Work (2015) and implemented according to their definition:

$$\hat{y}^i_{k,t} = a_1 y^i_{k-1,t} + \cdots + a_p y^i_{k-p,t} + c_k \tag{3}$$

$\hat{y}^i_{k,t}$: predicted delay of train i on trip t at station k
$y^i_{k-1,t}$: actual delay of train i on trip t at station $k-1$
$y^i_{k-p,t}$: actual delay of train i on trip t at station $k-p$
$a_1, ..., a_p$: factors accounting for relationship of train delays among current and past stations
c_k : intercept term which allows for constant delays on a section

p is a control variable that indicates the number of delays measured at previous stations that should be included in the model. It is the only parameter value that needs to be chosen prior to training the model. The model then trains the parameters $a_1, \dots, a_p$ and c_k for every station k along the route of train i individually.

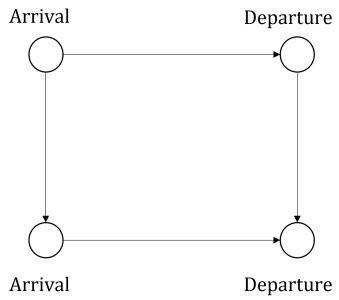
This model does not include possible interactions with other trains using the same infrastructure but rather models the delay at some station k as weighted sum of the actual delays at the p previous stations on trip t and a constant parameter c_k that accounts for the fact that there might be a constant component associated with the delay at station k.

2.3 Modelling Knock-On Delays: The Interacting and the Binary Model

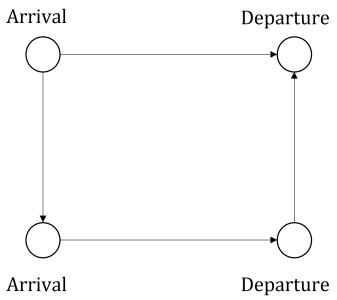
Train operations are not independent of each other - a fact that might lead to knock-on delays under given circumstances. There are two main classes of knock-on delays: route conflicts due to shared infrastructure and waiting for connecting trains (Daamen *et al.* (2008)). The type of knock-on delays that are modelled in the scope of this thesis comprise knock-on delays due to route conflicts only. Route conflicts can happen at any moment at which trains share the same infrastructure, i.e. both in stations and on sections between stations. The models implemented in this thesis only account for knock-on delays that happen within stations because data available does not comprise information about train operations on sections (section 3.1). There are several possible knock-on situations within stations which are presented by Corman and Kecman (2018) and displayed in figure 2. There is one specific type of knock-on delays within stations that is modelled in the scope of this thesis, namely the one depicted in figure 2(a). Through movements of trains (depicted in figures 2(c) to 2(e)) are not considered because the data provided does not comprise information about them (section 3.1). The knock-on type shown in figure 2(b) is neglected because it accounts for the fact that train order might change within the station, which cannot happen in the case study as described in section 3.1. Knock-on types as depicted in figure 2(f) model connection knock-on delays rather than route conflicts which is why they are not part of the scope of this thesis. A simple situation in which knock-on delays as shown in figure 2(a) happen is the following: If track infrastructure within a station is simple, i.e. if there is only one track per direction, a knock-on delay could possibly happen if two trains driving in the same direction are scheduled tightly. The moment the first scheduled train is late, it has a possible influence on the follow up train as the follow up train cannot switch to another track to escape the knock-on delay and has to wait until the necessary track infrastructure is available.

According to the models implemented, there are two criteria that need to be fulfilled in order for a train j to be considered to have a knock-on effect on train i. First of all, a train j is only considered to have a knock-on effect if it is scheduled tightly before train i at station k. A tight schedule is defined to be less than 5 minutes headway between the scheduled arrivals of the trains. Furthermore, if train j fulfils the first condition, it needs to have a certain delay at station k so that it is considered to have a knock-on influence. If train j arrives on time, it can be assumed that no knock-on effect happens because train schedules are designed so that trains do not experience knock-on delays if operations run on time.

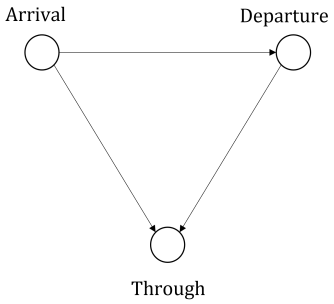
Figure 2: Types of knock-on delays in stations by Corman and Kecman (2018)



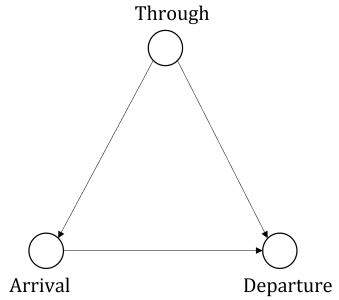
(a) both trains with scheduled stop



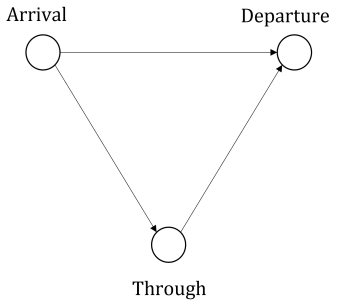
(b) both trains with scheduled stop and overtaking



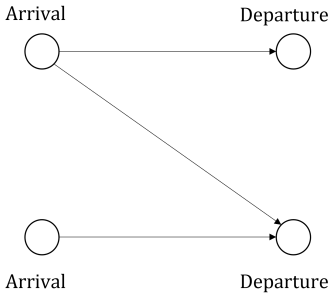
(c) second train without scheduled stop



(d) first train without scheduled stop



(e) overtaking



(f) both trains with scheduled stop and connection

There are two parameters that need to be chosen when it comes to modelling knock-on delays: First of all - just as in the standard model - a parameter p needs to be chosen, marking the number of delays that should be included in the model. The remaining parameter is needed in formulating knock-on delay conditions. It concerns the delay threshold parameter marking by how much train j must be delayed at least in order to have a knock-on effect on train i. It is assumed that only one train j can have a knock-on influence on train i. If therefore more than one train fulfils the necessary criterion, only train j that is scheduled closest to train i is considered to have a knock-on effect and is therefore included in the model.

This thesis considers two different approaches to model knock-on delays. The formula for the model accounting for knock-on delays - called the interacting model - is the same as was developed by Wang and Work (2015). Its mathematical formulation is the following:

$$\hat{y}^i_{k,t} = a_1 y^i_{k-1,t} + \cdots + a_p y^i_{k-p,t} + c_k + \phi^i_{k,t} \tag{4}$$

$\hat{y}^i_{k,t}$: estimated delay of train i on trip t at station k

$y^i_{k-p,t}$: delay of train i on trip t at station k-p

a_p : relationship of delays among the current and past stations

c_k : intercept term which allows for constant delays on a section

p : control variable marking how many previous stations should be included in the model

$\phi^i_{k,t}$: knock-on delay

The term for the knock-on delays is modelled by Wang and Work (2015) as follows:

$$\phi^i_{k,t} = \sum_{(j,\tilde{k},\tilde{t})\in\Omega_{i,k,t}} b_j y^j_{\tilde{k},\tilde{t}} \tag{5}$$

$\Omega_{i,k,t}$: set of train-station pairs contributing to the delay of $y^i_{k,t}$

$y^j_{\tilde{k},\tilde{t}}$: delay of tran j at station $\tilde{k}$ during trip $\tilde{t}$

With $\Omega_{i,k,t}$ being defined as: "If train j is scheduled at a [sic!] the same or neighboring station $\tilde{k}$ within an hour of train i at station k, then the delay of train j at station $\tilde{k}$ is considered as part of the regression". The delays of the trains j that possibly have a knock-on influence on train i are included as linear terms in the regression. The interacting model developed in the scope of this thesis defines $\Omega_{i,k,t}$ differently, namely according to the criteria listed above. In case no train fulfils the criteria, $y^j_{\tilde{k},\tilde{t}}$ becomes 0. To this effect, the formula implemented in this thesis can be more accurately described by:

$$\hat{y}^i_{k,t} = a_1 y^i_{k-1,t} + \cdots + a_p y^i_{k-p,t} + c_k + \phi^i_{k,t} \tag{6}$$

$$\phi^i_{k,t} = \begin{cases} b_j * y^j_{k,\tilde{t}} & \text{if some train } j \text{ fulfils the conditions} \\ b_j * 0 & \text{else} \end{cases} \tag{7}$$

The second modelling approach - called the binary model - is a conceptual derivative of the interactive model. Its goal still is to account for knock-on delays by means of a linear regression, however, the modelling approach has situational meaning and handles delay prediction differently than the interacting model. In case a knock-on delay happens according to the conditions defined above, the delay of train i at station k is assumed to only depend on the delay of train j at station k that causes the knock-on delay. In case no knock-on delay is detected, the delay of train i at station k is modelled with the

standard modelling approach as described in section 2.2. In this case the delay of train i is assumed to not be influenced by any other train j but only depend on its own current delay situation. With these facts in mind, the model is described as follows:

$$\hat{y}^i_{k,t} = \begin{cases} b_j * y^j_{k,\tilde{t}} & \text{if a knock-on delay happens} \\ a_1 y^i_{k-1,t} + \cdots + a_p y^i_{k-p,t} + c_k & \text{else} \end{cases} \tag{8}$$

Both modelling approaches are subject to several assumptions and restrictions. They mainly concern the conditions under which a knock-on delay is assumed to happen as there are many more factors than the ones defined above that determine whether a knock-on delay happens or not. The reason why these factors are not included in the model is that their modelling is very complex and therefore lies beyond the scope of this thesis. Even the finding of all influencing factors is difficult which is why only an inchoate discussion of possible influencing factors will be given in the following paragraph.

First of all, the model assumes that train orders never change. If the scheduled arrival of train j therefore is prior to the scheduled arrival of train i at station k, train j is also assumed to arrive first at station k in either case. The model does not account for the case in which train order changes along the route due to a major delay of train j. Secondly, the model detects a delay as soon as the delay of train j exceeds a certain threshold. It is not accounted for the fact, however, that knock-on effects also depend on the delay of train i. In case train i is delayed as well and specifically even more delayed than train j, the possibility of train j having a knock-on influence on train i becomes very unlikely. The model also assumes that train i does not have the possibility to switch track in order to escape a knock-on delay, meaning that train i and train j must share the same infrastructure. The model is therefore bound to stops with a simple track infrastructure of one track per direction. As soon as there is more than one track available, modelling complexity increases drastically as train i and train j do not necessarily share the same infrastructure anymore. This restriction is accounted for by choosing a suitable station for the knock-on delay modelling as explained in chapter 3.1. Last but not least, only one-directional knock-on delays are considered. It is not accounted for the fact that also forthcoming trains possibly lead to a knock-on delays if track infrastructure right before or after the station is single-track so that trains have to cross within stations. This restriction also is accounted for by choosing a suitable station to model knock-on delays.

3 Results

3.1 Data Set

The open-source data set is provided by Schweiz (2022) and has been filtered and pre-processed for the sector between Zurich and Chur by the Institute for Transport Planning and Systems at ETH Zurich. It spans the whole year 2021 and contains arrival and departure information about all train services stopping at any station in the sector including both local and intercity trains. There is no information about train movements between the stations available. Also freight trains that possibly have an influence on passenger trains are neither included in the data set nor considered in the model. Relevant information stored in the data set include the run ID which is an individual ID for each run of each train, the planned arrival time at a station k, the arrival delay at station k given to the second, the start stop of each run and the name of station k. A comprehensive list of all variables available in the data set is given in the appendix in section A.1. As described in chapter 2, this thesis deals with arrival delays only, departure data therefore is neglected.

Table 1: Stops included in the data set per train service

S2	S8	S24
Zürich HB	Zürich HB	Zürich HB
Zürich Wiedikon	Zürich Wiedikon	Zürich Wiedikon
Zürich Enge	Zürich Enge	Zürich Enge
Thalwil	Zürich Wollishofen	Zürich Wollishofen
Horgen	Kilchberg	Kilchberg
Wädenswil	Rüschlikon	Rüschlikon
Richterswil	Thalwil	Thalwil
Pfäffikon Sz	Oberrieden	
Altendorf	Horgen	
Lachen	Au ZH	
Siebnen-Wangen	Wädenswil	
Schübelbach-Buttikon	Richterswil	
Reichenburg	Bäch	
Bilten	Freienbach SBB	
Ziegelbrücke	Pfäffikon Sz	
Unterterzen		
(only on weekends and holidays)		

Table 2: Scheduled arrivals and departures of trains in Enge

Train	Scheduled arrival	scheduled departure
S2 (direction South-North)	xx:05 / xx:35	xx:06 / xx:36
S2 (direction North-South)	xx:23 / xx:53	xx:24 / xx:54
S8 (direction South-North)	xx:17 / xx:47	xx:18 / xx:48
S8 (direction North-South)	xx:11 / xx:41	xx:12 / xx:42
S24 (direction South-North)	xx:02 / xx:03	xx:32 / xx:33
S24 (direction North-South)	xx:26 / xx:56	xx:27 / xx:57

For the case study conducted in the scope of this thesis, only a fractional amount of the whole data set is investigated. More specifically, one local train services is analysed, namely the S2 service running between Zürich Flughafen and Unterterzen. Any conclusions that are drawn are therefore valid for local train services only. To be able to draw universal conclusions, a comparison of different train types with different stopping patterns would need to be conducted which is beyond the scope of this thesis. The reason why local train services rather than intercity or inter-regional trains are investigated more closely is that local train services tend to stop much more often at stations with a simple track infrastructure which is crucial for the modelling of knock-on delays as explained in chapter 2. The S2 service was furthermore chosen because it stops at a station that is predestined for knock-on delays to happen, namely Zürich Enge. Zürich Enge is the exemplary station

Figure 3: Situation in Zürich Enge with red lines marking track infrastructure of public transport (source: map.geo.admin)

for which knock-on delays are modelled in the scope of this thesis. A general plan of location of Zürich Enge is shown in figure 3. Zürich Enge has a simple track infrastructure with one track for each direction and double-track infrastructure on the adjacent track sections. Due to the lack of backup capacities, trains driving in the same direction therefore may experience knock-on effects given that train services are scheduled tightly. Forthcoming trains can be neglected in the analysis because trains driving in different directions do not share the same infrastructure neither in the station nor on the adjacent track sections. There are three local train services stopping in Zürich Enge, namely the S2 service, the S8 service travelling between Winterthur and Pfäffikon Sz and the S24 service travelling between Thayngen and Zug. An overview of their scheduled route that lies in the sector between Zürich and Chur is given in table 1. The actual routes of the trains are longer, however the data set does not include information about stops located outside the sector between Zürich and Chur. A comprehensive list of the schedule of all three train services is given in the appendix in section A.2. The scheduled departures and arrivals of the three train services in both directions in Zürich Enge are listed in table 2. All services run a half-hourly schedule. The scheduled arrival and departure data shows that the S2 and the S24 service are scheduled within three minutes of each other in both directions which may lead to the S24 service having a knock-on effect on the S2 service. Even though the S2 and the S24 service are scheduled tightly in both directions, a knock-on delay case study is only conducted for the South-North direction of travel. There is a simple reason why: in North-South direction, the S2 and the S24 service drive the same route, i.e. the route Zürich HB - Zürich Wiedikon - Zürich Enge as shown in table 1. The chance for a knock-on delay to happen in Zürich Enge and not already at one of the preceding stations therefore can be assumed to be very small. In South-North direction however, the S2 service drives from Thalwil straight to Enge whereas the S24 service drives from Thalwil via Rüschlikon, Kilchberg and Zürich Wollishofen to Zürich Enge. Both services still use the same track infrastructure but they do not share the stopping pattern. It is therefore much more likely that a knock-on delay happens in Zürich Enge in South-North direction than in North-South direction of travel.

So far, it has been mentioned that there possibly is a strong interaction between the S24 and the S2 service. The S8 service has not been included in the analysis yet. Table 2 shows that the S8 service is not tightly scheduled with neither the S2 nor the S24 service and therefore is unlikely to trigger a knock-on effect on or experience a knock-on effect due to neither of the other two train services. Figure 4 and table 3 undermine this. The S2 and the S24 service clearly show the greatest correlation both visually and numerically whereas the S8 service does not show significant correlation with neither of the two other train services.

Table 3: Correlation of the delays of the train services in Zürich Enge

	Correlation
Delay of S2 vs delay of S8 in Enge	0.26
Delay of S2 vs delay of S24 in Enge	0.56
Delay of S8 vs delay of S24 in Enge	0.12

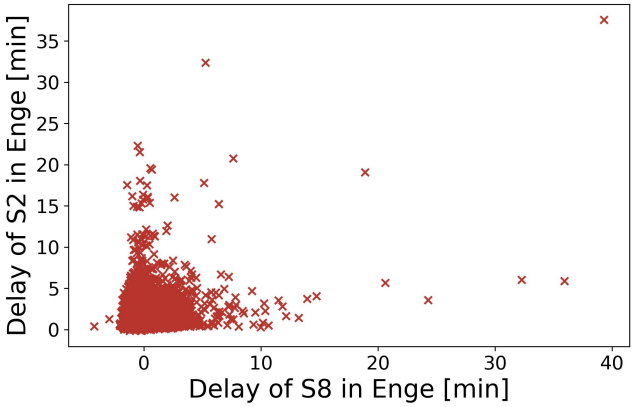
Modelling knock-on delays in Zürich Enge comes with one restriction: disregarding the three stopping trains, there are two trains passing the station Zürich Enge, namely the S25 service from Zürich HB to Linthal and the Regio Express from Zürich HB to Chur. Even though these trains do not stop in Zürich Enge, they still use the local track infrastructure and thereby possibly affect the delay of the stopping trains. As the data set does not contain information about the movement of trains between stations, these effects cannot be included in the model leading to a possible decline in prediction accuracy.

3.2 Data Preprocessing and Experimental Setup

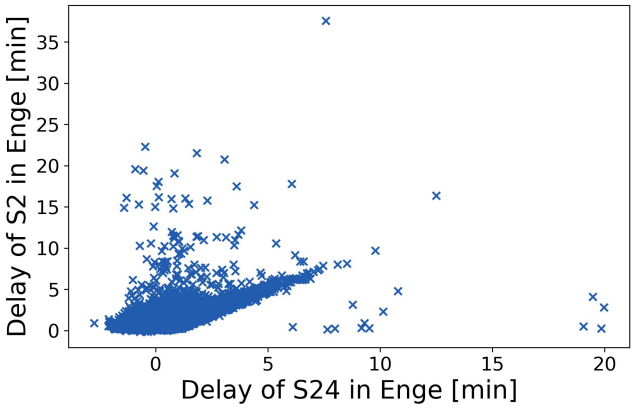
Preprocessing the data includes splitting the original data set into subsets based on the train service and the direction. Each train can run either in North-South or in South-North direction, therefore resulting in two subsets per train service. The North-South direction will from now on be referred to as downward direction and the South-North direction as upward direction. The data is not filtered temporally, e.g. by time of the day or weekdays and weekends. The model is rather trained and tested with all data that was sampled spanning the whole year of 2021.

All models that are implemented are tested and trained with a k-fold cross validation. K is set to be 5, i.e. the model is trained and tested with a training set containing 80% of the data and a test set containing 20% of the data whereat every subset generated by the k-fold cross validation is used four times to train the model and once to test it. Statistical values to assess the performance of the model and operating figures to interpret it are the mean absolute error (MAE) and its standard deviation, the median absolute error and the values for the parameters trained by the model. The model was trained by means of the ordinary least squares method provided by the statsmodels library (statsmodels (2022)).

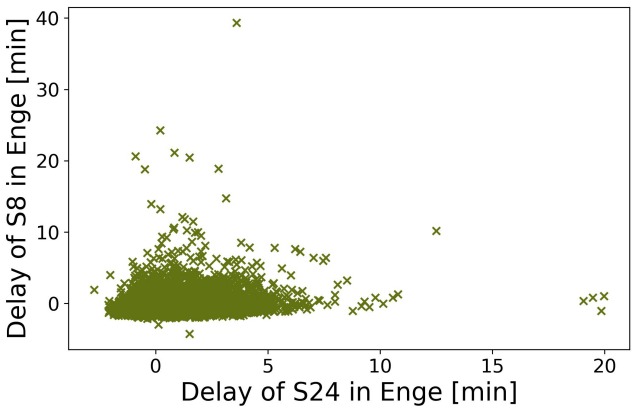
Figure 4: Delay relations



(a) Delay of S8 vs. delay of S2 in South-North direction



(b) Delay of S24 vs. delay of S2 in South-North direction



(c) Delay of S24 vs. delay of S8 in South-North direction

3.3 Implementation and Results of the Standard Model

Implementation

The standard model is analysed and assessed for the S2 service in upwards direction of travel as this also is the interesting service direction for modelling knock-on delays later (section 3.1). The only parameter chosen prior to training the model is the p-parameter indicating how far back on the route actual arrival delays are included in the model. The model is analysed and assessed for $p = 1, 2, 3$. If p is chosen to be 1, only the delay of station $k - 1$ feeds into the model. If p is chosen to be 2, the actual arrival delay at station $k - 1$ and $k - 2$ feeds into the prediction of the arrival delay at station k and if p is chosen to be 3, all arrival delays from station $k - 1$ to station $k - 3$ are included. Model parameters $a_1, ..., a_p$ and c_k are trained individually for each station k along the route of the S2 service as listed in table 1.

Results

First, the influence of p on the prediction accuracy will be analysed. Table 4 lists the MAE of the standard model depending on the p value with $p = 1, 2, 3$. The results clearly show that there is neither an improvement nor a deterioration of the results when choosing different values of p. Figure 5 displays the values of the trained parameters $a_i, i = 1, 2, 3$ and c_k across the whole route of the S2 service travelling in upward direction for $p = 3$. Parameter behaviour is very similar for $p = 1$ and $p = 2$ respectively. The corresponding plots can be found in section A.3 in the appendix. Figure 5 shows that c_k oscillates wildly between -1.5 and 1.5. While a_1 on the other hand remains stable at a level of 1, a_2 and a_3 are close to 0 over the course of the whole route. a_2 and a_3 are the factors multiplied with the delay at station $k - 2$ and $k - 3$ respectively. These two delays therefore barely influence the prediction result due to the dimension of a_2 and a_3.

As listed in table 4 The standard model of the S2 service travelling in upwards direction has an average MAE of 0.29min across the whole route contained in the data set whilst the MAE of the baseline approach is 0.85min. The standard model therefore manages to

Table 4: MAE of standard model with different p-values

Train Service	MAE for p = 1	MAE for p = 2	MAE for p = 3	MAE Baseline
S2 upwards	0.29min	0.30min	0.30min	0.85min

Table 5: Information about the section Thalwil-Zürich Enge of the S2 service

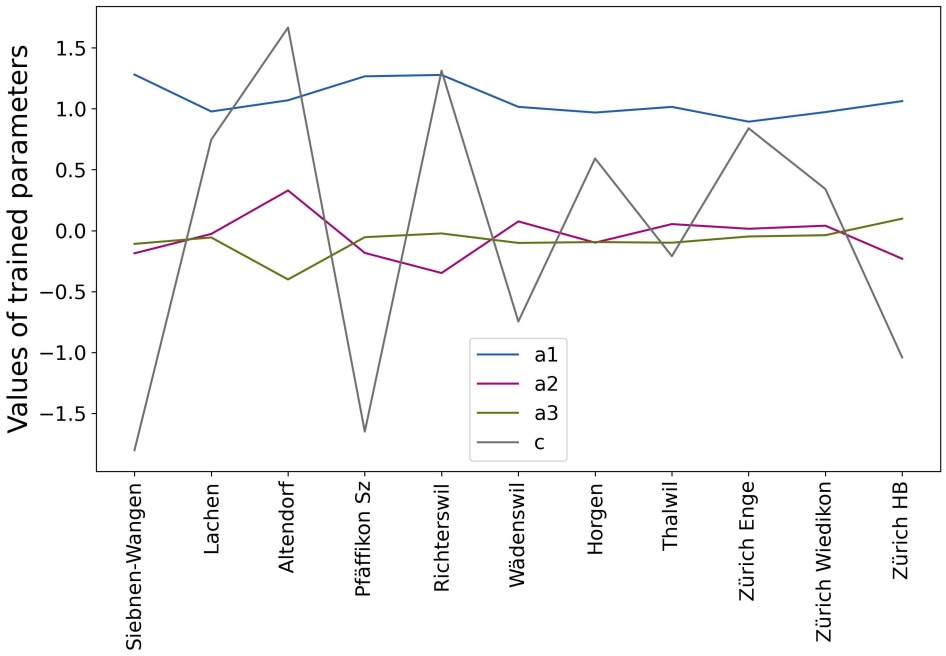
	MAE	a_1	c_k
Baseline Approach	0.85min	-	-
Standard Model	0.65min	0.87	0.84

improve prediction accuracy on average by about 65%. The improvement however differs depending on station k as shown in figure 6. While the MAE of the baseline approach oscillates from station to station with three main peaks in Siebnen-Wangen, Pfäffikon Sz and Zürich HB, the standard model prediction remains stable at a MAE level of about 0.25 minutes. Only for the stations Ziegelbrücke, Zürich Enge and Zürich HB, the MAE of the standard model spikes briefly to a MAE of 0.6 minutes. Possible reasons are discussed in section 4.

3.4 Results of Models Accounting for Knock-On Delays

The results presented in the following two sections do not comprise the whole route of the S2 service anymore but only the section from Thalwil to Zürich Enge due to the inclusion of knock-on delays (section 3.1). The MAE and parameter values of the baseline approach

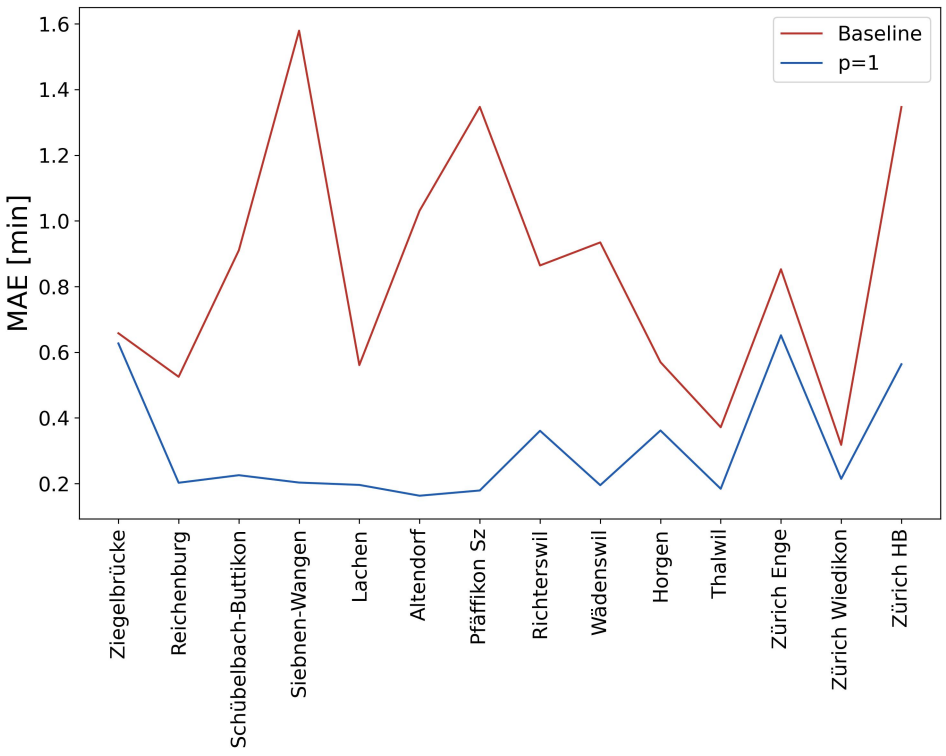
Figure 5: Trained parameters of the standard model for p=3



and the standard model for this specific section are listed in table 5. They are useful when it comes to evaluating and interpreting the results of the interacting model and the binary model.

As described in section 3.1, the train possibly causing knock-on delays is the S24 service. To include knock-on delays, the delay of the S24 service in Zürich Enge therefore needs to be known. There are two possible approaches to get the delay of the S24 service in Zürich Enge: In the first one, the actually measured arrival delay feeds into the model. This is a valid approach because the S24 service is assumed to arrive in Zürich Enge before the S2 service in any case due to the assumption that train order never changes. The arrival delay of the S24 is therefore always known prior to the S2 service arriving in Zürich Enge. This approach comes with the disadvantage that due to the tight schedule of the S2 and the S24 service, delay predictions for the S2 service can only be made as soon as the S24 service has arrived in Zürich Enge which only allows for very short-term predictions. To counteract this inconvenience, instead of waiting for the S24 service to arrive in Zürich Enge, the delay of the S24 service can be predicted by means of the standard model based on the actual arrival delay of the S24 service in Zürich Wollishofen, i.e. the station preceding Zürich Enge on the route of the S24 service (table 1). The S24 service takes around 5 minutes to run from Zürich Wollishofen to Zürich Enge. Predicting the delay

Figure 6: MAE of the S2 service in upward direction trained with the standard model



therefore allows delay estimates to be made up to 5 minutes earlier. The disadvantage of this approach however is that the predicted value of the arrival delay naturally is not as accurate as the measured delay. The losings in prediction accuracy that are caused by estimating the delay of the S24 service are analysed and discussed in section 4. In case the delay of the S24 service is predicted, the parameters trained by the standard model for the section Zürich Wollishofen-Zürich Enge are needed: a_1 corresponds to 0.99 and c_k is -1.46.

3.4.1 Implementation and Results of the Interacting Model

Implementation

When accounting for knock-on delays, the crucial parameter is the delay threshold indicating by how much the S24 service needs to be delayed at station k so that the model detects a knock-on effect on the S2 service. Section 2.3 suggests that a p parameter needs to be chosen as well, but section 3.3 investigated the effect of p on the model performance and came to the conclusion that p does not influence prediction accuracy which is why p is chosen to be 1 in any case.

The interacting model is only trained on runs for which the model assumes a knock-on delay to happen. By choosing different delay thresholds, this training set changes. The model was tested for every full minute in a defined delay threshold interval. The lower bound of this interval was chosen to be 0, meaning that the model assumes a knock-on delay to happen as soon as the S24 has some kind of delay in Zürich Enge, no matter how small. The upper bound was chosen to be 3, which corresponds to the scheduled arrival difference of the S2 and the S24 service. As soon as the S24 is thus delayed by more than 3 minutes in Zürich Enge, the model assumes a knock-on delay to happen. The model was tested for the delay threshold parameter being 0, 1, 2 and 3.

Results

The results of the MAE and the respective improvements compared to the baseline approach and the standard model are listed in table 6. The model performs best for the delay threshold being 0 and almost as well for the delay threshold being 1. Prediction accuracy of the model however decreases significantly for the delay threshold being 2 or 3. For the best performing model, predicting the delay of the S24 service rather than taking the actually measured delay leads to an increase in the MAE of 0.01 minutes only.

Table 6: MAE and improvement of the interacting model

Threshold [min]	MAE [min]	Improvement compared to baseline approach	Improvement compared to Standard Model
0 without prediction	0.42	51.28%	36.25%
1 without prediction	0.43	49.32%	33.69%
2 without prediction	0.70	17.58%	-7.85%
3 without prediction	1.85	-117.40%	-184.46%
0 with prediction	0.43	49.90%	34.46%

Table 7: Parameter values of the interacting model

Threshold [min]	a	b	c
0 without prediction	0.72	0.65	0.25
1 without prediction	0.65	0.68	0.28
2 without prediction	0.60	0.43	1.23
3 without prediction	0.65	0.16	2.72
0 with prediction	0.72	0.64	0.26

The statistical values for delay prediction of the S24 service are listed in table 8: The bias of the prediction is negligible, the MAE and the median absolute error do not differ significantly, suggesting that there are only few outliers. The parameters that were trained by the interacting model are listed in table 7. While values for a remain stable for all thresholds, b drops significantly and c shoots up for the delay threshold being 2 and 3.

3.4.2 Implementation and Results of the Binary Model

Implementation

The only difference between the binary decision model and the interacting model lies in the mathematical formulation. The binary decision model is therefore assessed for the same delay threshold values as the interacting model. Parameter b is trained only on runs for which a knock-on delay is assumed to happen, parameters a and c on the other hand are trained on the rest of the data. When testing the model, the prediction approach is chosen depending on whether for the respective test run a knock-on delay is detected or not.

Table 8: Statistical values of the prediction of the S24 service depending on different thresholds

Threshold [min]	Mean Error [min]	MAE [min]	Median Absolute Error [min]
0	-0.003	0.166	0.132
1	-0.008	0.169	0.136

Table 9: MAE and improvement of the binary model with and without prediction of the delay of the S24 service

Threshold [min]	MAE [min]	Improvement compared to baseline approach	Improvement compared to standard model	Improvement compared to interacting model
0 without prediction	0.46	45.8%	29.1%	-11.19%
1 without prediction	0.38	55.6%	41.9%	12.58%
2 without prediction	0.42	50.8%	35.6%	39.93%
3 without prediction	0.61	28.4%	6.3%	71.07%
1 with prediction	0.39	53.7%	39.4%	8.64% %

Results

The MAE and respective improvements of the binary decision model compared to the baseline approach as well as the standard model and the interacting model are listed in table 9. When comparing the result of the binary model to the results of the interacting model, the improvement is calculated with respect to the result of the interacting model at the same threshold level and not with respect to the best performing delay threshold of the interacting model. The binary model performs much more stable than the interacting model. Only for the delay threshold being 3, model performance decreases significantly. The best performance is achieved at a delay threshold of 1 minute. Predicting the delay of the S24 service instead of taking the actually measured value at this delay threshold value hardly influences model performance. Again, table 8 lists the statistical values of the prediction of the S24 service. Looking at the parameter values, they remain quite stable for all thresholds (table 10). Only c increases significantly for greater thresholds. Table 11 shows the MAE and the median for the two different parts of the model individually. Performance of the part of the model that accounts for knock-on delays is significantly worse than the part that does not. The difference in the MAE depending on whether the delay of the S24 is predicted or not however is rather insignificant. All MAEs seem to be contorted by outliers as the median absolute errors are remarkably lower than the MAEs.

Table 10: Parameter values of the binary model

Threshold [min]	a	b	c
0 without prediction	0.95	1.02	0.43
1 without prediction	0.90	1.00	0.46
2 without prediction	0.86	0.98	0.61
3 without prediction	0.85	1.06	0.84
1 with prediction	0.90	1.00	0.47

Table 11: MAE and median for the two different parts of the binary model

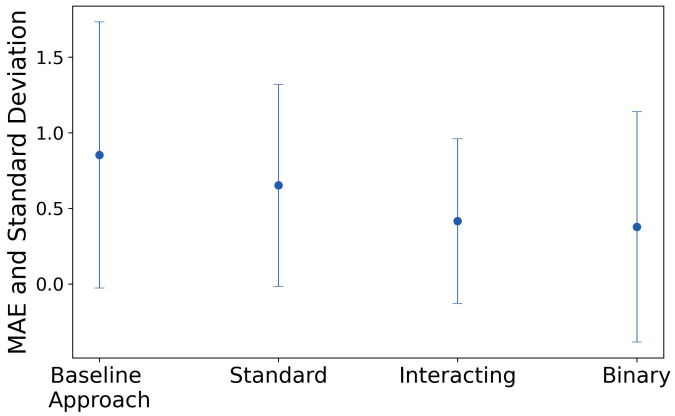
	MAE [min]	Median Absolute Error [min]
Model detects knock-on delay	0.49	0.32
Model detects knock-on delay (and delay of S24 is predicted)	0.53	0.35
Model does not detect knock-on delays	0.31	0.23

3.5 Comparison of All Models

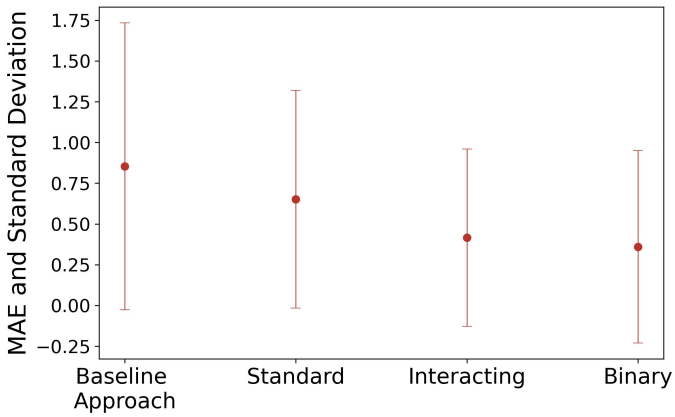
When comparing all three models, the best performing parameters for each model are chosen, i.e. a delay threshold of 0 for the interacting model and a delay threshold of 1 for the binary model.

MAE and Standard Deviation

Figure 7(a) displays the MAE of each of the models and their standard deviation. The performance of the interacting and the binary model is both visually and numerically clearly better than the one of the baseline approach and the standard model. Figure 7(a) furthermore illustrates that prediction accuracy of the interacting model and the binary model is very similar but that the binary decision model comes with a significantly larger standard deviation. While the interacting model has an MAE of 0.42min and a standard deviation of 0.54 minutes, the binary decision model has an MAE of 0.38 minutes and a standard deviation of 0.85 minutes. Figure 8 gives an explanation for the high standard deviation. It displays the errors of one validation test set that is exemplary for model behaviour. The MAE and the standard deviation of the binary model are visually contorted by gross errors of about 15 minutes. Removing this outlier improves the MAE of the binary model from 0.38 minutes to 0.36 minutes and reduces standard deviation from 0.76 minutes to 0.59 minutes. These outlier-free results are visualised in figure 7(b) and listed in table 12.



(a) MAEs of the models and their standard deviation



(b) MAEs of the models and their standard deviation without outliers

Median

Knowing that results may be contorted by outliers, the median absolute error, which is insensitive to outliers as opposed to the MAE, should be considered in the analysis as well. The median absolute errors of all models are listed in table 12. With the median absolute error being significantly better than the MAE for all models, it is implied that every model is to some extent contorted by outliers. Outliers mark runs on which factors that could not be captured by the model significantly influenced the true delay of the S2 service.

Bias

The error displayed in figure 8 is calculated by subtracting the estimated delay from the true delay. As there are outliers with a positive sign only, all models tend to underestimate delay and never overestimate it significantly. Moreover, all models show a positive bias

Table 12: Median, MAE and standard deviation of the models

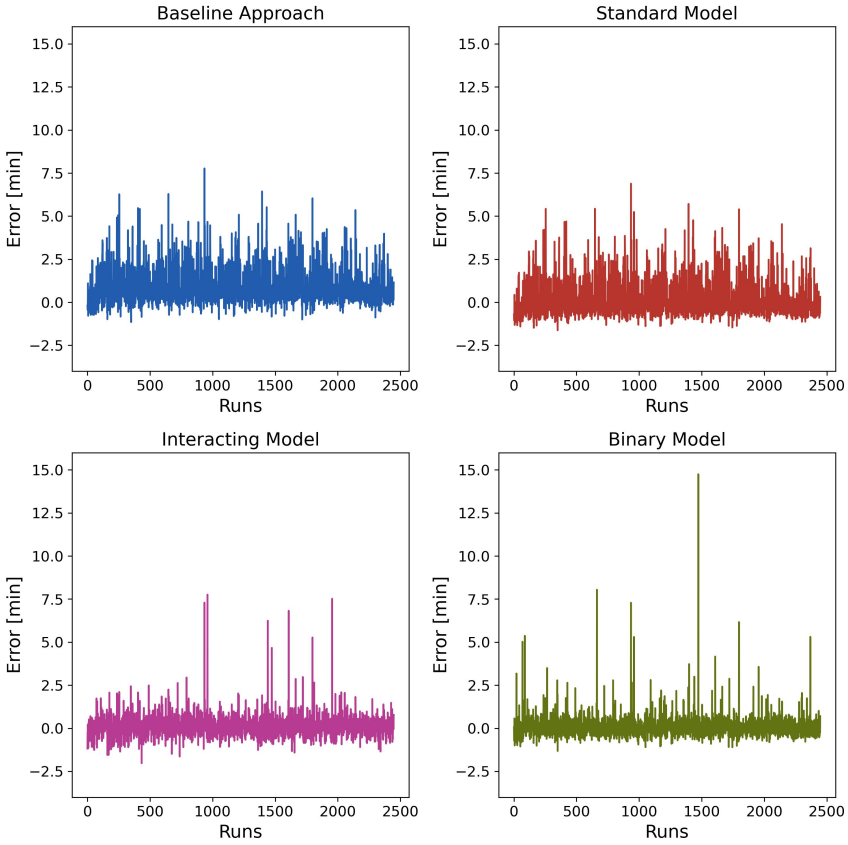
	Baseline Approach	Standard	Interacting	Binary
MAE [min]	0.85	0.65	0.42	0.38 (outlier-free) 0.36
Standard Deviation	0.88	0.66	0.54	0.76 (outlier-free) 0.59
Median [min]	0.57	0.50	0.30	0.26

which is listed in table 13. This bias is calculated by averaging all errors associated with the model. The closer to 0, the less biased the model. All four models feature positive bias whereat the standard model and the binary model are less biased than the interacting model. The baseline approach is strongly biased with a mean error of 0.85.

Table 13: Mean error of all models

	Baseline Approach	Standard	Interacting	Binary
Mean Error	0.85	0.09	0.14	0.08

Figure 8: Prediction errors for the S2 service in Zürich Enge



4 Discussion

4.1 Discussion of the Standard Model

The Value of p

The standard model showed neither model improvement nor deterioration contingent on the p value. Wang and Work (2015) came to the same conclusion. They found that p does not affect the performance of the model and that with a higher order of p the performance rather decreases than increases. One possible explanation is that the information about the delay at some station $k-2$ does not contribute to the prediction accuracy anymore once the delay at station $k-1$ is known. The delay at station $k-2$ directly influences the delay at station $k-1$ due to the correlation of consecutive arrival delays on a trip t of train i. The effect of the delay at station $k-2$ is therefore already captured in the delay at station $k-1$. If the delay at station $k-2$ as well as the delay at station $k-1$ were included in the model, the effect of the delay at station $k-2$ would feed into the model twice, once directly by including the delay at station $k-2$ and once indirectly by including the delay at station $k-1$. This theory can be substantiated by looking at the values shown in figure 6. Parameters for a_2 and a_3 tend towards 0. The delays at stations $k-2$ and $k-3$ therefore got barely any influence on the prediction result. In case one of the two parameters is significantly larger than 0, the other one is significantly lower than 0, leading to their effects cancelling each other out. The question of how the arrival delay at stations $k-2$ or $k-3$ influence the delay at station k becomes interesting in case the model should be used to estimate delays with a prediction horizon of more than one station. This question however is not investigated in the scope of this thesis.

Parameter Values and MAE

As already described in section 3, the prediction accuracy of the standard model is better than the prediction accuracy of the baseline model at every station. In few special cases, the difference between the baseline approach and the standard model is particularly significant, namely for Siebnen-Wangen, Pfäffikon Sz and Zürich HB: At all these stations, the MAE of the baseline approach is significantly higher than the MAE of the standard model. An explanation is given by the parameter values that were trained by the standard model as displayed in figure 5. Figure 5 shows that a_1 remains stable at a value of around 1, indicating that the delay at station $k-1$ feeds into the model as a whole. This is also the case for the baseline approach, which is based on the assumption that the delay

stays constant between stations. c_k however has a value of around -1.5 for the stations of Siebnen-Wangen, Pfäffikon Sz and Zürich HB. For reasons, which will not be discussed here any further, the S2 service is therefore able to decrease the delay on these sections by about 1.5 minutes. The MAE of the baseline approach (figure 6) at the stations Siebnen-Wangen, Pfäffikon Sz and Zürich HB is around as high as these 1.5 minutes because the baseline approach naturally does not manage to capture this 1.5 minute delay decrease. The standard model, on the other hand, manages to include this effect in the prediction which is why prediction accuracy of the standard model is significantly better at these three stations.

While the reasons for the spikes in the MAE of the baseline approach were clarified, the question remains why the standard model spikes at the stations Ziegelbrücke, Zürich Enge and Zürich HB. The spike in Ziegelbrücke will not be discussed any further because no conclusive statement about possible reasons can be made. Usually, Ziegelbrücke is the terminal stop of the S2 service. Only on weekends and holidays, the S2 service runs all the way until Unterterzen. This leads to data on the section Unterterzen-Ziegelbrücke being scarce and contorted by holiday travelling behaviour, which is why a more coherent and thorough analysis of this station and influencing factors would be needed to be able to make a statement about the reason for the bad model performance. This lies beyond the scope of this thesis and will not be discussed here. One possible reason for the spike in Zürich Enge might be the interaction with other trains, i.e. knock-on effects, that take place at this station and that naturally cannot be captured by the standard model. As both the interacting and the binary model manage to decrease the MAE significantly compared to the standard model, knock-on effects are very likely to be one of the reasons why the standard model is not able to predict delays in Zürich Enge as accurately as for other stations on the section. As far as Zürich HB is concerned, it is a very complex station and one of the most important transport nodes in Switzerland leading to many factors such as other trains, track geometry and technical disturbances that all possibly have an influence on the delay of a train service. A linear regression only including the delay of the S2 service in Zürich Wiedikon to predict the delay at Zürich HB therefore seems to be too simple to capture all effects influencing delay propagation on this section.

4.2 Discussion of the Interacting Model

The interacting model shows good performance for delay thresholds of 0 and 1 minutes (table 6). The MAE for the best threshold, in this case 0 minutes, decreases only by 0.01 minutes in case the delay of the S24 service is predicted (table 6). There are several factors

contributing to this phenomenon. Firstly, the bias of the delay prediction of the S24 service is negligible as the mean error is very close to 0 resulting in an unbiased prediction. Secondly, the MAE and the median absolute error both are small, suggesting that delay predictions are in fact accurate. Due to the fact that this prediction is multiplied with $b = 0.68$, the effect of the inherent discrepancy of the predicted and the measured delay is shrunk even further.

The reason why model performance drops significantly for delay thresholds of 2 and 3 minutes can be explained by looking at the parameter values that were trained by the model (table 7). With a delay threshold of 2 or 3 minutes respectively, the interacting model was trained on runs only on which the S24 service preceding the S2 service was delayed by more than 2 or more than 3 minutes. While the a parameter remains stable, the model does not capture the likely knock-on effect of the S24 service on the S2 service in the b parameter - which shrinks given a higher threshold - but captures it in the constant c parameter. The model is tested on all runs, not only those on which presumably a knock-on delay happened. For a threshold of 2 respectively 3 minutes, a constant factor of 1.23 or 2.72 is therefore added to every prediction independent of whether a knock-on delay happened or not. This leads to estimates that are contorted by this factor c in case no knock-on delay happened. There is a total of 12'244 runs of the S2 service that are used to train and test the model. In case a threshold of 2 minutes is chosen, there are 1762 runs remaining on which the S24 service preceding the S2 service was delayed by more than 2 minutes. For a delay threshold of 3 minutes, there are only 633 runs remaining. On most test runs no knock-on delay therefore has taken place, however, the constant factor c is still added to the estimated delay leading to bad prediction results. If the model captured the influence in factor b instead of factor c, the results would not be contorted as b only influences the prediction result in case the model assumes a knock-on delay to happen. In all other cases, b is multiplied with 0 and therefore does not influence the result.

The parameter behaviour also explains why model performance is better for thresholds of 0 and 1 minutes than it is for thresholds of 2 or 3 minutes. The c parameter is significantly lower for a delay threshold of 0 or 1 minute (table 7). Predictions for runs without knock-on delays are therefore not contorted by the factor c. The delay prediction rather depends in equal parts on the delay of the S2 service and the delay of the S24 service as the trained values of a and b are about the same. As the results obtained by the interacting model improve prediction estimates compared to both the baseline approach and the standard model, it can be stated that the interacting model manages to at least partially capture the interdependence of train services in Zürich Enge.

4.3 Discussion of the Binary Model

The performance of the binary model is much more stable than the performance of the interacting model. Parameter values also remain much more stable for all four delay thresholds. These results suggests that dividing delay estimates in two cases, one of them accounting for interdependence between trains and one only considering the train's own delay, is more accurate than trying to model both effects in the same formula as it was done in the interacting model.

Model Performance Taking into Consideration Different Delay Thresholds

Model performance is the best for a delay threshold of 1 minute with an MAE of 0.38 minutes. For the same reasons as already stated in the previous section, the prediction accuracy is not aggravated by predicting the delay of the S24 service. Undermining the reasoning from the previous section, table 11 states that the MAE of the part of the model that accounts for knock-on delays only increases by 0.04 minutes when predicting the delay of the S24 service instead of taking the actually measured value. This MAE increase is not significant enough to have a drastic influence on model prediction accuracy because only a minority of the test-run delay estimates are predicted with the part of the model accounting for knock-on delays. This increase of 0.04 minutes therefore does not feed into prediction accuracy as a whole but is extenuated due to the fact that the majority of delay estimates are performed under the assumption that no knock-on delay took place.

The MAE at a threshold of 1 minute is significantly better than the MAE at a delay threshold of 0 minutes. One possible explanation is that there is a significant amount of delays of the S24 service that lie between 0 and 1 minutes. Totally, there were 12244 runs of the S2 service registered in the data set. On 9211 of these runs, the preceding S24 service was delayed by more than 0 minutes but only on 4649 runs, the preceding S24 service was delayed by more than 1 minute. 4562, i.e. almost half of the runs for which the model detected a knock-on delay in case the delay threshold was chosen to be 0, were therefore runs on which the S24 service was delayed by some value between 0 and 1 minutes. It can be assumed that no knock-on effects take place yet if the S24 service is late by a few seconds only. As the majority of detected knock-on runs however falls into exactly this category, the knock-on phenomenon cannot be captured correctly by the binary model for a threshold of 0 minutes.

Just as for the interacting model, model performance is by far the worst for a delay threshold of 3 minutes. Parameter values of $a = 0.85$ and $c = 0.84$ and the MAE of 0.61

minutes at a delay threshold of 3 minutes resemble the results of the standard model a lot. The standard model has an MAE of 0.65 minutes and parameter values for $a = 0.87$ and $c = 0.84$ in Zürich Enge. So far, it has been shown that there is an interaction between the S24 service and the S2 service in Zürich Enge under given circumstances as both the interacting and the binary model managed to improve prediction accuracy significantly. The standard model cannot capture these interactions by definition. The fact that the a and the c parameter of the binary model at a threshold of 3 minutes behave exactly the same as the a and c parameter of the standard model shows that the threshold is chosen too big.

Results showed that for the best threshold, i.e. a threshold of 1 minute, the part of the binary model that accounts for knock-on delays performs significantly worse than the part that does not account for knock-on delays. These results suggest that knock-on delay detection is not sufficiently refined yet or that the phenomenon of knock-on delays is not simple enough to be captured by one linear parameter only.

Outliers

Figure 8, which is representative for model behaviour, shows that the binary model tends to have gross outliers that do not occur in other models. These outliers occur because the model estimates delay to be much lower than it actually turn out to be. There is one specific situation which results to these outliers in the binary model: In case the delay of the S24 service exceeds the chosen threshold, the model detects a knock-on delay and assumes the delay of the S2 service to only depend on the delay of the S24 service. If the S2 service now is more significantly delayed than the S24 service, the model has no chance to estimate the delay of the S2 service correctly because the delay of the S2 service in Thalwil does not influence model prediction. This problem could possibly be solved by adding another delay recognition condition to the model. It would include that the difference of the delay of the S2 service in Enge and the delay of the S24 service in Wollishofen or Enge respectively must not exceed a certain threshold. The threshold for this maximum delay difference however would need further analysis and investigation.

The gross errors that occur in the binary model do not appear in the other models because they include the delay of the S2 service in Thalwil when predicting the delay in Zürich Enge. However, there are outliers that may appear in the other models as well. Outliers are inevitable when it comes to train delay prediction. In train operations, there are always unforeseen events that may cause a train to be unexpectedly delayed. These events are random at times and therefore cannot be predicted. Other events causing outliers

could possibly be detected by the model by including more influencing factors and/or increasing model complexity.

4.4 Discussion and Comparison of All Models

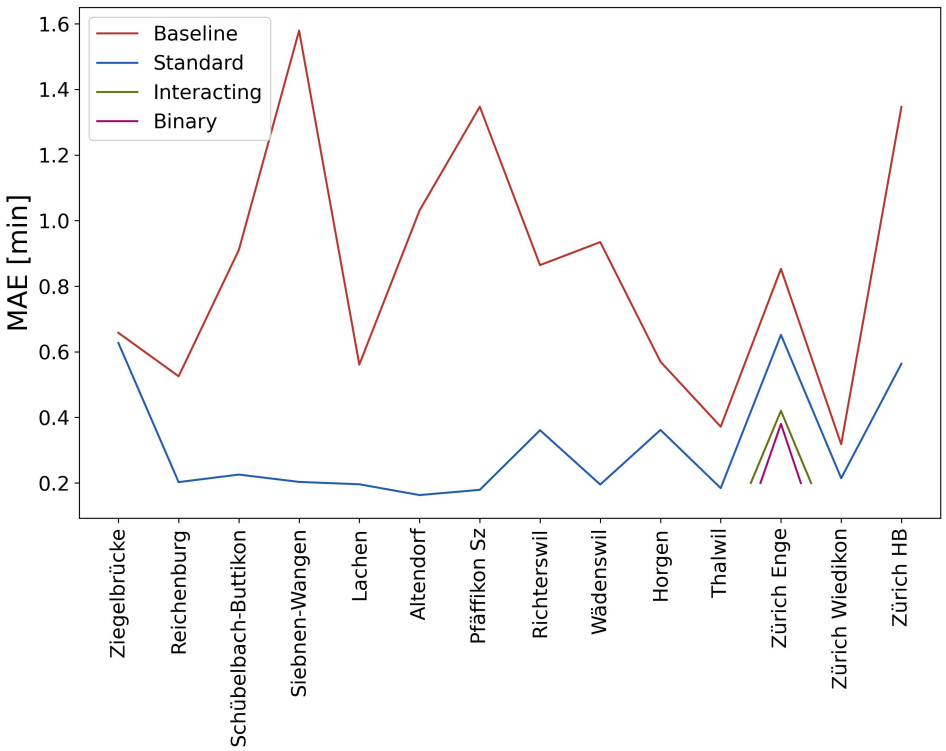
It can be stated that both the interacting and the binary model perform significantly better than the standard model. Performance of the interacting model and the binary model however is very similar. While the MAE as well as the median absolute error and the mean error of the binary outperform the equivalent values of the interacting model, the standard deviation of the interacting model is lower than the standard deviation of the binary model even when gross outliers are removed. Still, the binary model should be favoured over the interacting model as the crucial parameter in train operations is the absolute prediction error. Prediction accuracy of the binary model is 0.38 whereas prediction accuracy of the interacting model is 0.42. The model of preference therefore is the binary model.

But even by eliminating gross errors from the binary model, the MAE of the binary model is still greater than 0.3 minutes and therefore exceeds the average MAE of 0.29 minutes achieved by the standard model on average on the whole section (table 4. The performance of the binary model therefore is still worse than the average performance of the standard model. This suggests that either knock-on effects are not fully captured yet by the model or that there are more effects influencing the delay in Zürich Enge than considered by the model. As mentioned in section 2, the interacting and the binary model come with quite some assumptions. Especially only considering the delay of the S24 service and not considering the delay of the S2 service when detecting knock-on delays is a big restriction. Refining conditions in the binary model or including further influencing factors could possibly improve model performance.

5 Conclusion

This thesis aimed at analysing how the model of Wang and Work (2015) performs on a Swiss data set and how knock-on delays could be modelled more accurately. Even though trains in Switzerland are very punctual in general, the baseline approach showed that simply assuming constant delay propagation does not lead to accurate predictions. Already the standard model leads to a significant improvement compared to the baseline approach (figure 9). Both the interacting model and the binary model further improve prediction accuracy (figure 9) in Zürich Enge. Still, results showed that the binary model - the model that achieved the best results - does not tap the full potential and that outliers could be further reduced by improving modelling conditions. Apart from improving modelling conditions, there are several other aspects that require further investigation. Factors other than knock-on delays might influence delay propagation such as weather, peak hour overloads or temporary restrictive factors like construction sites. Another aspect that might be interesting to investigate is increasing the prediction horizon of delay estimates, i.e. predicting delays further than one station in the future. The earlier a delay is known, the more convenient. Making accurate predictions as far as possible into the future would therefore be desirable.

Figure 9: Model Improvement in Zürich Enge



One last thing to keep in mind is that the model was tested for one specific case study only. While the model's prediction is promising for the case study discussed in this thesis, its interoperability is limited. Knock-on delays are modelled for a simple case only. More complex knock-on situations would need more thorough investigation.

In conclusion, it can be said that even in a reliable network as the Swiss railway network, already a standard linear regression can improve delay prediction quality significantly. Approaches modelling knock-on delays furthermore show great potential and can increase prediction quality again significantly under suitable modelling conditions.

6 References

Corman, F. and P. Kecman (2018) Stochastic prediction of train delays in real-time using bayesian networks.

Daamen, W., R. M. P. Goverde and I. A. Hansen (2008) Non-discriminatory automatic registration of knock-on train delays.

LITRA (2022) Warum 93% aller züge der sbb pünktlich sind, https://litra.ch/de/aktuelles/blog-warum-93-aller-zuge-der-sbb-punktlich-sind/. Last access 25.05.2022.

Pongnumkul, S., T. Pechprasarn, N. Kunaseth and K. Chaipah (2014) Improving arrival time prediction of thailand's passenger trains using historical travel times.

SBB (2022a) Gemeinsam unterwegs in eine klimafreundliche zukunft, https://company.sbb.ch/de/ueber-die-sbb/verantwortung/nachhaltigkeit.html. Last access 25.05.2022.

SBB (2022b) Pünktlich für sie unterwegs, https://company.sbb.ch/de/ueber-die-sbb/verantwortung/die-sbb-und-ihre-kunden/puenktlichkeit.html. Last access 25.05.2022.

Schweiz, O.-D.-P. M. (2022) Open-data-plattform mobilität schweiz, https://opentransportdata.swiss/de/dataset/istdaten. Last access 16.05.2022.

Spanninger, T., A. Trivella and F. Corman (2020) Approaches for real-time train delay prediction.

statsmodels (2022) statsmodels, https://www.statsmodels.org/stable/index.html. Last access 30.05.2022.

Walker, D., B. Adamek and U. Stückelberger (2018) Fakten und argumente zum öffentlichen verkehr der schweiz.

Wang, R. and D. B. Work (2015) Data driven approaches for passenger train delay estimation.

Webster, M. (2022) Definition of knock-on effect, https://www.merriam-webster.com/dictionary/knock-on-effect. Last access 13.05.2022.

A Appendix

A.1 Variables Contained in the Data Set

- RUN_ID: individual ID for each run in the dataset
- EVENT_ID: individual ID for each event meaning each stop in the dataset
- DATE: date on which the run happened (yyyy-mm-dd)
- JOURNEY_ID: same function as RUN_ID but different format
- OP_NAME: name of the operator, e.g. SBB
- TRAIN_TYPE: type of the train, e.g. IC
- LINE_TEXT: specific name of the train, e.g. IC3
- START_STOP: start stop of the run
- END_STOP: end stop of the run
- STATION: name of the station at which the delay was measured
- STATION_ID: ID of the station at which the delay was measured
- STATION_NR: number of the station in the course of the run
- EVENT_TYPE: departure or arrival
- TIME_PLANNED: planned arrival or departure time
- TIME_REAL: actual arrival or departure time
- DELAY: delay given as difference between TIME_PLANNED and TIME_REAL
- DIR: direction of travel, either up (South-North) or down (North-South)

A.2 Schedule of the Train Services

Table 14: Routes of the S2, the S8 and the S24 service

S2	S8	S24
Zürich Flughafen	Winterthur	Thayngen
Zürich Oerlikon	Effrektikon	Herblingen
Zürich HB	Dietlikon	Schaffhausen
Zürich Wiedikon	Wallisellen	Neuhausen
Zürich Enge	Zürich Oerlikon	Andelfingen
Thalwil	Zürich HB	Winterthur
Horgen	Zürich Wiedikon	Weinfelden
Wädenswil	Zürich Enge	Märstetten
Richterswil	Zürich Wollishofen	Müllheim-Wigoltingen
Pfäffikon Sz	Kilchberg	Hüttlingen-Mettendorf
ALtendorf Sz	Rüschlikon	Felben-Wellhausen
Lachen Sz	Thalwil	Frauenfeld
Siebnen-Wangen	Oberrieden	Islikon
Schübelbach-Buttikon	Horgen	Rickenbach-Attikon
Reichenburg	Au ZH	Wiesendangen
Bilten	Wädenswil	Oberwinterthur
Ziegelbrücke	Richterswil	Winterthur
Unterterzen	Bäch	Kempthal
	Freienbach SBB	Effretikon
	Pfäffikon Sz	Bassersdorf
		Zürich Flughafen
		Zürich Oerlikon
		Zürich Wipkingen
		Zürich HB
		Zürich Wiedikon
		Zürich Enge
		Zürich Wollishofen
		Kilchberg
		Rüschlikon
		Thalwil
		Oberrieden Dorf
		Horgen Oberdorf
		Baar
		Zug

A.3 Trained Parameters for Different p Values

Figure 10: Trained parameters of the standard model for p=1

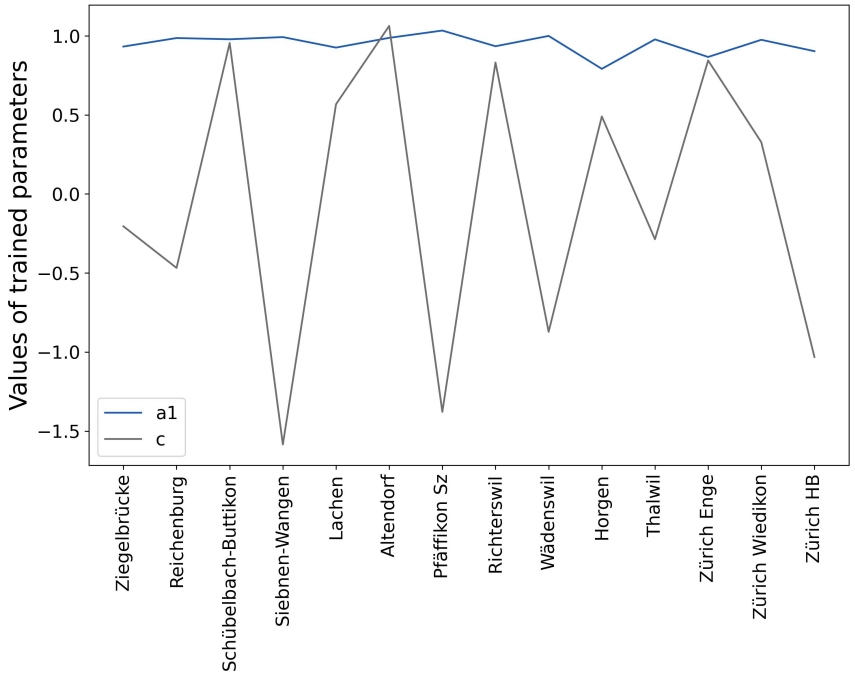


Figure 11: Trained parameters of the standard model for p=2