Demand Data for Dynamic Passenger Assignment within the Swiss National Rail Model

Livio Kaeslin, SBB Personenverkehr
Johannes Lieberherr, SBB Personenverkehr
Wolfgang Scherr, SBB Personenverkehr

Conference paper STRC 2014

STRC 14th Swiss Transport Research Conference
Monte Verità / Ascona, May 14-16,2014

Demand Data for Dynamic Passenger Assignment within the Swiss National Rail Model

Livio Kaeslin
SBB Personenverkehr
CH-3000 Bern

+41 79 592 41 68
livio.kaeslin@sbb.ch

Johannes Lieberherr
SBB Personenverkehr
CH-3000 Bern

+41 79 519 71 36
johannes.lieberherr@sbb.ch

Wolfgang Scherr
SBB Personenverkehr
CH-3000 Bern

+41 79 944 95 63
wolfgang.scherr@sbb.ch

25. April 2014

Abstract

In order to develop the time-of-day distribution functions of demand for a dynamic passenger assignment model we implemented a local regression approach with 6th degree polynomial correction terms on top of an initial smoothing. The major achievement in comparison to global regression approaches is that we better replicate the observed passenger volumes not only during the peak hours, but also for the off-peak period, which in earlier regressions (without the polynomial local regression) had been underestimated systematically throughout the Swiss national rail model. The chosen local approach is highly flexible and thus enables the control of extreme corrections.

Keywords

travel model – dynamic assignment – public transport – railway – demand data – local regression – time distribution

1. Introduction

SBB applies a passenger assignment model on a national scale, that is used in mid-term and long-term planning of rail service and timetables. Over the past few years, rail demand has rapidly increased in Switzerland and capacity constraints have appeared. Consequently, it is required that the SBB passenger models can explain and predict capacity constraints which appear typically at peak hours during the day. To respond to these new requirements, the assignment model was extended from a static, average-day model to a timetable-based, dynamic model which takes capacity restraints into account. Hence, also the input data of the assignment – demand and supply – need to be dynamic in nature. On the demand side this means that input data need to contain the information *how many* passengers want to travel *when* on *which origin-destination pair*.

From data mining in the large passenger travel data bases that originate from national fare clearing surveys we develop a highly-stratified demand model, with a fine segmentation of origin-destination pairs. For each origin-destination pair we estimate time-of-day distributions of demand representing the desired departure time of passengers.

It is important to emulate passenger volumes not only during the peak hours but also during the off-peak period. That is because capacity decisions (e.g. “how many rail coaches or train units will be used at which service frequency during off-peak period”) are based on the demand forecast from the model also in the off-peak period. In other words it is important to emulate the demand well over the entire day.

This paper is organized as follows. In chapter 2 we introduce our national rail model “SIMBA”, with network and assignment. Chapter 3 describes the development of the demand input for the dynamic assignment model with the local polynomial regression approach. Some artefacts and control mechanisms are discussed and illustrated. Chapter 4 provides a validation of the assignment results in the Swiss national rail model 2012 with count data and a comparison with an earlier used approach. In chapter 5, we draw our conclusions and share perspectives of further model development.

2. The Swiss National Rail Model "SIMBA"

For over 15 years, the passenger division of SBB maintains a passenger travel model, called "SIMBA[1]". SIMBA is applied to forecast passenger demand for future rail infrastructure, future timetables and service plans. In this section, we give a brief introduction to this model in three parts network, assignment and demand. Figures 1 and 2 show examples of model outputs from SIMBA.

2.1 Network Model

The public transport supply in our model consists of two networks: The network of links and the network of lines.

The network of links is based on the link database of SBB Infrastructure and contains the information length, owner, country, … In addition to the rail network there are complementary model links defined if there is regional public transport supply with buses. With those abstract bus supply links the demand can be distributed in a better way on the nearby rail stations.

The lines in our model contain information about the operator, the departure time, the interval, operating time, travel time, journey time, stops, and stop time. In our models we have three different kinds of timetables.

1. Regular interval timetable. It contains the regular interval supply and is only applied in a previously used static assignment.
2. Extended regular interval timetable. In addition to the regular interval service, this extended timetable includes shortened train runs at the boundaries of the operating time and additional train runs during the peak period. This timetable provides realistic capacities and is therefore used for dynamic assignment.
3. Annual "HAFAS" timetable. This time-table is based on the real-world passenger information system. Every train is contained with exact departure time. Also on this version of the timetable, dynamic assignment is applied.

[1] Standardisierte Integrierte Modellierung und Bewertung von Angebotskonzepten

We can provide the first two kinds of timetables not only for the existing state but in a consistent way also for the mid-term and long-term future. The HAFAS timetable however is available only for actual and historic states and cannot be used for long-term planning.

2.2 Dynamic Assignment Model

We use a timetable-based assignment. The assignment is performed in the software package Visum, with some complementary python scripts. The set of reasonable connections are found by a branch-and-bound algorithm as described in Friedrich (2001). The impedance of every connection is measured by the usual dimensions as in-vehicle time, number of interchanges, adaption time, … In SIMBA we use a dynamic assignment without and one with capacity restraints. For the latter one we use a penalty function depending on the load factor of the vehicles. The chosen penalty function is described in Lieberherr (2012). In the same paper the algorithm of the capacity restraint function is explained and convergence to an equilibrium state is shown.

While a timetable-based assignment algorithm is dynamic in its nature, it is rare in real-world modeling practice, that passenger models are calibrated and applied in a truly dynamic way, i.e. to forecast changes in passenger volumes over a 24-hour period. One example is Scherr (2009). In SIMBA, we apply the dynamic assignment in 10-minute intervals over a 24-hour day, with two cases: average weekday and average weekend-day. The route-choice impedance is Box-Cox-transformed and connection choice is modelled by a LOGIT model. All parameters in the assignment models has been estimated with stated preference data (e.g. Lieberherr (2009)) or optimized by comparison with counts and revealed preference data (e.g. Lieberherr (2014)) such that the modeled volumes are sufficiently accurate for a train-section level analysis.

2.3 Passenger Demand Model

For our dynamic assignment the demand is given by 144 OD-matrices, one per each 10 minute-slice of a 24-hour day. The development of these matrices will be discussed in the next chapters. These matrices are stratified uniformly in 10-minute slices. That allows us to extract the adaption time which we need for our impedance function.

Present-state demand is developed based on OD-survey data. To forecast demand changes we use an elasticity model, which allows to forecast passenger demand in reaction to variations in railway supply as well as in reaction to exogenous factors such as population growth, economic growth, and supply in car-transport.

Figure 1 SIMBA network model with passenger volumes from dynamic assignment

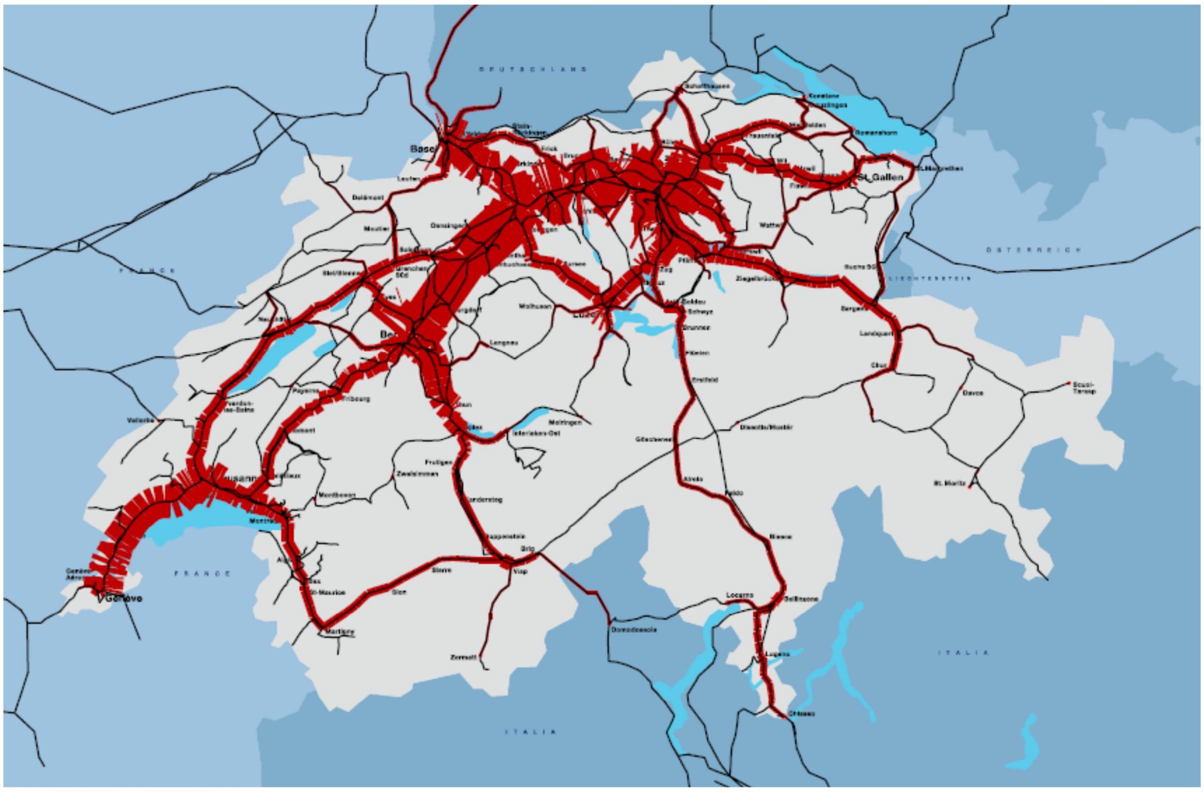


Figure 2 Use of dynamic assignment in a planning study – station-to-station passenger volumes per train and volume/capacity ratios.

link : from station >> to station	train departure time at first station															
	06:04	06:19	06:34	06:49	07:04	07:19	07:34	07:49	08:04	08:19	08:34	08:49	09:04	09:19	09:34	09:49
CO >> TA	0	1	3	6	17	26	40	38	64	37	51	13	14	6	7	3
TA >> MI	1	2	6	9	21	33	46	45	68	40	53	15	15	8	8	4
MI >> PT	2	3	9	12	27	38	52	51	76	46	56	19	19	10	9	5
PT >> VS	20	30	56	73	118	127	141	133	144	110	105	64	55	38	33	22
VS >> CR	30	47	91	123	197	215	234	226	218	192	144	110	87	62	50	39
CR >> GD	31	50	95	127	204	224	242	233	227	201	152	115	91	64	51	40
GD >> TU	34	59	107	144	229	264	274	252	252	217	160	125	104	70	55	44
TU >> CH	36	61	113	149	239	273	293	260	263	223	164	127	105	71	56	45
CH >> SE	37	64	117	154	245	283	311	268	272	230	167	131	108	74	59	50
SE >> GE	37	62	111	144	219	245	255	248	236	214	145	123	91	70	54	49
GE >> LA			158	232	303	296	308	288	245	156						
LA >> BA			132	203	253	260	257	255	205	137						
BA >> CH			137	215	259	277	267	269	210	146						
CH >> EV			96	165	181	204	187	212	146	107						
EV >> CH			65	101	123	132	128	140	100	70						
CH >> AN			26	40	51	52	51	71	41	27						

Legend:

- < 25% seat capacity used
- 50% seat capacity in use
- 100% seat capacity in use
- 100% seat and 50% standy capacity in use
- Total capacity (seat & standy) in use

3. Demand Data Input

The demand data for dynamic assignment consist of 144 demand matrices, one for every ten minutes of the day. Thus we need two elements:

- Time-distributions of the desired departure time, which is specific for each OD-pair. In short, we call this "time-of-day distributions" in this paper.
- OD-matrices (number of passengers who want to travel), potentially disaggregated by trip purpose.

In this section of the paper, we describe how time-of-day distributions are developed in three steps:

1. Raw distributions obtained from survey data
2. Smoothing
3. Peak correction:

The 3rd step is the focus of this paper: A global regression approach has been replaced by a local regression method to obtain the correction factors.

3.1 Data Context

The empirical basis of our demand data is the national fare clearing survey. The sample has seven million data records each year. After having extrapolated and calibrated the survey data, we add departure time and trip purpose to each data record. We aggregate the survey data such that we obtain the demand for each OD-pair and each minute of the day. The underlying time-distribution of the observed demand in the survey has strong peaks, which are caused by the exact departure times of trains in the existing timetable. Thus for each OD-pair there is an 1440-dimensional demand vector (with an entry for every minute of the day). We aim to get 144 demand matrices, one for every ten minutes of the day, as data input for the dynamic assignment in SIMBA, the Swiss national rail model. To accomplish that, we want to prepare the demand for each OD-pair: For each OD-pair we need a 144-dimensional demand vector with an entry for every 10-minutes. This demand vector shall represent the times desired by the passengers and thus should be independent of the train departure times in the existing timetable. At the same time this vector shall replicate the highs and lows of passenger demand. The first step to reach this goal is to smooth the demand curve. The smoothing eliminates the highly timetable-driven peaks but also flattens the planning relevant peak hour demand. After the smoothing there will be a need for a correction such that the peak hours and off-peak hours are properly emulated.

3.2 Previous Procedure

3.2.1 Smoothing

We apply to the 1440-dimensional demand-vector of each OD-pair a simple triangulation smoothing. Let $D(i)$ be the demand at minute $i \in \{1, \ldots, 1440\}$ and n the width of the triangle. We get the smoothened demand $D_{smooth}(i)$ at minute i as follows

$$D_{smooth}(i) = \sum_{j=i-n}^{i+n} D(i) \left(\frac{1}{n} - |j - i| \cdot \frac{1}{n^2} \right)$$

In Lieberherr (2008) there were studied different triangle-widths and the width $n = 60$ was considered the most appropriate. Thus we chose $n = 60$ for our smoothing.

The off-peak hours look very good after the smoothing, but an unwanted side effect is that the peak hours get too flat. Therefore we apply a peak correction which is determined by regression.

3.2.2 Regression

Already in Lieberherr (2008) there is a regression proposed which tries to emulate the peak hours, but systematically underestimates the off-peak hours. It is a global regression, i.e. a regression that affects the time distribution as a whole.

3.3 Revision of the Procedure

To better emulate the off-peak hours and still emulate the peak hours well or even better, in general it is possible to revise the smoothing or the regression. Two necessary conditions have to be fulfilled:

1. For each OD-pair the total demand over the whole day must not change.
2. For each OD-pair the demand in the unsmoothed peak hour should equal the demand in the peak hour after the regression.

Several smoothing methods have been considered. In Shumway (2011) in chapter "Smoothing in the Time Series Context", several smoothing methods in the time series context are proposed. A smoothing method that keeps the peak hour demand would be nice such that the problematic subsequent regression would not be needed anymore. If such a smoothing cannot be found, the regression will have to be revised.

All of the proposed smoothing methods found in the literature do not solve the problem. Generally we can say that a global smoothing always flattens the peaks. But there are also non-global smoothing methods proposed like the “Smoothing Splines”, which smooths the curve interval-wise. The Problem with that method is that the total demand will not necessarily remain the same. That we cannot accept. Another solution could be using “Kernel Smoothing” interval-wise. But it would lead to discontinuity at the interval boundaries. That we cannot accept either, because we want the curve to be independent of the time-table and that requires a smooth curve.

After this analysis of smoothing methods we realize that a new regression method is required. We know that off-peak hours are nicely emulated after the smoothing. Thus our approach will be a local regression around the peak hour which leaves the off-peak period curve unaffected.

3.3.1 New Regression Method

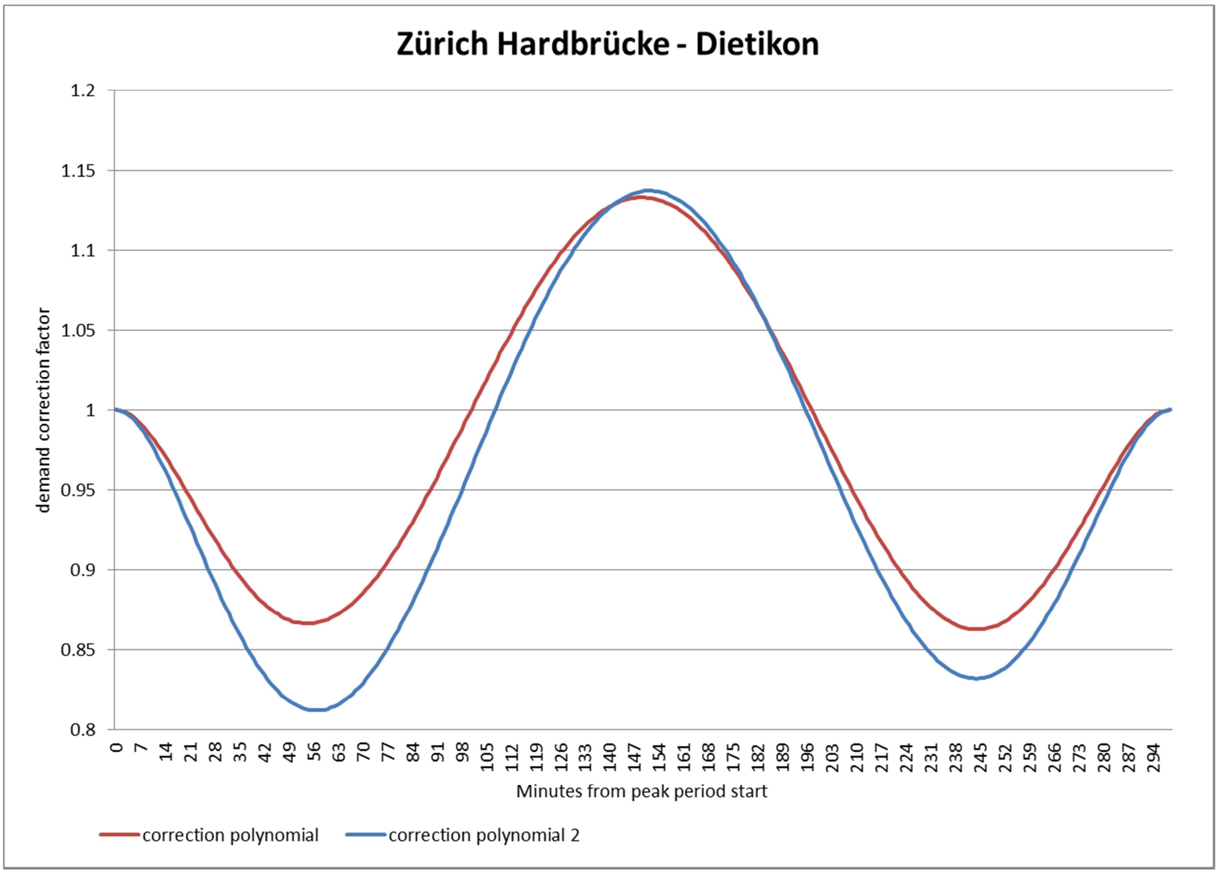
With a local regression around the peak hour we want to move demand from around the peak hour into the peak hour such that the demand during the peak hour equals the unsmoothed peak hour demand. We consider an approach with a polynomial correction term function with values in $[0, \infty[$.

The following conditions should be hold in such a regression:

1. The demand in the unsmoothed peak hour should equal the demand in the peak hour after the regression.
2. The total demand during the interval affected by the local regression should remain constant.
3. The correction function should be 1 at the boundaries (two conditions).
4. The derivative of the correction function should be 0 at the boundaries (two conditions).

Knowing the shape of polynomials, we consider appropriate a polynomial correction function of 6^{th} degree, because it can have five local extrema: One maximum in the middle to emulate the peak hour demand and one minimum on each side to compensate the demand used in the middle. The other two extrema are used at the boundaries to fulfil derivative conditions.

Figure 3 Two polynomial correction function of 6th degree with five local extrema each



We now explain which time interval should be affected by the local regression. If the time interval is defined too short, there will be almost no demand on the sides of the peak hours to compensate the additional demand needed during the peak hour. Furthermore the time interval should contain the whole peak until it gets flat again. That is because the peak as a whole suffers under the flattening caused by the smoothing and for that needs regression. If the time interval is defined too long, there will be not enough off-peak period left (which is well emulated by the smoothened curve) and the regression becomes almost global. We found that five hours are the most efficient interval length for this purpose.

We look for a polynomial function h of 6th degree, i.e. $h: \mathbb{R} \to \mathbb{R}$ with

$$h(x) = a \cdot x^6 + b \cdot x^5 + c \cdot x^4 + d \cdot x^3 + e \cdot x^2 + f \cdot x + g, \text{where } a, b, c, d, e, f, g \in \mathbb{R}.$$

On a closed interval $[m, n]$ with $m, n \in \mathbb{R}$ and $m < n$ the function $h_{|[m,n]}: [m, n] \to [0, \infty[$ will return positive correction factors as values of the function at integer points. We choose $m = 0$ and n$= 299$. Each number in $\{0, 1, \ldots, 299\}$ represents one minute in a 5-hour time interval around the smoothened peak hour. Thus the peak correction will be applied for the entire 5-hour time interval, with the 300 correction factors contained in $h([m,n] \cap \mathbb{N}_0)$.

The minute s_0 is the starting minute and minute $s_{299} = s_0 + 299$ is the end of the peak period, the smoothened peak hour lays between minute $s_{120} = s_0 + 120$ and minute $s_{179} = s_{120} + 59$.

The correction function h will be applied as follows to obtain the demand after regression $D_{reg}(s)$ at minute s as follows

$$D_{reg}(s) = \begin{cases} h(s - s_0) \cdot D_{smooth}(s), & \text{if } s_0 \leq s \leq s_{299} \\ D_{smooth}(s), & \text{else} \end{cases}.$$

We need seven conditions to determine the polynomial function. Six conditions are already written above. The second condition can be split into two conditions: The remaining demand after having emulated the peak hour is distributed on the right and the left side of the peak hour proportionately to the prior smoothened demand on both sides. Like that we have the certainty that the asymmetry of the peak period is emulated well.

Now we have to translate the above conditions into mathematical equations

1. Suppose the unsmoothed peak hour lays between minute us_{120} and minute $us_{179} = us_{120} + 59 = us_0 + 179$ and the peak hour after the regression lays between minute s_{120} and minute s_{179}. The first condition means
$$\sum_{i=120}^{179} D(us_0 + i) = \sum_{i=120}^{179} D_{reg}(s_0 + i),$$
where $D_{reg}(s_0 + i) = D_{smooth}(s_0 + i) \cdot h(i)$ for $i \in \{0, 1, \ldots, 299\}$.

2. The second condition means
$$\sum_{i=0}^{299} D_{smooth}(s_0 + i) = \sum_{i=0}^{299} D_{reg}(s_0 + i).$$
and can be split into (as explained above)
$$\sum_{i=0}^{119} D_{reg}(s_0 + i) = \left(\sum_{i=0}^{299} D_{smooth}(s_0 + i) - \sum_{i=120}^{179} D(us_0 + i)\right) \cdot p_1,$$
$$\text{where } p_1 = \frac{\sum_{i=0}^{119} D_{smooth}(i)}{\sum_{i=0}^{119} D_{smooth}(i) + \sum_{i=180}^{299} D_{smooth}(i)}$$
and

$$\sum_{i=180}^{299} D_{reg}(s_0+i) = \left(\sum_{i=0}^{299} D_{smooth}(s_0+i) - \sum_{i=120}^{179} D(us_0+i)\right)\cdot p_2,$$

$$\text{where } p_2 = \frac{\sum_{i=180}^{299} D_{smooth}(i)}{\sum_{i=0}^{119} D_{smooth}(i) + \sum_{i=180}^{299} D_{smooth}(i)}.$$

3. Boundary conditions:
$$h(0) = 1 \text{ and } h(299) = 1.$$
4. Derivative boundary conditions:
$$h'(0) = 0 \text{ and } h'(299) = 0.$$

If the peak period is overlapping midnight one has to be careful because the regression has to be treated especially with modulo operations. But the idea remains exactly the same.

For each OD-pair and each peak hour the polynomial correction term function will be constructed as follows

1. We search the peak hour in the unsmoothed demand curve, i.e. the consecutive 60 minutes with the most demand. The total of peak demand is stored.
2. We search the peak hour in the smoothened demand curve similarly and store the total as well.
3. We determine the demand in the smoothened peak period and calculate the remaining demand for the peak period by subtracting the unsmoothed peak hour demand.
4. We distribute the calculated remaining peak period demand proportionately to the prior smoothened demand on both sides of the peak hour.
5. The above conditions now have concrete values and the parameters of the polynomial function h can be obtained by some matrix-multiplications, which solve the linear equation system, given by the conditions.

Classically there are two peaks: One in the morning (AM) and one in the evening (PM). Both are important for us. So we would like to apply the regression method to both of them.

If the first peak hour starts AM then another peak hour is searched PM and vice versa. If the two peak periods do not overlap, the regression will be applied to both. Otherwise just the strongest peak period (which contains the consecutive 60 minutes with the most demand) will be handled in the regression step.

If an OD pair has two peaks, in AM and PM, and with similar magnitude, it is theoretically possible that the unsmoothed peak hour is in the AM, but the smoothened peak hour is found in the PM or vice versa. This case is prohibited by our algorithm.

There are two reasons, why we require a demand of at least 50 passengers per average weekday for the regression:

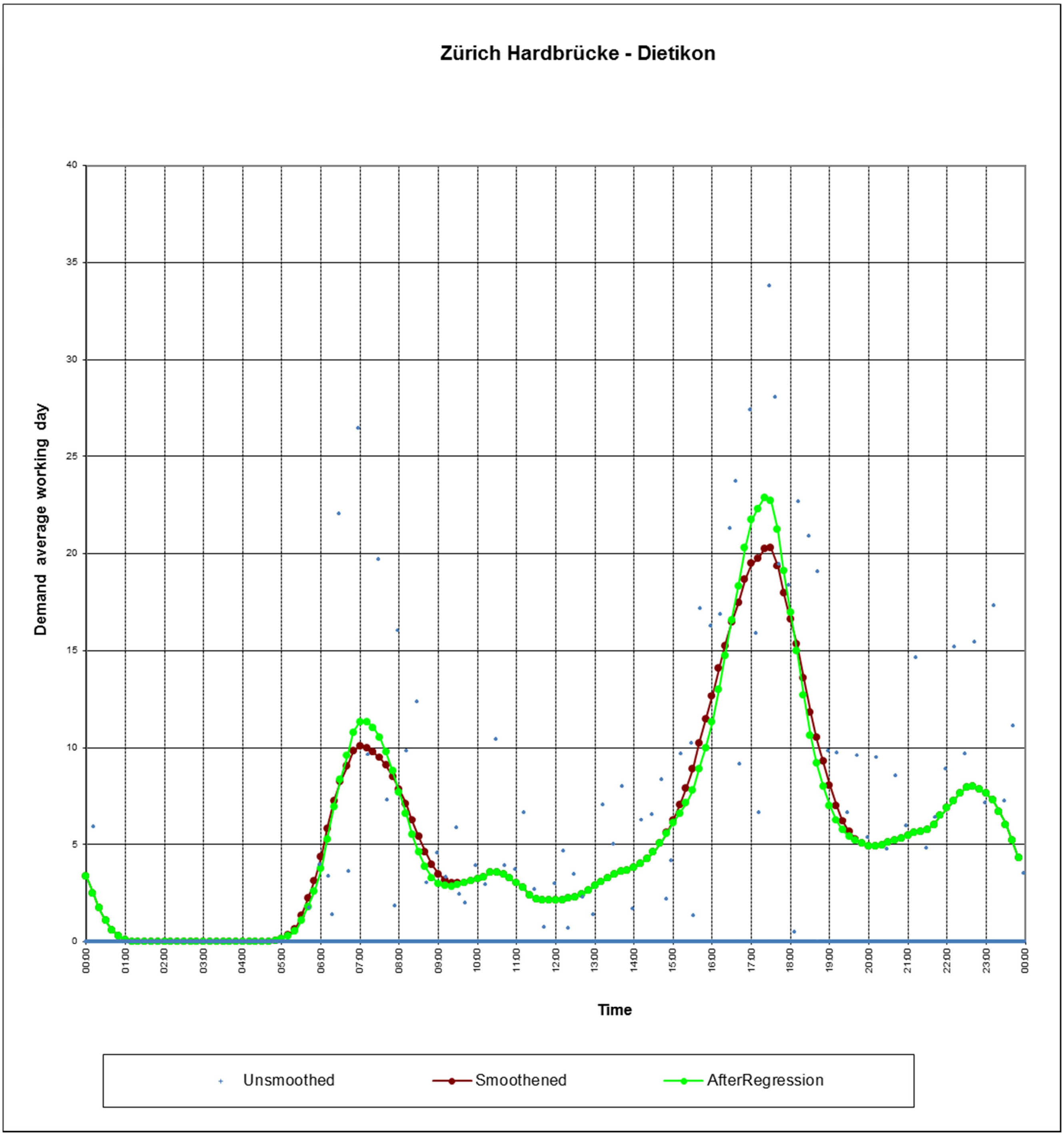
- Performance: There are in total 210000 OD-pairs with non-zero demand on average working day in the Swiss national rail model. It just takes too long to apply the polynomial regression to all OD-pairs. there are around 5000 OD-pairs with above 50 passengers, but these 5000 cover 76% of the total demand.
- Infrequently travelled OD-pairs show low sample sizes in the survey, that are insufficient to determine realistic time distributions.

The bound of 50 persons per average working day is not well studied, but chosen intuitively.

3.3.2 Examples

For the OD-pair Zürich Hardbrücke – Dietikon we find the peak hour in the unsmoothed demand curve in the morning between 16:43pm and 17:42pm with a total demand of 131.2 persons per average working day. The peak hour in the smoothened demand lays between 16:52pm and 17:51pm with a total demand of 117.9 persons per average working day. Thus the considered peak period will be between 14:52pm and 19:51pm. Some linear algebra leads us to the correction polynomial in Figure 3 for the first peak period. The smoothened AM peak hour is found between 6:45am and 7:44am. A second regression will be applied in this second peak period. The correction polynomial 2 for this peak period is also shown in Figure 3. In Figure 4 the whole demand curve of the OD-pair is shown before and after regression. The unsmoothed demand of this OD-pair is also shown. The demand curve is actually a 144 dimensional demand vector with an entry every ten minutes.

Figure 4 Demand curve Zürich Hardbrücke-Dietikon unsmoothed, smoothened and after regression



3.3.3 Control Mechanisms

This local regression approach is highly configurable and thus different control mechanisms can be applied locally. In this chapter there shall be given some examples.

For a given OD-pair we want to know, when do how many persons want to leave if there would be services at any minute. But the demand in our data is always influenced by the timetable. The question is now which influence of the timetable is too strong and thus has to be considered an artefact. For example the exact departure time is considered an artefact and thus we smooth the demand curve.

Further artefacts can be:

- temporal moving in the timetable of single trains in a train line, e.g. a train in the morning leaves one minute later than usually during the day: Like that two trains of this hourly service may lay within 60 minutes.
- Additional international trains might raise the demand within one hour more than we want.
- Too few services such that persons cannot choose the hour when they want to travel.

It is a difficult question, which effects are to be considered an artefact. But it is clear that there are timetable artefacts as the ones described above. To reduce the influence of such artefacts we apply a cap on the correction polynomial: If the unsmoothed peak hour demand is more than 50% of the 5-hour peak period demand, the ratio of the peak hour demand is considered too extreme and probably influenced by artefacts. That is why we then only emulate the 50% of the peak period demand during the peak hour. But never we lower the peak hour demand due regression, i.e. if there is no maximum in the middle of the correction polynomial we will not apply it. In 28.6% of OD-pairs with demand greater than 50 passengers per day this control mechanism triggers in the main peak period in the Swiss national rail model 2012. Like that 2.65% of the total peak hour demand is capped. The median ratio of the unsmoothed peak hour demand in the 5 hour peak period is 43.1%.

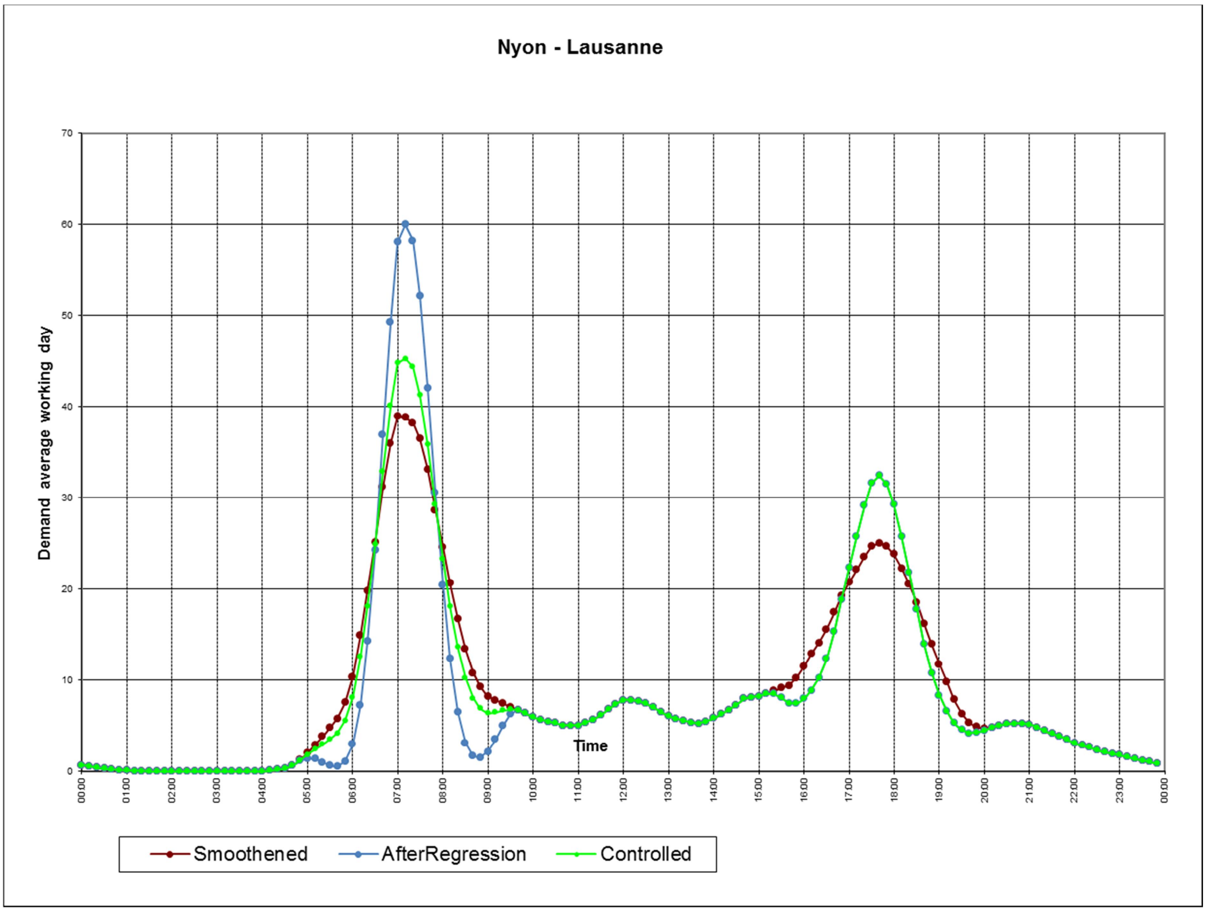
Table 1 Totally capped peak hour demand with different caps

Cap	50%	55%	60%
Capped peak hour demand	2.65%	1.69%	1.08%

Another applied control mechanism is: If the minimum of the correction polynomial is less than 0.5, the correction is considered too extreme and thus influenced by artefacts and the regression will not be applied. With a cap of 50% in the Swiss national rail model 2012 in 133 OD-pairs that control mechanism has triggered.

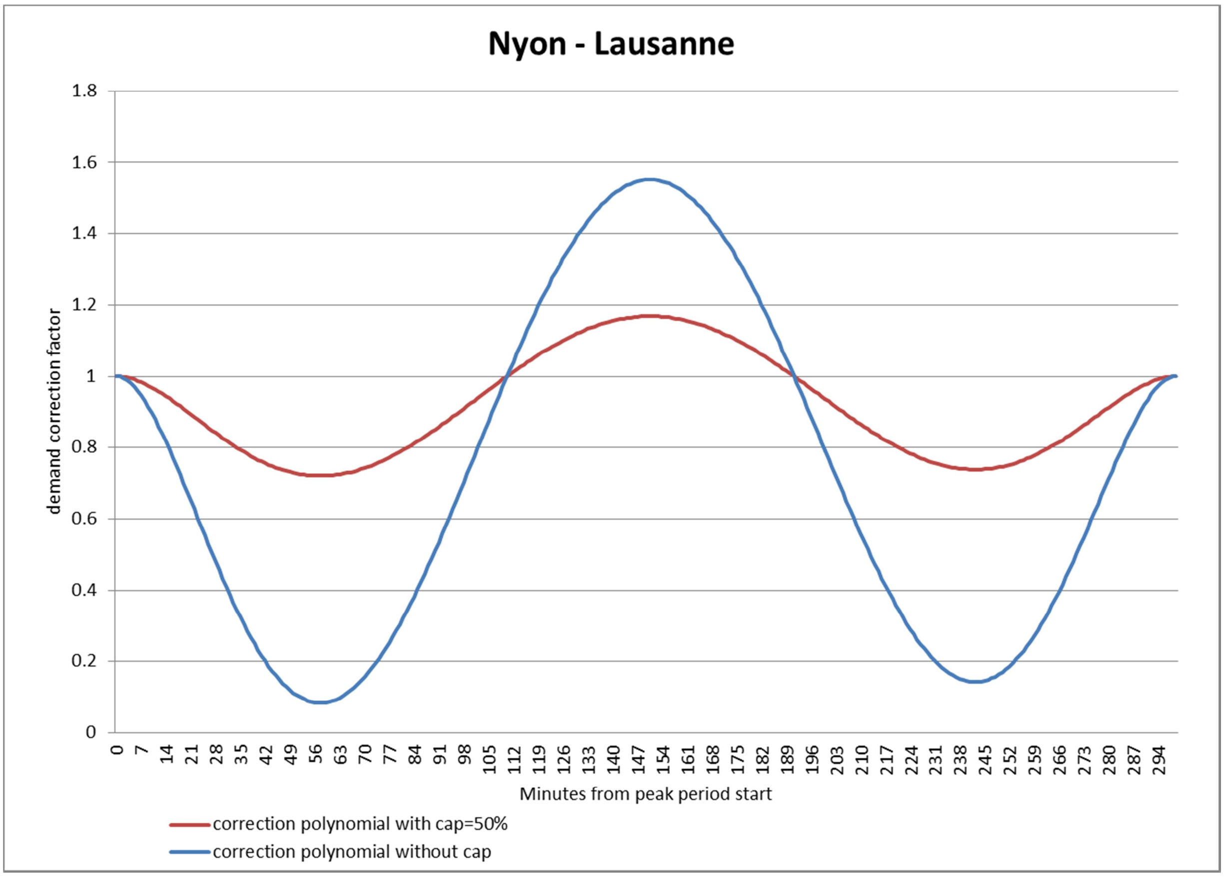
Those control mechanism will be illustrated with the OD-pair (Nyon, Lausanne). In this OD-pair the peak hour demand ratio is 64% of the peak period and thus the correction polynomial without control mechanisms is quite extreme. See the correction polynomial in Figure 6. This extreme correction can also be seen in the demand curve after regression (See Figure 5). The demand next to the AM peak hour is almost fully taken away to compensate the very strong peak hour correction. That we do not want.

Figure 5 Demand curve Nyon-Lausanne before (red) and after regression without control mechanisms (blue) and with control mechanisms (green)



The implied control mechanisms trigger in this example. Thus with the cap of 50% just 50% instead of 64% of the peak period demand will be emulated during the peak hour. This leads to a less extreme correction polynomial which does not trigger the second control mechanism anymore, because the minimum of the correction polynomial is now greater than 0.5. This correction polynomial is also shown in Figure 6. Also the controlled demand curve after regression in Figure 5 looks less extreme and thus more realistic after having applied the control mechanisms.

Figure 6 correction polynomial of the peak period of OD-pair Nyon-Lausanne with (red) and without (blue) control mechanisms



How has it come that in this OD-pair the peak hour demand ratio is 64% of the peak period? Is there an explainable artefact? Considering the timetable of 2012 we find the artefact quickly: The InterRegio (IR) train from Geneva to Lausanne and Brig normally departs in Nyon at XX:50 and XX:27 but in the morning there is an exception and the train leaves at 06:50am, 07:10am and 07:47am (instead of 07:50am, see Figure 7). This increased concentration of trains causes an unnatural attraction for passengers to travel within the 60 minutes. This is clearly a timetable artefact and justifies the use of control mechanisms.

Figure 7 The timetable artefact is caused by train IR1717 due to the departure time 07:47am in Nyon.

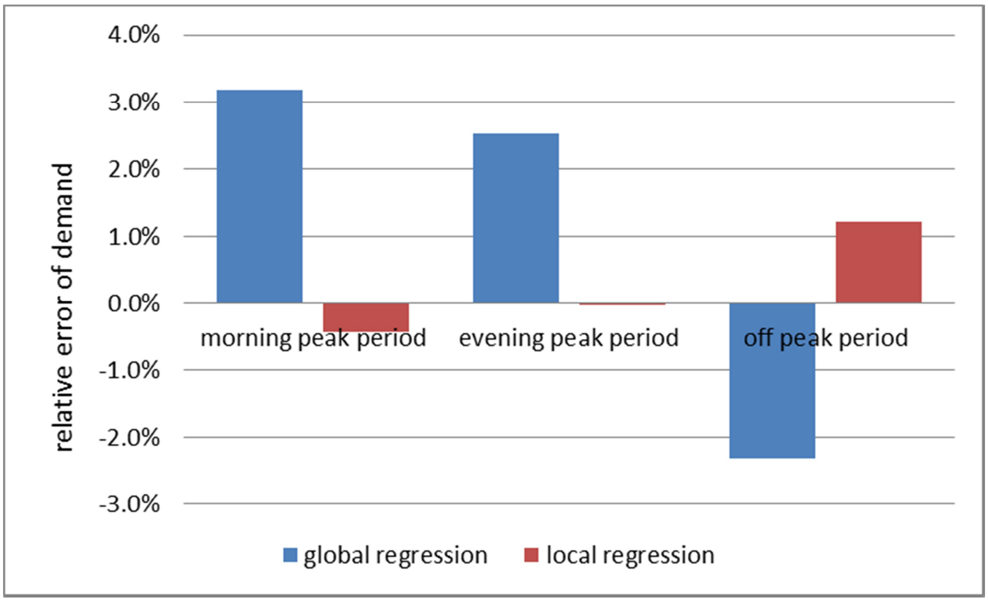
		RE 2615		IR 1715	IC 715 R FA	S3 12317	RE 2717	IR 1417		IR 2517	ICN 617	S4 12419	RE 2617	IR 1717	
Genève-Aéroport		611		627	636			647		701	705			724	
Genève	o	618		634	643			654		708	712			731	
				IR Genève-Aéroport-Lausanne-Brig	IC Genève-Aéroport-Lausanne-Bern-St. Gallen			IR Genève-Aéroport-Lausanne-Brig		IR Genève-Aéroport-Lausanne-Bern-Luzern	ICN Genève-Aéroport-Neuchâtel-Biel/Bienne-Basel SBB			IR Genève-Aéroport-Lausanne-Brig	
Genève 28		620	Ⓐ633	636	645		Ⓐ649	656	703	710	714		720	733	Ⓐ733
Genève-Sécheron			635						705						735
Chambésy			637						707						737
Les Tuileries			639						709						739
Genthod-Bellevue			640						710						740
Creux-de-Genthod			644						714						744
Versoix			646						716						746
Pont-Céard			647						717						747
Mies			650						720						750
Tannay			652						722						752
Coppet	o	630	Ⓐ656				658		726				729		Ⓐ756
Coppet		631					659						730		
Nyon 3	o	636		649			704	709			726		735	746	
Nyon		637		650			705	710			727		736	747	
Gland		642						715					741		
Rolle		648					714						747		
Allaman	o	652											751		
Allaman		653				708						Ⓐ739	752		
Etoy						710						741			
St-Prex						712						743			
Morges 3	o	659		704		716	723	727			740	747	758	801	
Morges		700		705		717	724	728			741	748	759	802	
Morges-St-Jean												749			
Lonay-Préverenges						719									
Denges-Echandens						721									
Renens VD 3		707				726	732	735				Ⓐ755	807		
Prilly-Malley						11 728						757			
Lausanne 4	o	712		715	718	732	Ⓐ737	740		743		Ⓐ801	812	812	

4. Validation with Passenger Counts

The local regression approach has been validated with count data for the year 2012. To validate, the resulting time-of-day distributions have been applied in a dynamic assignment in SIMBA and the obtained passenger volumes are compared to the counts.

During the morning peak period (departure time between 6:00 and 8:59), the evening peak period (departure time between 16:00 and 18:59), and also during the off-peak period (rest of the day) the results with the new local regression are better emulated than with the old global regression approach. This can be seen in Figure 8. With the earlier global regression approach peak periods where over-estimated and the off-peak period was under-estimated. The local regression approach emulates the peak periods very well and leads to a weak over-estimation of the off-peak period.

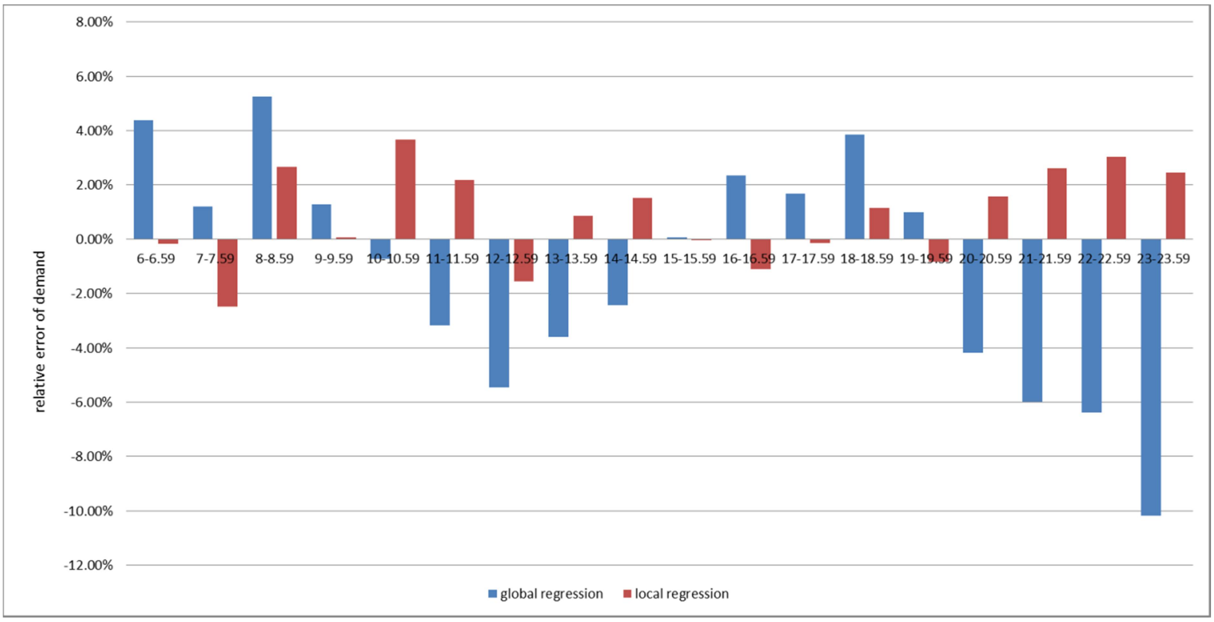
Figure 8 Relative error of person kilometers from counted data in SIMBA period-wise



The high configurability of the local regression approach helped us to better emulate the peak periods using control mechanisms. The off-peak period is better emulated with the new method because the regression does nearly not affect the off-peak periods and thus emulation mostly equals the smoothened demand curve which is good for off-peak periods.

In Figure 9, we show the relative error to counted person kilometers for each hour of the day. Overall, the new regression method is better than the old one. The underestimation of the model with local regression between 16:00 and 16:59 and between 19:00 and 19:59, can be interpreted that the correction polynomials compensated the PM peak hour demand mostly during those hours.

Figure 9 Relative error of person kilometers from counted data in SIMBA hour-wise.



5. Conclusions and Perspective

Over the past decades, SBB has been facing high growth rates in passenger demand. Consequently, capacity issues occur during peak hours and many projects aim to increase transport capacity in future timetables. In that context, it is imperative that the SBB travel model SIMBA, which is used to evaluate all future projects and timetables, represents the time dynamics of passenger demand. In particular, it is important that the model understands demand variations during the day and replicates well when and how peak loads occur today and in the future.

To better replicate the dynamic distribution of passengers during a 24-hour day, new distributions of desired departure time have been developed on the basis of empirical survey data, by a three-step method. Using local regression, correction factors are computed as 6th degree polynomial that are applied on top of the output from a basic smoothing step. Validation on passenger count data shows that this method has improved considerably peak period and off-peak period emulation.

SBB continues to improve our national travel model SIMBA. Among other projects of model development and model improvement, one current effort that builds on the work presented in this paper, aims to categorize the OD pairs and their time-of-day-distributions in order to find factors that explain the time-of-day distributions to enable us to forecast how time-of-day-distributions will change in the future.

6. References

Friedrich, M., Hofsäß, I., Wekeck, S., (2001). Timetable-based Transit Assignment Using Branch & Bound Techniques. Transportation Research Records, No. 1752, p. 100-107.

Lieberherr, J., (2008). Ganglinie pro Relation. SBB Personenverkehr, internal technical report, Bern.

Lieberherr, J, (2014). Tagesgangumlegung in SIMBA. internal report, Bern.

Lieberherr, J., Pritscher, E., (2012). Capacity-restraint railway transport assignment at SBB-Passenger. Conference paper, STRC 2012, Ascona.

Lieberherr, J., Riedi, M., (2009). Bericht zur HVZ-Studie, internal technical report, Bern.

Scherr, W. and Fisher, I. (2009). Regional Application of 24-hour Dynamic Transit Assignment. Conference paper, 12th TRB Transportation Planning Applications Conference, Houston, Texas, USA.

Shumway, R., Stoffer, D.,(2011). Time Series Regression and Exploratory Data Analysis. Springer, New York.