12-2014

The effect of infrastructure changes on railway operations.

Martin J.R. Schmidt 1966-
University of Louisville

THE EFFECT OF INFRASTRUCTURE CHANGES ON RAILWAY OPERATIONS

By

Martin J.R. Schmidt
Dipl. Ing. , Universität Hannover, 1993
MBA, Fachhochschule des Mittelstands, 2008

A Dissertation
submitted to the J.B. Speed School of
Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Industrial Engineering
University of Louisville
Louisville, Kentucky

December 2014

THE EFFECT OF INFRASTRUCTURE CHANGES ON RAILWAY OPERATIONS

By

Martin J.R. Schmidt
Dipl. Ing. , Universität Hannover, 1993
MBA, Fachhochschule des Mittelstands, 2008

A Dissertation Approved on

December 1, 2014

By the following Dissertation Committee:

____
Gail W.DePuy, Dissertation Director

____
William E. Biles, Committee Member

____
Gerald W. Evans, Committee Member

____
Thomas D. Rockaway, Committee Member

DEDICATION

This dissertation proposal is dedicated to my parents

Mr. Johannes W.A. Schmidt †

and

Mrs. Jutta I.R. Schmidt, geb. Borchardt,

who gave me all support for my education.

ABSTRACT

THE EFFECT OF INFRASTRUCTURE CHANGES

ON RAILWAY OPERATION

Martin J.R. Schmidt

December 1, 2014

This paper makes use of standard simulation programs in combination with the tools of applied statistics to simulate railway operations. The purpose of the use of this tool is to evaluate and compare different possible kinds of railway infrastructure, like different types of signaling procedures, different network configuration or operational procedures. A railway system is a logistic network and because of the demand for improved railway operation, much work has been undertaken lately in this scientific field. However the author postulates the hypothesis based on a literature review that in a lot of these works there is a lack of full application of statistics. With this paper the author makes use of standard simulation programs for detailed simulation of railway operation especially with respect to the signaling and operation procedures. Additionally the influence of delays, which occur during real life railway operation is taken into account for a first time. This allows statistical evaluation of the results based on statistical significance. Also sensitivity analysis could be performed. It is demonstrated, that the results of such simulation runs show superior results when compared to other techniques not taking into account the variability. Additionally, procedures were

developed to find the capacity of a railway network with the help of additional software tools.

In this work the software package ARENA is used to simulate the operation of trains in railway networks. For this approach two major obstacles have to be solved: the simulation of train travelling times and the simulation of block rules used in railway operation. By introduction of visualization the confidence in the results of simulation, even for stakeholders not familiar with this technique, is increased. In this paper it is shown that with ARENA it is possible to calculate the capacity of different railway networks (scenarios). The results, which are calculated using quasi steady state simulation without variation, are similar to those obtained with other calculation methods. Additionally in one scenario the rule of thumb for the quotient between theoretical capacity and practical capacity in a railway network is confirmed by simulation including random variation. It is also demonstrated that OptQuest, an additional software package available for ARENA, is a suitable tool to find near optimal timetables in a scenario including delays.

The results of this work may be not only of interest for railway operators, but also for operators of other automated transport systems. Such systems may be unmanned transport vehicles in a factory, transporting goods between different manufacturing stations. But also for automation of road traffic the results may be of interest.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 MOTIVATION

1.1 Operation of railway transport

Since more than 175 years the system of railway exists and has played a major role in development of nations. Today a lot of different other means of transportation across the nations and continents are used, like the road transport, aviation and the use of boats on artificial waterways and rivers. Nevertheless, in all these cases the goal is the same: to transport goods and persons over longer distances at the same time with an acceptable speed and for affordable costs. In all cases, the risks associated with the transportation must be sufficiently low, the risk for damaging or losing the transported goods and persons as well as the risk for the environment influenced by the transportation.

But different from some other inventions from more than one century ago, railway systems still exist worldwide and play a major role in supporting the logistic needs in a globalized economy. In some parts of the world there is actually a renaissance of railway transportation. This trend is influenced mainly by two factors, the increasing awareness in public about possible negative environmental impacts of the rising transportation needs and increasing prices for transportation. Due to some special characteristics of railway systems, this will be explained later on in more detail, the railway system is very efficient with respect to fuel consumption and area needed for operation.

The railroad is characterized by a transportation vehicle, suitable to house the goods and people which need transportation. This railway vehicle is moved along fixed routes with the help of tracks. These tracks were first made from wood, but since the loads increased soon after the start of railway transportation they are made out of steel. For the ease of movement on these rails, the vehicle has wheels also made out of metals. So the contact area between track and vehicle is a contact between two metals. An advantage of railroad is the fact, that the resistance against movement is extraordinary low between such a steel-wheel and a track made out of steel. So the vehicle can be moved very easily across the track, only with a friction of the power that would be necessary for transportation of the same load by road with the same speed. The coefficient of adhesion, which describes this physical phenomenon, is for the railway on average eight times less than in highway traffic (Hansen/ Pachl 2008 page 17). But there is also an obstacle behind this physical principle. The acceleration and deceleration of the vehicle must take into account for this low resistance otherwise there would be a strong tendency of the wheels to slip. If this slipping occurs, there is no longer a possibility to exchange the large forces between track and railcar necessary for the operation. Otherwise additional wear and tear will be a result. Additionally, the wheel must be held on the track, so that no derailing occurs. By special means in modern railway systems, like sophisticated track elements like automated signals and points it is guaranteed that such derailment occurs very seldom.

Another major characteristic of railway systems is the strong interconnection between the infrastructure (track and signaling as well as personal in the office and at the track like dispatchers and signalers) and the transport vehicle. The train can only use a defined infrastructure and the operation is strongly dependent on the exact matching of train operation as well as infrastructural operation. In road transport it is for example possible for the driver to change the transportation direction or chosen way during the travel himself, perhaps make a pause due to severe traffic or changing weather conditions. Contrary to this in most traditional railway systems the track is already fixed for most of the travel distance when the railway vehicle journey starts. For example in passenger service a timetable is published by the railway operator. This timetable informs the possible users of the trains, where and when there will be a possibility to travel to a given destination and how long the travel will take.

1.2 Terminology and visualization, fundamentals in railway operation

As railways have been in service for a long time, regional differences in railway operation procedures exist as well as differences between the different wordings from country to country, sometimes even different between the companies in the same country. But as the trend for globalization did not exclude the railway operation, there was also a strong countermovement to standardize the different wordings and use of terms amongst railway experts. So there have been in the past some efforts to standardize railway language. One has to mention the efforts of the International Union of Railways (Union Internationale des Chemins de fer, UIC), which published some dictionaries. Furthermore the expert groups within UIC, which is an organization out of several railway companies, initiated and published scientific works (in books and on several conferences as well) to further increase the exchange and harmonization of words and ideas between experts (for example Winter, P. ; Compendium on European Rail Traffic Management System ERTMS).

One example of such effort is the guide 406 capacity edited by UIC (UIC 2004), where some basic concepts of railway operation are described. Furthermore some basic definitions are laid down in this guide. One very basic concept is the introduction of capacity in this guide. Here also a basic calculation method of capacity consumption is given.

A very basic concept used in railway operation is to visualize the movement of trains across the railway network. Such visualization is called a train diagram. In the North American terminology, this is called a string line graph. On the one axis, the way is displayed, whereas on the other axis the time is drawn. On the axis with the way, the position of the stations is marked, so this axis is called the station axis. The actual positions of the trains are marked on this diagram, mostly connected as lines with changing angles according to the different speeds at which the trains travel along the tracks. As security of railway operation means, that in no case two trains are on a single track at a given time on the same position, it is very easy to detect conflicts in the time-way diagram. A nonconformance to this principle of one track with one train only at a given time is easily visible as the crossing of two train-lines in a single track layout. For a double track line, crossings may occur, but the trains with crossing lines must run on different tracks. Two examples of a way-time diagram are given in Figure 1.

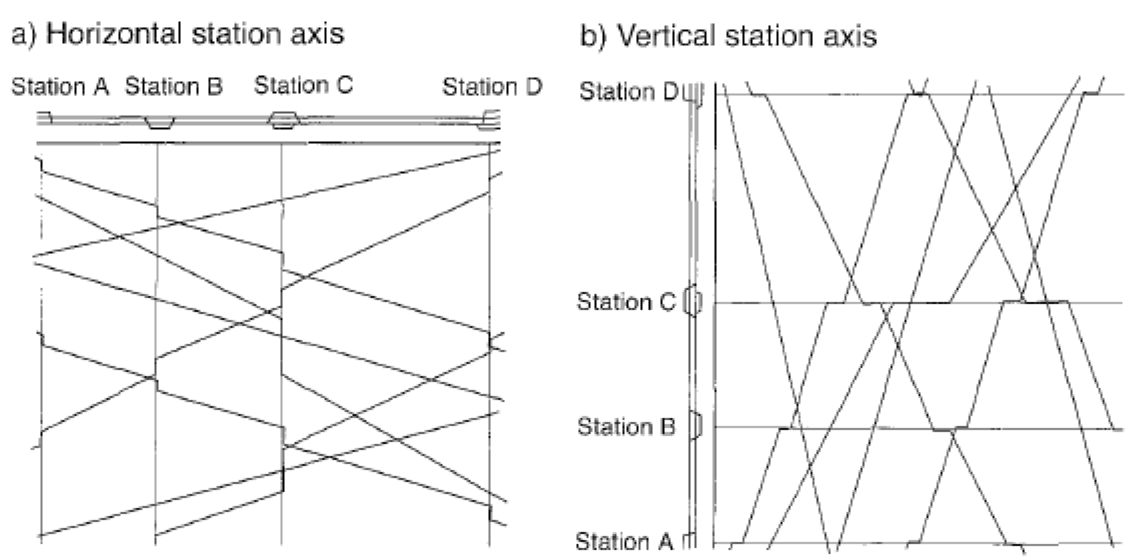


Figure 1: Example of train diagrams (Hansen/Pachl , 2008 page 15)

As shown in the diagrams, the position of the trains is drawn on a way-time-plane. In the middle of each diagram you see an overtaking which happens at station c. A slower train, which makes stops at every station, is waiting. Meanwhile a faster train is going in the same direction, passing this train without stop at station C.

The zero point on the time-scale is chosen arbitrarily. Also the choosing of horizontal and vertical axis is somewhat arbitrarily.

Additional diagrams may be used to display the routing of trains inside a station or in a marshaling yard.

A very basic tool for safe railway operation, which is in use since the beginning, is the concept of block separation. In a network with block separation, the driver of the train is only allowed to travel if he or she gets an authorization to drive. This authorization is valid only for a predefined distance, which is called the block. The signal to drive is given only if three different prerequisites are fulfilled: the track ahead must be free of obstacles like other trains, there must be a protection behind the train against other train movements and there must be a protection against trains running inside the block section from the other direction. Only if these requirements are all fulfilled can the train driver get permission to start travelling. Such permission was given in a written way in the beginning of railway operation, but then soon changed to visual permission methods with signals like flags, mechanical signals or light signals (see Theeg, Vklasenko 2009 page 180 ff.). Nowadays, such movement authorities are sometimes also given in verbal form using radio or telephone (see Theeg, Vlasenko, 2009 page 58). If there is no valid movement authority, the train driver has to stop, as this standstill is the safe state of railway operation (Theeg, Vlasenko 2009, page 29).

1.3 Internal and external factors as sources of disturbance in operation flow

When looking at railway operation from a scientific standpoint, one could have a first impression that railway operation may be totally deterministic, so that operation and performance (in the past as well as for the future) could be described purely by mathematical equations. Several books have been published containing complicated mathematical equations which attempt to describe in total the movement of trains on the track. Most of the underlying equations are differential equations. In the past, a lot of geometry and graphic displays as well as coefficient´s tables were used to make the results of such calculations available and visible for the engineers in the operational work. Examples of such efforts are the train diagrams for the description of train movements along the railway routes and the braking curves based on reference curves with characteristic coefficient dependent on train´s individual characteristics. For more information see Wende (Wende, 2003, in German) .

Additionally, in the railway system, there is only a limited number of actors. Nevertheless the railway system may be comprising thousands of different parts. At first there is a number of moving vehicles, second there is one known track network on which the trains are dependent, and third there is one management (that may contain a lot of different persons with functions like dispatchers, signalers etc.), using given procedures -fixed in operational handbooks- together with signaling infrastructure and mobile networks to manage the traffic. As there are fixed procedures, in the first sight the actual situation should be known completely and the progression in the future may be predictable or computable.

But when one considers the actual operation of railways all over the world, there is one major difference between mathematical approach and reality: the variance in operation.

There is always a difference between theoretically possible movement and real movement on the tracks, mainly in the form of delay between the theoretically possible timetable and the actual measured timetable of operation.

One main reason for this difference is the disturbance which results from events which occur outside the railway system but have an impact on the railway network. Another reason is the variance in the system´s performance itself. The main effect of such variance is decreasing performance. As the main key performance indicator of railway systems is the punctuality, the adherence to published or prefixed timetables, the dependencies resulting from the variances on this indicator is of main interest not only for the customers of the railway but also for the operating company. As in the railway system, early departure, meaning departure before the timetabled or the agreed time is not wanted and therefore forbidden in most cases, the influence of the variance is always a delay.

In his work, Goverde (2001) developed a scheme to categorize the delays with respect to their sources. At first, he differentiates between primary delays and secondary delays. He defines primary delay as „the deviation from a scheduled process time caused by disruption within the process (Goverde, 2001, page 33), whereas a secondary delay is defined as „deviation from a scheduled process time by conflicting train paths or waiting for delayed trains" (Goverde, 2001, page 34).

For the sources of the underlying variance as reasons for primary delays he defines four different categories, according to their source: disturbances due to malfunctioning of infrastructure, train operators (manmade disturbance), variances due to railway traffic management and external factors, see Table 1 (based on Goverde, 2001, page 34).

Infrastructure: Technical Malfunctioning, Maintenance & Construction	
Rail network	Tracks; Switches; Structures (tunnels, bridges)
Electrification	Supply; Catenary
Signaling	Signals; Interlocking; Train detection (track circuits, axle counters); Automatic level crossings
Train Operators	
Rolling stock	ATB-application; Malfunctioning traction, engine, brakes, running gear, doors
Personnel	Driver and conductor behavior (experience, routine, discipline, stress, illness)
Logistics	Loading/unloading; Catering
Train circulations	Shunting; Cleaning; Braking test
Passengers	Volume alighting and boarding; Supporting disabled; Aggression; Nonpaying passengers
Railway Traffic Management	
Systems	Disposition; Traffic control; Communication; Automatic Route setting
Personnel	Dispatcher behavior (experience, routine, discipline, stress, illness)
Plan	Timetable bottlenecks; Rolling stock scarcity; Crew scarcity
External	
Weather	Frost; heat; wind; sight; lightning; slipperiness (leaves on track)
Vandalism	track obstruction
Environment	Incidents at level crossings; animals on tracks; trespassers on tracks; suicides

Table 1: Sources of primary delay in railway operation (Goverde, 2001)

So, there are many different possible reasons for a primary delay in the railway operation.

For the secondary delay, only three different reasons are known (Goverde, 2001 page 34): hindrance of trains via signaling system due to a slow train ahead; a conflicting train movement at a junction or crossing and the scheduled transfer at stations waiting for a delayed connecting train.

As a result of these high numbers of reasons for delay, it is quite astonishing that high differences in the numbers between different railway companies. In Japan for example, the resulting delay in the network is measured within seconds (Hansen, Pachl; 2008, page 171) ; whereas for most railway operators in Europe even a delay of up to five minutes is still thought to be in time (Hansen, Pachl; 2008, page 171).

One approach for reduction of delay is adding buffer times in the timetable, which is then either used up by the (random) delays or by just waiting at stations. This buffer procedure is only possible however only for a limited percentage of delays, as outliers like major external incidents resulting in long delays will not be handled with this approach. As this approach is easy to implement, most railway companies have adopted this procedure in their timetables to a certain extent differing between companies. In most cases, there is a fixed percentage or a fixed buffer time included in the published timetables. For example in the Netherlands in the operation handbooks there is a rule to introduce a buffer time at layovers for 5 minutes for local trains (Goverde, 2001, page 49). This approach does not have a direct impact on the sources for delay but reduces only the effect of the delay on performance.

A second approach is to reduce the sources themselves. For example in France on high speed lines there is an electronic detection of foreign objects on the track to automatically stop the approaching train in a safe distance to reduce possible damage and additional delay. Another example is enhanced maintenance effort on infrastructure and trains to reduce the number of failures. Additionally the railway operating companies set high goals for reliability of new trains for their suppliers.

1.4 Latest developments for renovation of railway transport

Railway networks are often considered to be a strategic issue, therefore most of the railway networks in Europe have been in state ownership, at least for a long time of their existence. Therefore there was only a little effort undertaken to fully consider all costs. For example some of the costs were considered as inevitable as they were supposed as necessary for national defense. As in the last decades the state deficits were growing and growing in most countries, there was a movement towards privatization of railway networks. Additionally the need for national defense was reconsidered and the tactical reserve in logistics provided by railway networks were no longer deemed necessary. The governments thought more about the reduction of deficits as well as of reducing the number of state employees. Additionally governments were no longer in fear of a sudden war, especially in Europe, so the strategic issue of maintaining a large railway network was no longer considered a state affair. This came alongside a movement towards individual transportation, as the numbers of cars and trucks and their transportation capacity were growing much faster than that of railway transportation. Therefore railway transportation was considered in some governments and even from public view as outdated, not up-to-date and even obsolete. Together with the old fashioned outer appearance of many of the trains (due to the long usage intervals) the railway systems were no longer considered as competitive.

The private companies (which took over some services) proved, that railway systems could be competitive, when modern railway vehicles were used and speed and reliability of service was improved. In some countries, this was demonstrated not only from private companies but even from state owned railway operators which very successfully adopted these strategies of management from industry usage. Currently there is a renaissance in railway service: the number of transported persons is growing as well as the tonnage of transported goods. But since the maintenance of the infrastructure is not in line with the growing demands, there is a need for updating the infrastructure to match the transportation needs of a modern globalized society.

Therefore, during the last years there have been major research efforts in different countries to develop new methods in railway operation. Some examples are the more sophisticated signaling systems in the US, the American Railroad´s Advanced Train Control System (ATCS), and the European Train Control System (ETCS). ETCS for example was developed in Europe with the aim to reduce the overall costs in railway operation while enhancing safety and reliability. Some experts even expect increased capacity in railway networks after introduction even with the same track infrastructure (Stanley, Peter; 2011).

The ETCS is based on the idea, that through the development of a standardized control system the overall costs could be lowered. The ETCS has one main characteristic; it is an upgradable system, so at first a simple system, similar to that already in place in most European railway systems and named ETCS level 1, could be introduced. ETCS level 1 is a blocking time model similar to the conventional line side signaling system (see VIA, 2008, p. 7). There is also a second level – called level 2 - , which is also based on a blocking time model. But in contrast to the level 1, level 2 does not consider fixed starting points for braking, but considers an individual start for deceleration based on the braking characteristics of the train (VIA, 2008, p. 8). This second level has some similarities to signaling systems already used in some European countries for high speed tracks. In the concept of ETCS there is also a third level fixed. This level 3 uses a moving block model. Some authors are of the opinion, that this level 3 may have a positive impact on track throughput and capacity, whereas others are sure, that this level will never be adopted, due to the fact that some aspects for safe operation of level 3 are not fully solved yet. In one book (Theeg, Vlasenko, 2008 page 58) the authors state, that “such train separation in relative breaking distance is a rather academic idea with no realistic chance of being adopted”.

For the introduction of ETCS there are many reasons discussed in the public. At first, the railway systems in Europe are split into several separate networks, based on the different signaling systems that are in use. Altogether there are almost twenty different systems, most of them separated by country borders, but some of them used also in parallel in the same country. As the demand for longer and faster railway travels increases, the use of different signaling systems is becoming more and more impractical. In the past, locomotives and their crew have to be changed at such signaling borders between different signaling schemes. This takes much extra time and therefore slows down traffic speed unnecessarily. Therefore a unified signaling system is advantageous, since such extra processing time could be avoided and extra staff necessary for these procedures eluded. Additionally, one could consider to include the different technical installations necessary for the different signaling schemes into one locomotive and to train the staff for the different procedures associated with the different signaling systems. But it is not possible to install all systems on one locomotive due to incompatibilities and restrictions in space. Also, the training of crews is very expensive and may result in decreasing security of railway operation due to the possibility of increasing numbers of human based errors.

Another reason for introduction of ETCS is the standardization of service, which is estimated to generate economic benefits, resulting from a win-win-situation for infrastructure companies as well as for the suppliers. The companies which are manufacturing such signaling equipment will benefit from standardization, since the numbers of potential applications are increasing and a shift towards this new system will produce additional demand because of the replacement even of equipment, which was still functional and which has not already reached its end of lifespan. The companies running the infrastructure will also benefit from the standardization, since the number of potential suppliers being in strong competition is increasing. Also the dependency on the original manufacturer for maintenance of the equipment will decrease. This will result in lower maintenance cost in the long run. Furthermore the higher amount of similar installations may produce new business models with benefit for the railway companies. For example the founding of centralized, external training centers for the education of personnel needed to handle the signaling procedures would be beneficial. Additionally the establishment of centralized centers for spare parts could save fixed money and space for the individual company. Some researchers also see a benefit in the usage of modern technique due to possible lower number of failures in service than with the traditional signaling procedures, even if there has been an evolution since the beginning of railway operation.

Last but not least there is a strong discussion among experts, if the introduction of the new ETCS will increase the capacity of an existing railway network alone. For such justification the UIC has initiated two different research projects. Bothe projects were handled by the VIA, the Institute for traffic-related science at the Rheinisch-Westfälische Technische Hochschule (RWTH) in Aachen/ Germany. The results of these projects were compiled in two generic study reports, one named the Influence of ETCS on line capacity (VIA, 2008), while the other is bearing the title Influence of ETCS on the capacity of nodes (VIA, 2010).

These works are of fundamental interest, since they show, that there is estimated a higher capacity of railway nodes and lines after the shift over towards ETCS (especially when adopting the ETCS level 2), however there still remain a need for further evaluation.

Especially the aspects of statistics and randomness seem not completely introduced in the two different calculation methods for capacity (the calculation method invented from Schwanhäußer called STRELE-formula and the calculation by compression of a given timetable according to UIC Code 406) the authors of these studies have used.

The results of these studies have also influenced the simulation studies performed to justify the need for a new layout in a major railway station in Germany. After long discussions in the German public about the pros and cons of the new concept Stuttgart 21, it was finally agreed to compare the concept Stuttgart 21 as proposed by the German Railway with the existing layout of the railway station in Stuttgart. This comparison should be based on the calculation of the infrastructural capacity. As the comparison is only possible by simulation, such simulation should be made to either prove or neglect the higher capacity of the new layout. As the tool of simulation is normally mainly used by specialists, it was agreed by the different groups of stakeholders to perform an audition on the results of such simulation runs. The results of these audits were published (SMA, 2011) and demonstrated that the results (somewhat higher capacity for the new layout of the station) were based on the actual state of art in railway simulation. But nevertheless the critics on the new concept Stuttgart 21 were not completely stopped. Some stakeholders were not convinced, since the results of the simulation runs were some kind of magic for them. So it became evident, that there is actually a lack of confidence into the result of simulation studies. Such confidence can only be gained, if the process of performing a simulation study becomes more transparent for all stakeholders and if the results of a simulation study are strictly based on statistical evidence.

Therefore we propose as the main course of action for a simulation study for railway operation with different, alternative types of infrastructure:

At first it is necessary to define the desired goal of the simulation study, together with a definition of borders of the railway network under study. In this work the goal is defined as a decision between multiple alternatives (the infrastructure changes) in railway operation. Additionally it is necessary to define, how the influence of randomness will be assessed (Note: it may be appropriate not to simulate all possible randomness; since in most railway operations if fatal failures occur, there will be another plan B to operate or shut down the railway network).

Then it is essential to come up with a simulation model suitable for the goal of the study. As all different alternatives need a model, it is beneficial to have a very universal model capable of simulating all alternatives. In this work the standard simulation software ARENA (available from Rockwell Software) is used to formulate the model and run the simulation. It is important to note all the assumptions and restrictions used in the formulation and in the execution of the model, since validity of simulation may depend on them.

Then, like in all simulation studies, the models used have to be verified and validated.

To include the randomness in the models, the easiest way would be by adding constraints at the borders of the model space e.g. primary delays introduced from outside the model space. If possible, real data should be used for this randomness, but it has to be kept in mind that it may be appropriate not to include very fatal events (like very long delays that will destroy the timetable completely if they occur). If necessary, additional calibration of the models has to be performed.

After this preparation, the simulation experiment with the different simulation models can be performed, using a meta-model approach. It may be necessary to run several repetitions for each sub-model to be able to test the different results for statistical relevance. This allows to

draw conclusions by comparing the different results of simulation runs (taking into account mean, variance and distribution) with an appropriate statistic tool, like the t-test.
Also sensitivity analysis may be used to discover the interdependencies between (input) assumptions and results of simulations.
In a final step the results of the simulations in the model space back have to be transformed back with respect to the goal of the simulation study.

For some of the steps, it may be necessary to go back several steps after revealing an error (e.g. to perform several iteration steps), as it will not be likely to get a satisfying result right with the first attempt. To increase the confidence, the use of additional aids like animations and sensitivity analysis seem appropriate. Therefore it may be useful and more appropriate to use standard simulation software, which is capable of providing these tools instead of using specialized software for which such additional aids are not available at the moment.

The contribution of this work is defined accordingly in four different steps: First a literature review of the actual status of railway simulation is given. Then in a second step the proposed standardized layout for a simulation study is developed. In a third step this layout is tested with several different railway networks and signaling procedures. This also includes different types of traffic. The method to develop a simulation model is based on the concept of discrete events. As the process of a train travelling across a railway network is a continuous process several calculations and transformation steps are necessary to develop a model. The model is aimed to simulate the network´s performance with respect to travel time and train speed at certain network points, the simulation of lately developed signaling procedures and the usage of the main infrastructure of the railway network, the tracks. The results of different simulation experiments are then contrasted to those of traditional simulation approaches known from other sources. The evaluation will be based upon a railway networks performance criterion. As example of such a criterion the term capacity is introduced. In a final step the results of the comparison are compiled, together with an outlook on further possible applications and developments of the new approach.

2 LITERATURE REVIEW

2.1 Simulation

Simulation is a mathematical tool, used in operation research, to evaluate different systems or operational practices with respect to special aspects, like performance, reliability or cost. Introductions into simulation methods may be found for example in Hillier, Liebermann 2010, chapter 20; Law, 2007 first chapter and in Kelton, Sadowski, Swets 2010 in the first two chapters.

There exist other methods like linear programming, decision analysis and queuing theory to evaluate different alternatives in advance and that are able draw similar conclusions than through simulation, but compared to these methods, there are several reasons which make the simulation approach unique and superior to such other methods.

At first, one can try to develop a set of mathematical equations which mimic the complete behavior of the system of interest. This mathematical representation is needed for linear programming and is also the underlying concept of the queuing theory. But then there is often the problem to develop a valid mathematical model for the representation of this real system. More and more systems which are of interest to researchers, especially the more complicated and interconnected and interrelated with other systems, are so complicated, that all attempts to develop such a complete mathematical model are failing. This can be due to some unknown constraints (sometimes even stochastic ones) of the different process, due to unknown influencing effects from the outer world or just because the resulting list of equations and constraints is becoming so enormous, that either the single best solution of the problem is not calculable (even with the large computational power available today), either there are several solutions that are not interpretable or the result is too sensitive to small changes of the constraints (which will be revealed in a sensitivity analysis) to be realistically adopted as method to control the underlying processes.

Second one can test and change the process, either the full real process directly or a smaller representation using a mechanistic model representation. For the first subcase, such testing in the real process may be very time consuming, therefore costly and sometimes dangerous or impossible. Such testing is for example not possible, when the effects of changes to the system should be evaluated, when the system itself shall remain fully operational during the evaluation period. If the testing is used in a smaller scale system for evaluation, than the costs of such testing may be substantially lower, compared to full scale testing, but the usage of the results for the small scale system may not be fully transferrable to the full system to other physical effects and constraints.

Third, computer simulation may be the only possibility to evaluate the effects of fundamental changes to the system, in a situation when suitable models and data exist do not exist and mathematical calculations are not possible, as no physical realization is needed. Additionally, the results of the simulation are often more transparent to the target group, as animations (often a byproduct of simulations) are easier to understand than sets of equations and calculations.

But there are also some disadvantages of simulation. The process for the development of a valid and verified model may be complicated, time consuming and expensive. Furthermore the results of a simulation may be misinterpreted, if the methods of the design of experiments including correct evaluation are not fully adopted. Also for a lot of real cases, the simulation is not able to give the one and only exact prediction of a performance; therefor it is sometimes evaluated by the process owners only an additional tool in system evaluation.

So in actual application, application of simulation will be one of several steps to evaluate different alternatives. Simulation will be one important source of information, but especially when costly decisions will be necessary; it will be one of several supportive methods. But as it has been in the case of the new railway station in Stuttgart in Germany (see SMA & Partner, 2011), simulation is gaining much more importance even in the public. There the result of a simulation study together with an audit on this study was the culminating point and solution of a conflict between supporters and opponents of a fundamental restructuring of a main railway station in the southern Germany.

2.1.1 Motivation to perform simulation in railway systems

Railway systems are a special case of logistic chains. Like in other logistic chains, for example in conveyer belts or in a warehouse complex with automated storage and retrieval systems, the process of handling goods and passengers is the main purpose. But in contrast to most of the other logistic chains, the railway network is very large in size. The largest railway networks can be as large as a whole continent. One additional characteristic of the railway system is, that there is a high degree of interdependency between the movement of trains due to the limits of infrastructure, which is very high in value and expensive in money and human resources to run and maintain. Additionally, the process of handling has in most other logistic chains only a support function, as the main source of value for the product lies in separate process steps, like storage, distribution or production. However in railway systems the transportation itself is the main process of the network, a value which must delivered to the customers in a reliable and safe manner, since this is the only value which the customer is willing to pay. Furthermore the railway systems are very often a public matter. First, there are a large number of employees working directly or indirectly for the railway. Second, the railway systems are considered a public affair, since cheap transportation is a fundamental need of the public and the high impact on nature due to the land needed for operation of railway systems make it necessary to support railway construction from government. In some cases, due to the strategic impact of railway operation, the railway networks are state owned or the networks have been in state ownership during a long time of existence.

As the demand on cheap and safe transport is rising in the modern world, a side effect of globalization, the railway systems are asked for new solutions. The search for such solutions, under the increasing pressure of economic constraints scoping also the increasing environmental awareness in the public, is the main task of railway companies in the modern world today. Therefore these companies, more than in the past, are now looking for the adaptation of already proven concepts for optimization of services. Such proven concepts can be found in the research of industrial engineering, where fundamental concepts have been developed since the middle of the last century and have also been successfully introduced and maintained in the industrial use. These concepts, which sometimes were named the industrial revolution of the information age, are evaluated to match with the needs of railway infrastructure management and government.

2.1.2 Classification of simulation studies

Simulation methods, very roughly spoken, can be classified with respect to the time domain of simulation as falling into one of two categories: the discrete-event simulation, where only at special, predefined points in time changes in the state of systems are allowed. These points may be the start and the end of a production-process or the appearance of new customers in a queue waiting in a shop for the service. The other simulation method is the continuous simulation, which is necessary when a system is changing its state continuously, as it is for example in a storage tank, where liquids are filled in and the final product as a mixture is drained off continuously or when a rocket is moving in the air. The continuous simulation is by far the more complicated method, which also needs much more computational effort, so in most application of simulation it is tried to transfer the physical world, which is unfortunately in most cases characterized by continuous processes, into a series of single events, which can be easily transferred into a discrete-event simulation (Kelton, Sadowski, Swets; 2010, page 7 and Law; 2007, page 6 ff.). For the simulation of a railway network, this means that the movements of the trains, the halting processes and other courses of action have to be transferred into a succession of events (see Hillier, Lieberman; 2010, page 936ff.). This classification is, as mentioned before, only very simplifying. Additionally, a lot of different classification schemes for simulation exist, some with respect to the desired aim of the simulation, the type of software used, the type of system to be modeled and so on.

For example, another characterization of simulation can be made according to the aim of simulation studies. Such a simulation study may be performed to gain strategic knowledge, which is knowledge necessary for decisions for changing processes that are measured in years and decades, like changes in infrastructure, personnel or running equipment in the case of a railway network. A second category may be a simulation study with a medium-term horizon, like some month. This may be the calculation of a seasonal timetable or the prediction of the need of personnel. The third category of simulation studies is then the short-term simulation study, which is performed to support the infrastructure management in the case of sudden events like accidents to maintain or regain a stable timetable.

A strategic simulation will be an example of a simulation study in the following chapters, but the application of the concepts developed is not restricted to strategic simulation studies. Also simulation studies with medium term or even short term aspects may benefit from the application of standard simulation software in combination with the application of statistics. But especially when the aim of a simulation study is to find a solution in short term problems, the main aspect is not in finding the optimal and best solution without constraints in time, but in finding a feasible solution in very limited time. For such short-term studies, the application of standard software tools may be too time-consuming at the moment, but perhaps in the future the computational possibilities may make standard simulation competitive against special railway calculation methods. Actually, in such cases of urgent decision needs the expertise of infrastructure managers is much more asked than the help of computers (see Railway timetable & traffic, page 188 following). But in the same book it is expected (page 189), that in the future there will be a need for simulation as a supportive decision tool even in short-term events.

In the railway research, another differentiation of simulation methods has been used with success. This divides the simulation methods into asynchronous and synchronous simulation methods.

In asynchronous simulation methods, at first only the traffic flows with high priority are simulated, like the high speed passenger traffic over the time period. Than in additional steps the different other traffic types are added, like the local passenger traffic or the goods traffic with low priority using the remaining resources after the preceding simulation steps. By such a simulation method, the interdependencies of the different traffic flows are not fully simulated, as for example the impact of low priority traffic on the high speed passenger trains is neglected. This may be an acceptable assumption in some cases, but this must not hold in all cases.

In contrast to these asynchronous simulation methods, which are very special for railway simulation, the more general simulation method is the synchronous simulation. Here are all entities and the processes are simulated as they would be handled in the real world. In the case of railway simulation all traffic is simulated together at the same time, having in mind all the different priorities. But especially in the case of railway networks, there is a special problem with such simulation techniques: the event of deadlocks in the simulation. Such deadlock may arrive, when a resource, like a track or a siding, is used by two different trains and the routes of these trains are in conflict with each other so that the traffic cannot be handled as expected without accident. Such deadlock can be handled in two different ways: either neglecting the whole simulation running into such a deadlock or trying to prevent the simulation into a deadlock by implementing some deadlock recognition and solving algorithms. In the latter case the solving algorithms can be based on the expert knowledge (see for example Goverde, Daamen, Hansen; 2008 and Ahuja, Jha, Liu; 2007). In a lot of simulation techniques the event of a deadlock is prevented either by choosing simple and deadlock free network configurations (for example two parallel tracks each with traffic in a single direction) or by constructing deadlock free timetables and using compression techniques only.

In the past, both simulation methods, asynchronous and synchronous, have found support among researchers, however it is expected (see railway timetable & traffic, page 187) that in the future synchronous simulation methods will gain more importance even in railway research, due to the increasing computational possibilities. Asynchronous simulation methods were developed following the traditional methods of finding a timetable with graphical methods using train diagrams. So there was only very small computational effort needed to apply these methods. For the synchronous simulation the higher validity of results is a major benefit.

For a railway simulation study, it is essential to define the aims of the simulation study well in advance, as the different simulation methods may have a strong impact on the calculated results and their validity. A key for a successful simulation study is also to define a key-performance-indicator at the very beginning. Such key performance indicator may be the waiting time of all customers during the hour with the highest traffic needs or the capacity of a new railway network with a specific kind of traffic. In railway research often the quality of service or the robustness of the timetable against disturbances is mentioned as one of the key performance indicators (see also SMA and Partner AG 2011). But for such performance indicators there exists no defined and accepted definition, so in the research these indicators are not meaningful enough. Throughout this paper it is attempted to use terminology which is based on accepted papers (literature published from the institution of railway signal engineers in the Eurail press for example) and definitions given by cooperation's of different companies (like UIC guide 406) or by papers developed in cooperation (like Eurostat 2009).

2.1.3 Fundamentals for a simulation study with stochastic influences and use of statistics for evaluating the simulation results

David T. Sturrock (Sturrock, 2010) gives some hints for simulation studies. Based on the author´s experience with simulation he gives eight steps. At first it is very important for him, that the performers of a simulation study define the objectives of their study, with respect to the stakeholders and their expectation towards a successful simulation work. In a second step it is important to get knowledge about the system which is to be modeled. In a third step a functional specification should be formulated to clearly define the deliverables of the work, in close cooperation with the study´s stakeholders. Then the reader is reminded that each simulation study is not only a single event and the result of one calculation, but is the result of a project work connected with decisions, changes, successes and failures. The important point for the author is that in this phase the performers of the study reflect their own work, looking critically at the failures not to lose the important tasks in the simulation project out of sight. Another important topic for Sturrock is to remind the users of simulation methods that the collection of correct data is essential for a successful simulation. Only after all these steps are performed, the author recommends to build the simulation model, always having in mind that each model is only an approximation and that the model has to be verified. The authors also stresses that the model´s results have to be validated against the performance of the real system. It is so important for him, that this validation gets its own process step. Instead of testing the model against reality, which is sometimes not possible, he gives two alternatives for validation: either to make the model as deterministic as feasible and evaluate the results or to use the experience of the stakeholders for validation. Only in the last step, the author recommends to experiment with the model, to analyze the results of simulation and to present the results of the simulation study. For the presentation of results, the author suggests to concentrate on three things: first the different alternatives considered in the study, second the conclusions or recommendations and third to give supporting information to the stakeholders to gain their confidence in the conclusions of the simulation study. He reminds the reader, that each simulation result is only an approximation of the reality, so the conclusions and recommendations should always have this fact in mind.

As the verification and the validation of the model are –according to Sturrock – very important processes of a simulation study, the possible techniques for these are discussed very often in literature about simulation. In his paper for the 2010 Winter Simulation Conference, Robert G. Sargent concentrates on verification and validation techniques. He states that for a confidence of the stakeholders in the model (and the results of the simulation study) it is essential, that the verification and validation are well documented to convince the stakeholders.

The possibilities for verification and validation can be divided according to Sargent into groups of usage: either subjectively or objectively. Objectively means in this concept using mathematical procedures and tests, for example hypothesis tests. In general a combination of different techniques is used for verification and validation.

The process of verification and validation is illustrated in Figure 2 (source: Sargent, 2010, page 170).

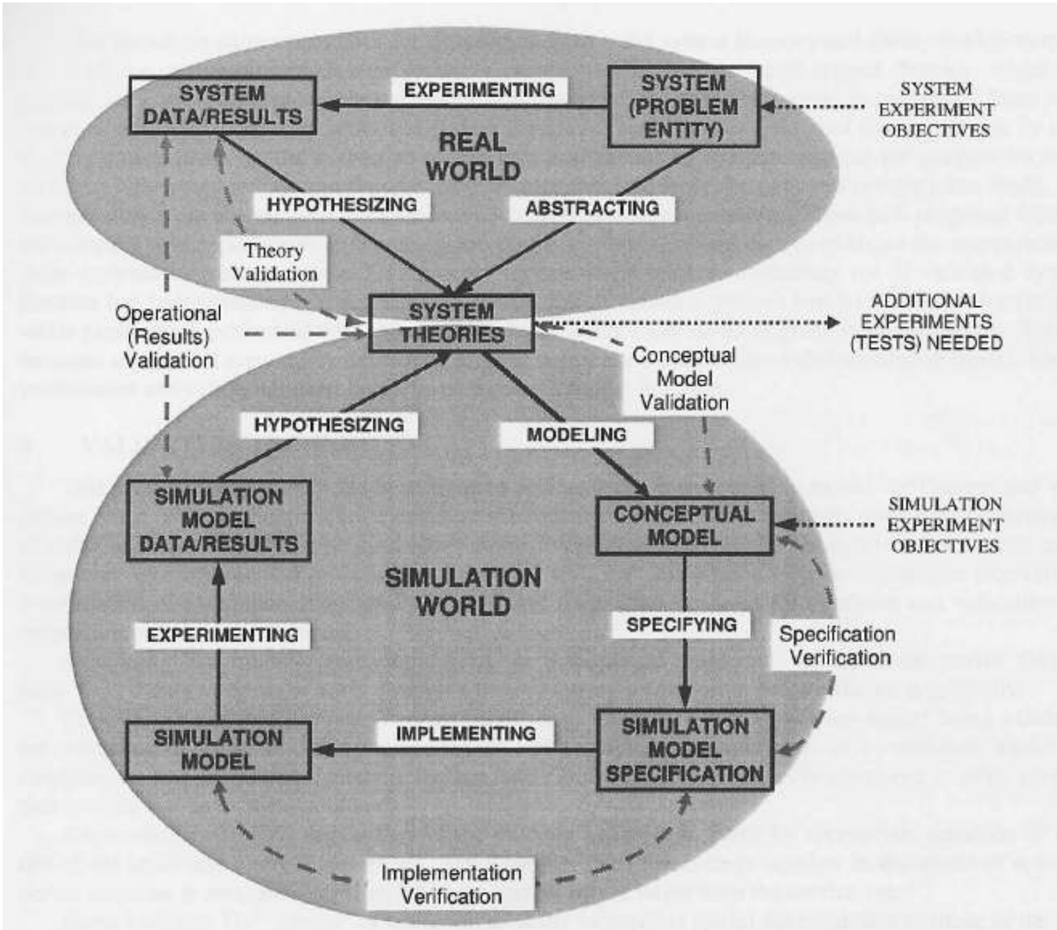


Figure 2: Process of model verification and validation (Sargent, 2010)

The verification is the process to assure that the software design and the specification for the simulation model itself are satisfactory. This can be called the specification verification. A further procedure is the implementation verification. Validation deals with the process of determining that the theories and assumptions which led to the model are consistent with the system theory and that the model is suitable for the intended purpose. This is called the conceptual model validation by Sargent, whereas the process of testing, that the model´s output has sufficient accuracy for the desired purpose of the simulation is called operational validation. Therefore validation is always in some kind dealing with the relationships in the real world, whereas the verification is dealing with the simulated world alone.

Testing the model can be performed in different ways. Sargent gives some examples in his paper. Animation and operational graphics can be useful as they may reveal operational errors in the model. The comparison of the model with other model can be also very useful. Sometimes degenerate tests and extreme condition tests may also very helpful. Validity checks for events, tracing of special entities and checks by experts with further knowledge about the simulated processes may also very valuable in the verification and validation. To compare the results with historical data is another possible technique, added sometimes by predictive validation that is using the model´s behavior to predict changes in the system. By using the sensitivity analysis the performer of a simulation may test the parameters of his or her model, eventually uncover inconsistencies in the underlying data collection. But also statistical tests like the calculation of confidence intervals are useful technique. Additional hypothesis tests are useful according to Sargent to demonstrate the ability of the model to predict the behavior in an acceptable range of accuracy. For convincing the stakeholders of the simulation study, a good documentation of the verification and validation process is helpful. But as all simulation studies are different, no general procedure can be given as the correct one. According to Sargent verification and validation is more a process than just performing one additional test. As this process is critical for the acceptance of a simulation study´s results, even third party certification of a simulation model is performed.

2.2 Railway operation, timetable and infrastructure

2.2.1 Literature about railway operation

In the book Railway Signaling and Interlocking (2009), different authors write about certain aspects of railway systems, especially those aspects which deal with safety of operation.

As a matter of fact, the railway system has some advantages (like relatively low investment needed to start operation, low energy consumption in service) compared to other transportation systems. But there are also some disadvantages associated with railway operation, like the fixation on the existence of tracks and the poor braking performance, as already at speeds higher than 30 km/h the braking distance is longer than the sight distance of the train driver.

To improve the performance of the railway, especially the speed of transportation, different systems have been developed in several countries around the world to overcome these disadvantages. The book does not only give the actual status of these systems, but also gives a historic review of interlocking systems, especially with regard to the reasons, why different approaches were developed.

At first, there is the British approach based on procedural safety (which was later adopted in many other parts of the world), as Britain was the first nation to introduce regular railway services. Later, there was developed a second approach, which has its origins in Germany, where a more mechanistic approach was invented, relying on automated procedures reducing the possibility of human errors. In addition to these concepts, there was a third approach developed in the United States of America, which was later influencing all other railway operations. In North American Railways the base for operation is laid in a centralized operation combined with new information delivery to the train driver like the usage of on board signalization.

As since some decades the use of computers and modern communication tools like mobile phones and the World Wide Web were introduced in industrial and private usage, there was also a strong desire to introduce these new technologies also in railway operations. Also the loss of attraction towards railway operation for passengers and industry (freight services) in the last decades made it necessary to enhance the performance and at the same time to reduce costs of existing railway networks by introducing such new technologies.

As the authors explain in the book, the introduction of new technologies , which are already tested and proofed in industrial use, could be a solution for the dilemma of decreasing demand and rising prices for the needed resources and the operation and maintenance as well.

The elements needed for such more automated systems are explained in different chapters, at first the basic railway operation processes, then the interlocking principles. In the next chapters the detection possibilities for trains in a network are explained, followed by chapters about movable track elements and signals. On the next chapters, the different train protection systems are explained, with the ETCS (European train control system) being the most prominent and actual one. The status of implementation of this control system into commercial service all over the world is also described. In the last chapters, some specific aspects of railway operation are described more in detail, like the control of marshaling yards and the problems to operate level crossings and hazards in railway systems.

In the illustrated glossary for transportation statistics (Eurostat 2009), harmonization of transportation statistics in the European Community should be achieved by introducing the same terminology in all sectors of transport, not only railway transport. The glossary was developed by an intersectional working group between UNECE (the United Nations Economic commission for Europe in Geneva) the ITF (the International transport forum in Paris) and the Eurostat (an office of the European Community for statistic affairs). Additional to the definitions for railway transportation, also terms for maritime and other modes of transportation (like road, via inland waterway, pipeline and air) are given. However, the first chapter of this leaflet deals with railway transport, starting with definitions for infrastructure.

Track: A pair of rails over which rail borne vehicles can run

Track gauge: Distance between a pair of rails measured between the inside edges of the rail leads (problem: large gauge, narrow gauges on some tracks, therefore no easy interchange, standard gauge is 1435 mm)

Running track: Track providing end-to-end line continuity designed for trains between stations or places indicated in tariffs as independent points of departure or arrival for the conveyance of passengers and goods

Sidings: Tracks branching off running tracks

Line: one or more adjacent running track forming a route between two points. Where a section of network comprises two or more lines running alongside one another, there are as many lines as routes to which tracks are allocated exclusively.

Conventional railway lines: All railway lines that are not classified as dedicated high speed lines or upgraded high speed railway lines.

Upgraded high speed railway line: Line where the allowed speed is close to 200 km/h for the main segments

Railway network: All railways in a given area

Maximum operation speed: The highest speed allowed on commercial service taking into account technical characteristics of the infrastructure

Railway station: A railway establishment which is either open or not to the public, generally staffed and which is designed for one or more of the following operations: - formation, dispatch, reception and temporarily stabling of rains; - stabling and marshaling of rolling stock; - boarding and alighting of passengers; - generally, where open to public, providing facilities for the purchase of tickets; - loading and unloading of goods

Halt: Stop-of point generally open to passenger traffic only and not usually staffed

Marshaling yard: station or part of a station especially equipped with a number of tracks or other equipment for railway vehicle marshaling (switching) operation

Railway vehicle: mobile equipment running exclusively on rails, moving either under its own power (traction vehicles) or hauled by another vehicle (coaches, railway trailers, vans and wagons)

Wagon or freight wagon: railway vehicle normally intended for the transportation of goods

Van: non traction railway vehicle forming part of passenger or goods train and used by the train crew as well as for the conveyance of luggage, parcels, bicycles, accompanied road passenger vehicles etc.

Coach: passenger railway vehicle other than a railcar or a railcar trailer

Railway enterprise: any private or public enterprise acting uniquely as a railway transport operator, an infrastructure manager or as an integrated company (combining the first two alternatives)

Railway transport operator: any public or private transport operator which provides services for transport of goods and/ or passengers by rail

Infrastructure manager: any enterprise or transport operator responsible in particular for establishing and maintaining railway infrastructure, as well as for operating the control and safety systems

Integrated company: railway transport operator also being an infrastructure manager

Railway traffic: any movement of a railway vehicle on lines operated

Shunting: operation of moving a rail vehicle or set of rail vehicles inside a railway station or other railway installations like a depot, a workshop or a marshaling yard

Railway vehicle journey: any movement of a railway vehicle from a specified point of origin to a specified point of destination

Train: one or more railway vehicles hauled by one or more locomotives or railcars or one railcar travelling alone, running under a given number or specific designation from an initial fixed point to terminated fixed point

Types of trains: goods trains, passenger trains, mixed trains and other trains (no payments to other third parties)

Railway transport: any movement of goods and/or passengers using a railway vehicle on a given railway network

Accident: unwanted or unintended sudden event or a specific chain of such events which have harmful consequences. Railway accidents are accidents in which at least one moving rail vehicle is involved.

Further definitions (from Railway timetable and traffic, 2008 page 14 ff.):

Train diagram/ traffic diagram: on most railways, train diagrams are used both as the basis for all planning of railway traffic and also as essential documents for the control of the current operation. The only exceptions are the North American railways where traffic diagrams (there called string line graphs) are used for capacity analysis and in very early stages of operation planning, while in current operation tabular sheets are preferred. The amount of traffic on a line is described in form of a time-, distance diagram that consists of a time axis and a station axis. Train movements are represented by train paths (time-distance graphs) with a train description inscribed on them. Traffic diagrams may be shown with a horizontal or with a vertical station axis (see Figure 1).

Blocking time: sum of the following times: time for clearing the signal, signal watching time, approach time (only for train without stop), time between block signals, clearing time and the release time. For an illustration, see Figure 3 (Source: Hansen, Pachl 2008; page 20).

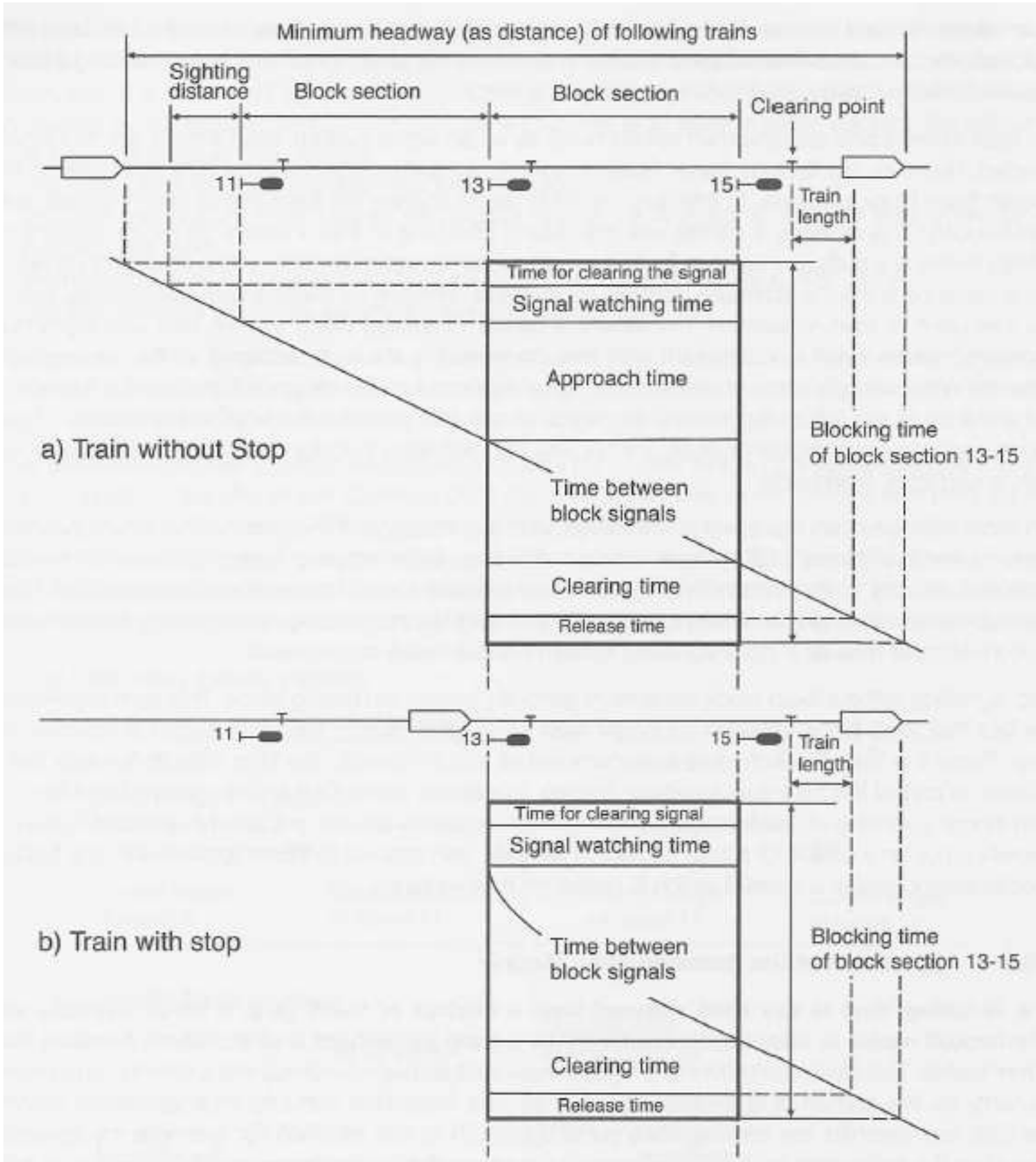


Figure 3: Blocking time of a block section (Hansen/Pachl, 2008)

The process of the determination of the minimum headway by blocking time stairways (from Railway timetable and traffic, 2008 page 21) is illustrated in Figure 4 (source: Hansen, Pachl, 2008; Page 21). The second train is moved as long on the time axis in the traffic diagram, until the blocking stairways of the second train hit those of the first train.

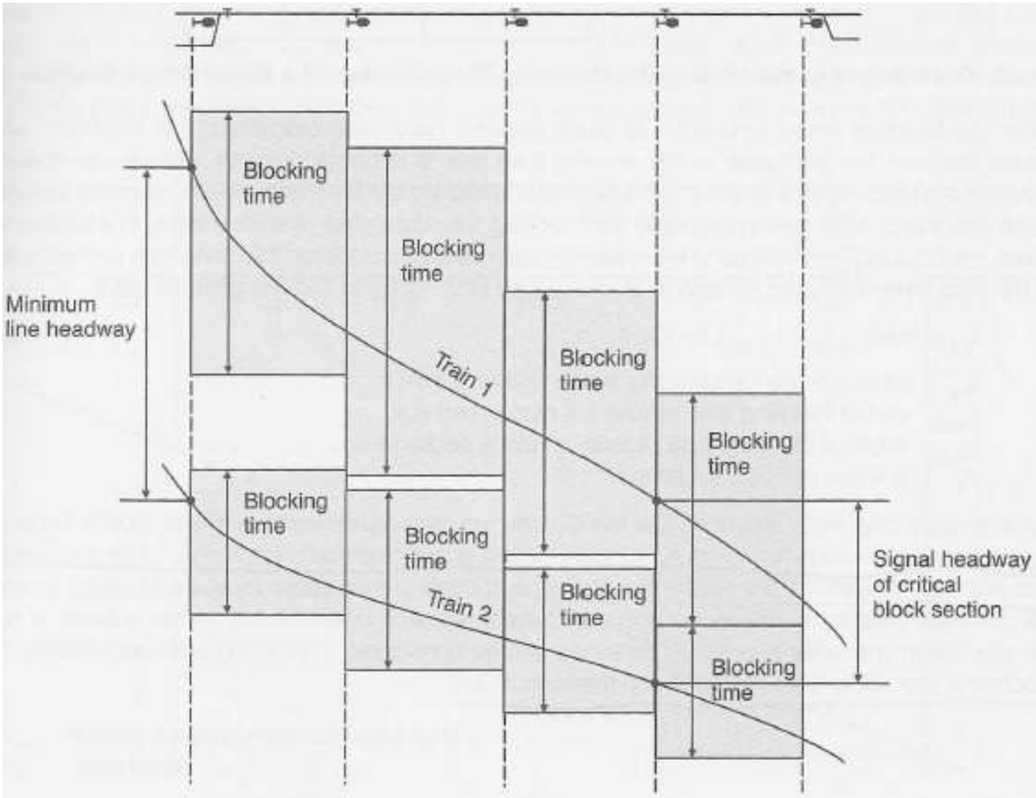


Figure 4: Minimum headway determination (Hansen/Pachl, 2008)

2.2.2 Attempts to describe the railroad operation in an abstract way

Wende, 2003; Fahrdynamik des Schienenverkehrs

The author, a previous professor at the University of Dresden, attempts to give comprehensive calculation rules for driving dynamics of railway vehicles: railcars, wagons and locomotives.

For this, he at first describes in this book the physical basics for train movement on railway tracks. Then, based on the equity of forces he develops a calculation scheme for the dynamics of a driving train: acceleration, sailing (travel without influence through motor acceleration or braking) and braking. Because the resulting set of mathematical equations is not easily solved - even if a solution is possible the given constants used in the equations are neither always given respective known nor really constant over the whole range - the author derives a set of simplified equations for evaluation. Additionally he gives several examples with values for comparison, mainly from German railway networks, mostly via diagrams or in tabular design. According to the author, the use of diagrams played and plays a significant role for the scientific and practical work on train dynamics, as some interrelations in this field are nonlinear, like the e-function and parabolic and quadratic functions. As a main problem for exact calculation of train movement (the mathematical description of the dependency of time and space of a train in the future) the author identifies the contact between wheel and track. For calculation of the train dynamics he also derives the differential equations that are suitable for computation the train movement with the use of calculators and computers.

A more specific approach, mainly concentrating on the aspects of calculating running times of trains between stations using computers is described in Hansen, Pachl 2008; page 58 to 82. Here in addition to the calculation of running times using a difference equation approach, also several different methods to calculate the running time in closed formulae derived from integration are presented. But the authors also discuss the fact that some basic assumptions for calculating this running time do not hold, especially regarding the train data. With the examples of the mass of the train, the mass factor and the weight of a tractive unit in the case of a diesel hauled locomotive the authors demonstrate, that even these basic data is in fact not constant but changing from day to day and may sometimes – as in the case of the diesel in a diesel hauled locomotive – even change during the trip. Therefore they also mention probabilistic approaches for calculating the running time, which take these variations into account using randomly generated data to calculate a most likely running time.

UIC Code 406 Capacity

Aim of this leaflet from UIC (Union de Chemin de Fer) is to enable infrastructure managers and others to carry out capacity calculations with standardized definitions, criteria and methodologies. Main requirement for evaluation with respect to capacity is the examination of already existing pre-constructed timetables for train-operation. A specific step for calculation is the compression procedure. In this process, all unnecessary times in the timetable are eliminated, but the times needed for train operation and other necessary occupation times. Then in a second step, any additional buffer times are added for timetable stabilization and for maintenance requirements, where needed. The ratio of the time if the compressed timetable divided by the full time of the timetable is then called the actually capacity of the network. This figure may additionally be compared with typical values, which have been derived from several real railway networks. If the actual value is lower than the typical value, there is a possibility for some left-over capacity, which can be assigned for further traffic and train paths. Otherwise, there is already an overload in the network, so there is a potential risk for delay propagation in the network. However even the authors mention in their paper, that this result is just an approximation for the capacity, which must be further developed through complementary analysis (as for example proposed in this dissertation).

Capacity: on a given infrastructure, capacity depends on the interdependencies between: - the number of trains per time interval, the average speed, the stability, the heterogeneity (my interpretation: the last two factors can be renamed as variance). Different views on capacity exist: - from market/ customer needs, - infrastructure planning, - timetable planning, -operations.

Constraints on capacity: - priority, - timetable structure, - capacity allocation process, - design rules, environmental protection, safety aspects, technical constraints, theoretical capacity.

Theoretical capacity: a theoretical maximum capacity expressed in terms of the maximum number of trains can be calculated by defining ideal circumstances.

Capacity: the capacity of any railway infrastructure is the total number of possible paths in a defined time window, considering the actual path mix or known developments respectively and the infrastructure managers own assumptions in nodes, individual lines or parts of the network with market-oriented quality. This must also take into account the infrastructure manager´s own requirements.

Corridor: All possible journey routes (main rout or alternative routes) according to market needs, between a defined source and target.

Route: consecutive lines and nodes as a whole, between a defined source and target.

Line: a link between two large nodes and usually the sum of more than one line section.

Nodes: points of a network in which at least two lines converge. Nodes can be stations or junctions. They can be differently sized, depending on the number of converging lines and their tasks.

Station: point of a network where overtaking, crossing or direction reversals are possible, including marshaling yards.

Junction: point of a network in which at least two lines converge and neither overtaking, crossing or direction reversals are possible.

Line section: the part of a line, in which the traffic mix and/ or the number of trains and the infrastructure and signaling conditions do not change fundamentally. It consists of one or more coherent sections, which are limited by two neighboring stations or nodes.

Part of a network: specific combination of nodes and lines to fulfill specific tasks and controls.

Relevant block sector: block section within the chosen line section, which determines the maximum headway along the entire chosen line section.

Calculation of capacity consumption:

Notes: This procedure is applicable on a single line section. In order to access capacity and bottlenecks for a line or a whole route, the capacity of every single line section shall be calculated. The highest value of capacity consumption on a line section shall determine the capacity consumption of the whole line or route. The capacity consumption varies according to time of day, weekday and season. Recommended time windows for evaluation shall be either peak hours (at least one hour) or a 24 hour daily total (determination of a representative weekday), although it should be noted that a full 24 hours is never possible due to reasons such as maintenance work. It is recommended to use a basis being a representative day (Thursday for example) over a peak day period at least two hours long.

1) Timetable

   A pre-calculated timetable (a real operational one or a case-study) for train operation on the particular infrastructure exists. This timetable is then analyzed afterwards from the view side of capacity consumption.

2) Compression

   All single train paths of the timetable are pushed together up to the maximum theoretical headway according to their timetable order, without recommending any buffer time, either by constructing graphical analysis or suitable tools for this case or by analytical calculation. During the compression process, neither the timetable running times, the given over takings, crossings nor stopping times (which are requested by railway undertakings) may be changed.

3) Calculation of times share

The basis for the investigation into capacity consumption shall be the time share of a timetable and of a compressed timetable using equation 1 and equation 2.

$$k = A + B + C + D \tag{1}$$

where

k: total consumption time [min]

A: infrastructure occupation [min]

B: buffer time [min] including heterogeneity time

C: supplement for single-track lines [min]

D: supplements for maintenance [min]

and

$$K = k * 100/U \tag{2}$$

where

K: capacity consumption [%]

U: chosen time window [min]

Journey time: a typical journey time shall consist of the sum of a basic time without time supplement, which is calculated according to line and rolling stock characteristics and a time supplement, which is assigned depending on the type of route (either route for passenger or freight trains) and which is used to compensate for smaller-scale irregularities in the course of train travel (smaller-scale delays, exceeding of stopping times etc.). These time supplements are also used to cover the usual speed reduction due to maintenance work. A typical value of this time supplement is about 5% of the journey time.

Timetable journey time: It is the sum of the typical journey time, additional times, which result from market requirements and additional times which result from timetable construction constraints.

For compression-based methodology, not only the timetable train paths on the whole line section are incorporated in the elementary occupation time, but also the occupation times, which are not related to train paths, must be incorporated. Those latter times are also called indirect occupation times.

Buffer time: Buffer times are those times, which are inserted between train paths in addition to the minimum interval between trains that arises depending on the signal systems used. They serve also to reduce the transfer of delays from one train to the next one (secondary delays).

Heterogeneity time: heterogeneity time is related to the different speeds between the individual trains on the line. It is the minimum time between two train paths as a result of different speeds of these train paths and the block system.

Congested infrastructure: an infrastructure is called congested (according to UIC leaflet 406), if either the infrastructure´s occupation is larger than the corresponding typical value or shifting of routes in practical service is so extensive that the market requirements are no longer be met. For typical values for capacity consumption for congested infrastructure, see Table 2.

Type of line	Peak hour	Daily period	Comment
Dedicated suburban passenger traffic	85%	70%	The possibility to cancel some services allows for high levels of capacity consumption
Dedicated high-speed lines	75%	60%	
Mixed-traffic lines	75%	60%	The capacity consumption can be higher, when the number of trains is quite low (smaller than 5 per hour) with strong heterogeneity.

Table 2: Standard values of capacity consumption for congested infrastructure (UIC leaflet 406)

When the operation of a railway network is the main reason for the simulation study, the aim of the study can be manifold. From just supporting function in verifying an existing main infrastructure like tracks and number of sidings for the traffic planned in the upcoming years to the concrete calculation of a preliminary timetable to handle a railway network situation for example in the case of a sporting event. These different tasks require also a different degree of abstraction in the simulation model. For the strategic issue of infrastructure planning, a more general model, often called a macroscopic model (see Hansen, Pachl 2008, page 156) may be sufficient. For such a macroscopic model it may be sufficient to model the tracks with their preferred direction of travel and the nodes of the network, together with the trains as infinitesimal short and travelling along the network with an average speed. For a more detailed study, down to the correct prediction of operation up to the minute like in the second example a much more detailed simulation model of the railway network is needed. In contrast to the macroscopic model, this may be called a microscopic model. Such a microscopic model may include the length of block sections, handling procedures, different types of signaling systems and the acceleration and braking characteristics of the different trains together with their length. In contrast to the more generalist view of macroscopic model, in a microscopic model the different relationships between trains become very important. For example, it is not allowed in a microscopic modeling for a train to enter a block section before the proceeding train did not leave this section and the signals have been switched. Therefore the model shows the same behavior as the real railway system. So every event, that may have influence on the operation of other trains must be simulated (and included in the list of events) like the start of a new route, the passage of block signals, the waiting times in stations or the waiting times in front of signals showing enforcing a stop signal. As mentioned before it is not sufficient for such a microscopic model to calculate running times between events from mean speeds but to calculate them depending on train capabilities, mainly motor and brake characteristics, the actual speed restriction in this track section, the actual speed at the start of the section and the signaled status of the railway network.

Whereas for a macroscopic model it may be sufficient to use rules of thumb to include disturbing factors for a microscopic model this approach may be too vague.

2.2.3 Influence of infrastructure on railroad operation

ETCS Implementation Handbook (UIC, 2008)

This handbook is aimed to support the implementation of European train control system (ETCS). It should support the infrastructure managers in their decision towards phase wise adaptation of the ETCS in their railway networks. At first, the different infrastructural components of an ETCS-implementation are explained.

The existing on board installations in the trains have to be modified, if not replaced at all. Additionally the signaling infrastructure of the track has to be changed. Also many new procedures in railway operation have to be modified to fit into ETCS. The original goal of the inventors of ETCS was, according to the authors, to ensure cost reduction by using the economies of scale also in the railway infrastructure. The idea behind this approach was that more installation of the same standardized signaling system should be cheaper than the installation and maintenance of a lot of different signaling systems. Also the introduction of a seamless trans-border service in the European railway networks seems to become easier, when there is only one signaling system in all countries. For this introduction two constraints must hold: ensure the level of safety in all phases of adaptation comparable or higher than the actual level and ensure at least the actual status of reliability and quality of service (including the capacity with respect to train operation). As a book from practitioners for practitioners, it sums up a lot of recommendations and findings of early adaptors of ETCS, added with some best practice examples. The functionality of the main different levels of ETCS is explained. For the train three different levels can be stated, whereas for the tracks and lines five different levels can be distinguished. The different possible combinations of these levels are explained together with possible optional enhancements. Also the different national parameters are explained, which were introduced into the ETCS-standards to reflect the actually very different approaches for signaling and operating procedures in the different European railway networks. Also the authors add a warning, as for some operations the capacity with ETCS may decrease compared to the conventional control and signaling system. The authors inform the reader, that this is mainly due to the more restrictive handling with braking curves in the automated ETCS. As the most unfavorable breaking condition is used always in ETCS (for safety reasons,) in some cases a skilled driver in the conventional system can be better off than the driver with the automated ETCS. This paper ends with recommendations for optional procedures and staff training. Also some examples of ETCS-installations are presented as well as an overview of the relevant document structure needed to define and describe the ETCS.

2.3 Simulation of railway systems

2.3.1 Description of different railway simulation methods

Synchronization and control of a multimodal transportation system (di Febraro, Genova, 1994)

In this paper, the authors describe the increasing a discrete event-simulation to predict the behavior of a transportation system with different means of transportation, like light railway, bus and underground urban railway systems. Their motivation is to enhance performance of the existing systems by using more optimal operation procedures like the present ones. Based on statistical data, the authors predict, that there is an increasing demand for synchronization between the different transportation systems in urban areas. For the evaluation with respect to the performance of the different operational alternatives on such multimodal networks the authors propose the use of discrete-event simulation. Also they propose to include the delays in such simulation. In their work, the authors have developed such a simulation tool and are possible to use the results to develop different control strategies for a given system with respect to delay reduction. However due to computational restrictions the results remain at a very fundamental level, as only main characteristics could be modeled with their approach.

Abril, Barbe et al. An Assessment of railway capacity, 2008

Based on an analysis of the transportation sector and its increase due to globalization due to the globalization especially in the first years of the second millennium, the authors see a need for the effective use of the existing railway infrastructure in Europe. The authors propose to perform capacity studies of the existing networks as a first step towards such effective and optimal use. After a short review of the different existing approaches for such studies, the authors present also some of the results of such studies using different of these mentioned methods. The main task of these studies was according to the authors to demonstrate the interdependencies of infrastructure and other factors on railway network capacity. As they present these studies, they identify a gap in the works done so far, mainly the lack of tools to develop optimal railway timetables for the optimal usage of the available capacities of a given railway network. Therefore they present a computer-based tool called the MOM system as a solution for this dilemma. With their approach, the authors try to combine different calculation methods for railway capacity in a new way. According to the authors, three different categories of methods for calculating the capacity exist: Analytical, Optimization based and Simulation based methods. According to the authors, the analytical methods aim at calculating mathematical formulae, which stand for the real performance of a system. With some of these methods, based on queuing theories, it is possible to incorporate some aspects of randomness and even the mix of trains and different signaling locations. However the authors think that such analytical methods are suitable for identifying bottlenecks in the railway network and as a good starting point for the other methods, only. For the second category, the optimization methods, the authors state that these methods use procedures like linear programming to develop optimal saturated timetables out of a set of equations, which describe the behavior of the railway network together with a set of constraints. The authors are of the opinion, that these methods are able to provide better solutions for the timetable problem, than with the use of analytical methods only.

The third category the authors describe is the simulation method. In simulation, the operation of a real-world system, in this case a railway system is imitated to evaluate the network with respect to its dynamic behavior. Simulation methods are often combined with other methods. The authors are of the opinion that this method the hybrid approach, as they name it, is the best choice. So they developed a system, that combines analytical methods with optimization based on the UIC leaflets on capacity (remark by the author: at the time this paper was published, an older version of the capacity leaflet of UIC was in use, so the term capacity and capacity consumption is used in a slightly different meaning in this actual paper based on UIC leaflet 406). After introducing this main idea, they illustrate the procedure with an example of a railway network. Within this example, the get a compact timetable, for which the authors themselves claim a disadvantage, since it eliminates „buffer time between trains. These schedule stacks could adsorb normal operational variances, so their elimination may cause serious reliability problems in the management of the infrastructure“. As they are aware of this disadvantage, they add a chapter to their work, where they discuss the several influencing factors on (theoretical) capacity, also adding some information about the benefit of adding stack time to increase reliability for the sake of losing some capacity.

In their final chapter, the authors discuss briefly the influence of ETCS and ERTMS on capacity. They demonstrate that for ERTMS there is an effect of train speed (travel time), braking time, block distance on the number of trains possible in a given time, which can be expressed as capacity. However they are not able to calculate the capacity, they give instead only qualitative statements. They conclude their work with some summaries of their findings:

they are of the opinion, that they demonstrated the trade-off between optimal capacity usage and the reliability and robustness of service. Additionally they admit that there is still a need for better prediction and calculation of capacity usage to evaluate the different influences on this trade off to find an optimal one.

Hansen et al. Railway Network Timetabling and Dynamic Traffic Management (2009/2010)

In his paper, Hansen attempts to give an overview on the different approaches to apply simulation methods for the benefit of finding an optimal railway timetable and for optimization of the traffic management. He differentiates between analytic (queuing) methods and stochastic micro simulation. He states that these methods can be useful to analyze the waiting times and the capacity consumptions in an actual timetable for a given railway network. Furthermore the author concludes based on his tests with these calculation methods, that methods which combine both approaches (he names them as combinatorial methods) in addition to stability analysis should be applied to get an optimal network timetable.

The author asks for such new approach, as a real-time simulation tool to support the train drivers and the infrastructure management as well in their reaction on disturbing factors. He claims that such early reaction on these disturbances is the most appropriate countermeasure to reduce secondary delays in railway operation. He estimates that the continuous update of information to the train driver will be an effective tool to reduce delay propagation. Based on his theoretical evaluations he comes to the result, that the actual methods used are not able not incorporate sufficiently the influence of the human factor (i.e. the driver behavior and the expertise of the infrastructure manager), so he calls for additional research in the field of railway simulation.

2.3.2 Simulation of railway systems in deterministic approach

Sergey Vakhtel, 2002 (Rechnerunterstützte analytische Ermittlung der Kapazität von Eisenbahnnetzen)

As we have seen in several literature before, also this author starts his work with the statement, that there is a lack of adequate methods to calculate the capacity of a railway network. He tries to justify his assumption by a literature review. In his paper, the author contrasts these old methods with a new approach, a system called ANKE (abbreviation for Analytic Network Capacity Determination), to investigate the relation of infrastructure and capacity for the German Railways. The system ANKE, which was developed by a group of researchers, is using according to the author an analytic approach to calculate the network´s capacity. He admits that in the future (when more computational resources may be available) simulation methods may be superior than the analytical methods, but for that time (the paper was written in 2002) the author is of the opinion that analytical methods are performing better, as handling is much easier, also less computational effort is needed and such analytical methods are able to provide the user with the exact position of a bottle neck in the network whereas simulation methods (with random influences) are only able to give a good estimate on the bottleneck´s position.

However, the author states that due to the manifold simplifications and assumptions that are necessary in the modeling process some aspects of infrastructure management could not be handled adequately, like solving actual routing problems or automatic timetable construction. In his work, the author describes the weaknesses of the methods based on Schwanheußer (a German researcher active in the field of railway network operation planning well known for his compression methods to determine capacity consumption), mainly based on the fact that these methods all strongly depend on the correct choose of the calculation method, which is not easy to choose when the solution is not known a priori. This choice has to be done with respect to the aim of the capacity calculation, which sometimes can change due to different stages in strategic and operational procedures. As in some methods the aspect of variation is not correctly taken into account (namely the dependencies of routing due to a given timetable) the results may be questionable.

Goverde and Odynk (2002) Peter

The authors describe in their paper a tool for evaluation and comparison of network timetables used to assess network performance indicators in a deterministic setting. The method which is called PETER is based on the application of a special algebra, the max-plus algebra. The performance indicators possible include cycle times, throughput of the network and the stability margins, which are related to the critical parts of the railway network. Timetable robustness can according to the authors also be analyzed by calculation of the cumulative recovery times and the delay propagation in an existing timetable under special stress situations like the (fixed and predetermined) delay of some trains in the network.

The authors state, that train operations are typically exposed to random variations in the train running times and the dwell times which results in primary delays. The calculation method PETER is based on a timed event graph which itself is represented by (max, plus)-recursions. As stated before, the max, plus algebra is a special algebra, where a system theory similar to the conventional system theory has been developed, the main difference are the different calculation procedures and the interpretation of the outcomes (translation into normal algebra and analyzing afterwards). Some main disadvantage of this approach is that such calculations depend on the existence of a periodic timetable in which all train departures repeat themselves at a regular (hourly) time interval. The authors claim however, that their calculation method is not restricted to such periodicity, but they give no evidence for this assumption in their paper. The authors define the throughput of a railway network as a key network performance indicator denoting a trade-off between maximum performance under ideal circumstances (In UIC 406, the throughput is defined as theoretical capacity) and the robustness against external disturbances. For this trade-off they state that is necessary to introduce buffer time into the timetable. Additionally the authors demonstrate the possibility of PETER to calculate delay propagation on an example of a part of the Dutch railway network. In this example, the transfer of primary delays of some trains on other trains as secondary delays can be observed. They also state, that by integration of buffer time, a timetable should be able to cope with a certain amount of primary delay without additional action by the infrastructure managers. For this ability they define another performance indicator, named the recovery time. This is defined as the time which is needed in a given network and a given timetable to recover from a given delay to normal (for example according to timetable) operation procedures. In their paper, the authors show the different results of calculations of such delay propagation in a network which is based on a part of the actual Dutch railway network. They analyze a given timetable with respect to delay sensitivity for a certain train relation using PETER. With their work, they are able to describe the train interdependencies between certain train paths and can present a calculation for the recovery time in the network, when a certain (given) amount of trains have a fixed primary delay at certain stations. However they are not able to demonstrate the ability to calculate with PETER in a more realistic environment with random delays.

VIA/UIC (2008) Influence of ETCS on line capacity of lines see following chapters.

Nils Nießen (2008): Leistungsgrößen für Gesamtstraßenknoten (key performance indicators for nodes and marshaling yards where large corridors join).

In this paper the author presents the results of work on an attempt to calculate the theoretical capacity of a large node in a railway network. One characteristic of large nodes is the fact according to the author, that it is possible to handle several routings in parallel. This is possible since through the large number of different sidings and tracks it is possible to handle them simultaneously and independently without higher risk of accidents. In his work, the author uses the calculation methods previously developed by other authors especially the STRELE-formulas from Schwanhäuser and enhances them for calculating the theoretical capacity of railway nodes. Based on the queuing theory, he uses the STRELE-methods for parallel server systems, but also taking into account that there is still interdependency between them. He first divides the large node into separate independent sub nodes, calculates the theoretical capacity for each sub node and then combines the result to a capacity of the total node. In his work he defines a waiting probability of trains in siding areas. This is the relative amount of trains, which have to wait at a siding area, resulting in adding secondary delay on the train schedule. He also gives an example of such waiting probability, see Table 3.

	Passenger stations	Freight stations
Standard value	P_w= 2.5%	P_w= 5%
Limit value	P_w= 5%	P_w= 10%

Table 3: Practical allowed waiting probability P_w in the railway network of DB AG (Nießen, 2008)

As described in UIC 406, also in this work the basis for calculation the key performance indicators of a railway network is the compression method. But in siding areas in contrast to tracks, there is a possibility of parallel handling with respect to time. Therefore the capacity (as well the theoretical and also the practical capacity) depends not only on the speed and timetable (together with the given infrastructure of course) but also on the routing. So the minimal (theoretical) headway for certain routings can be zero, because of the layout of the siding area. The author admits that simulation techniques may be an alternative way to calculate the capacity of large nodes but denies this possibility for his work due to calculation time restrictions. With other words, simulation is not suitable for real world problems, according to the author, because solutions calculations take far too much time. Instead he uses an analytical approach to calculate the waiting times with the help of an enhanced STRELE-formula. To get from the theoretical capacity as the maximal possible capacity just before collapse of the railway network to practical capacity he introduces adjusting factors. To verify these adjusting factors he compares his results out of analytic approach with simulation results for very simple example nodes. This comparison reveals that the newly developed factors will improve calculation results for the practical capacity of large nodes compared to the unmodified STRELE-formulas; however the author realizes still a difference between calculation and simulation. Unfortunately, there is an overestimation of the practical capacity compared to the simulation results, which the author is unable to reason.

Chr. Schmidt (2009) Beitrag zur experimentellen Bestimmung der Wartezeitfunktion bei Leistungsuntersuchungen im spurgeführten Verkehr (in German)

In this work, a dual approach to determine the interdependencies of infrastructure in railway networks and the quality of service (using the queuing waiting time as key performance indicator) is implemented in a computer system and further developed. The idea behind this approach is to use first several simulation runs with a model to calculate specific values which are characteristic for the railway network and then in a second step to use regression analysis to draw further conclusions. In the first chapters of her work, the author deals with the simulation of a railway network. She reminds the reader that for a correct simulation run, there has to be a warm-up phase in which the simulation objects – the trains – are inserted into the previously empty model, which represents the railway network of interest. Then there follows the simulation period, where trains arrive to and leave the modeled network. In this period the key-performance indicators are measured. In the last phase of a simulation run, the model area is emptied, as no further trains enter the model and all trains inside the model area leave. The reason for neglecting the warm-up period and the slow-down phase in the evaluation is because of the fact, that these periods are not representing real world situations but only necessary for calculating reasons. The author identifies the capacity of a railway network as a key performance indicator. For calculating this value she identifies two extremes in a railway simulation as of interest: the random timetable and the compression of an existing timetable to a new optimized timetable. She defines the random timetable as a timetable where trains of different categories arrive randomly at the borders of the modeled area, are handled inside the simulation model and leave the simulation afterwards. In this random timetable, it is possible in principle that deadlocks happen, due to conflicts in routing which cannot be solved in the simulation run. In contrast to this the existing timetable is having no such deadlocks, so after correct application of the compression method there should still be no deadlocks or conflicts, but there might be some scheduled waiting times added to the routes. In her work, the author uses a compression method which leads to a timetable with a regular pattern, the so called periodic timetable. Remark of the author: such periodic timetables are well known in railway literature; see for example (Goverde 2002) where the whole calculation method, the max-plus algebra, is based on such periodic timetables. In her work the author compares the results of both methods and she shows that with the railway network she evaluated the optimal capacity (remark of the author: according to UIC 406 this value would be named theoretical capacity) is lower for a random approach than for a periodic approach. The difference was calculated in the range of five percent of the theoretical capacity. So also in this simulation runs it has been demonstrated that variation - here the variation of the arrival times in the randomized timetables - has a decreasing effect on throughput of the system – here expressed in the diminished capacity of the system. To calculate the simulations a synchronous simulation method, the program system RAILSYS, has been used. The author decides to use only the random approach in the following calculations. But the variation of arrival times was the only source of variation which was used, so no variation of travel times and other random effects where introduced in the simulation runs. In another pre evaluation the author demonstrates, that there is an effect of the size of the simulated area and the simulated time on the capacity values as a main result of the simulation runs. But in her further paper not the capacity but the total waiting time of the trains inside the simulated area is used as a key performance indicator. In this work waiting time is defined as the difference between the (simulated) measured travel time, which can also be called as realized travel time, and the planned travel base time. The planned travel base time is the time, which results from infrastructure and train characteristics only, and is calculated independent from the existence of other trains in the railway network. The author states, that for the determination of waiting time with the dual approach it is essential to know the theoretical capacity; however for this case the waiting time is infinite, as the queues in this case are growing without border. Therefore it is not possible to calculate the waiting time in the case of operation at the network´s maximal capacity. The author states, that in previous papers with simple network models where only one feeding block was present it was assumed, that maximal capacity was reached, when the feeding block was occupied 100% of the time. Under this assumption, the maximal (theoretical) capacity equals to the number of trains leaving the system in a predefined time. Another possibility for reaching the point of maximal capacity in the network was to feed more and more trains into the system, until no further increase on the number of trains that leave the system per time period could be observed. Then it was assumed that the network was running at theoretical capacity. In contrast to these two older approaches the author of this paper proposes two alternatives. One is to look at the ration of entrance and outlet occupancy, the other possibility is to evaluate the dependency between waiting time and the time used for evaluation of the simulation run to estimate the theoretical capacity of a given network. These alternatives have the advantage that they can also be used when multiple feeding blocks in the system are existing (Sketches see page 71 to 74 of Schmidt 2009).

To determine a waiting time function with respect to the number of trains inside the simulated railway network (also called the occupancy of the network), it is necessary to look not only at the extremes (no waiting time if the number of trains inside the systems equals to zero and infinite waiting time when the number of trains is at theoretical capacity) but also values in between this interval have to be simulated to get an idea about the interdependency. Unfortunately the author does not give hints neither how many simulation runs must be performed nor which occupancy steps should be used to get the best results. Only the suggestion is made, that the data values should be distributed over the whole performance range. Nevertheless she comes to the conclusion, that for the approximation of waiting time function that logarithmic approximation procedure gives best results.

In a final chapter the author draws conclusions from looking at the approximated waiting time function. Based on her literature review she holds the view that the optimal performance point is easily calculated by searching for the value for which the waiting time neither decreasing nor increasing over the track occupancy. Finally she applies this method to several different test cases.

2.3.3 Attempts to integrate random effects in the simulation studies of railway systems

There are already some scientific papers published containing reports about simulations with integrated effects of variation published in literature. Some of them are recapped in the following lines.

Clymer (1995) System design and evaluation

In this paper, the author describes an expert system based approach for simulation. The simulation runs are used to derive optimal decision making rules for system control.

As an example, the author uses the operation control in a railway network. He characterizes the control of such a network as highly context sensitive and quite complex, compared to other systems. Therefore the decision making process is more difficult –according to the author - than in most manufacturing processes, where experts systems have been applied before. In his example network, which is a single track network with altogether five stations, there is one designated start station, one end station (at a distance of 100 miles away) and three overtaking stations in between. In an attempt to integrate random effects in the calculations, the author assumes the inter arrival times to be following a gamma distribution, additionally the travel speed of the two different train types on the track (freight trains and passenger trains) is also assumed to follow this distribution. The author develops algorithms in his work, which he tests in a simulation runs, where he is able to demonstrate the ability to control the railway network. As in this network, quite easily deadlock situation can occur, for example when two trains running in different direction try to occupy the same track at the same time, so the prevention of such deadlocks is one main task for the control strategy. However, the two different approaches which were automatically developed within a training phase do not show such difference in performance over those rules, which were constructed manually. In contrast to the similar results presented, the author claims a notable difference, especially with respect to finding solutions to unexpected events. The author tries to explain this hypothesis with the help of system theory. He thinks that as machine learning is based on expansion and operational models (top-bottom approach), it must be better than rule based (human experts) strategy, as this is using mechanistic models and is therefore always based on reduction methods (bottom up approach), according to the author.

M. Schachtebeck (2009) Delay management in public transportation; capacity, robustness and integration

In his paper- part of a dissertation- , the author first gives a definition for timetable construction and delay management. According to his definition timetable construction is part of the strategic planning phase to later operate a railway network successfully. In contrast to this, the delay management is part of the operational phase, where short-time decisions are necessary for successful operation with respect to the unexpected events occurring in real operation. In his work the author concentrates on research on delay management and neglects the timetable construction. With his contribution he wants to develop a method to support railway network management in operation. He tries to describe the interdependencies of a railway network in a set of mathematically equations with certain additional constraints. This set of equations is developed by looking at the network as a summation of several event-activity networks. By using this approach he is able to calculate the secondary delays in a given railway network, when primary delays and the fixed (planned) timetable is given. From the event-activity networks, several equations result. As all networks are dependent in a certain way, they are combined to a full set of equations which could describe the underlying rules of the full network. By searching for the optimum in this set of equation the author is convinced to get also to the optimal delay management rules. So he uses the well-known strategies of linear integer programming to derive the optimal point out of the set of equations. He is able to demonstrate the procedure on small scale networks. However in his next chapters the author admits that for real world problems it will become very complicated to solve the necessary set of equations and given constraints to find an optimum. So he attempts to use different heuristic approaches to get to an optimal delay management for larger network problems. He then compares the results of the different heuristics, some of them are priority-based, others use a relax-and-repair approach and most of them are combinations derived from these with different weights. Even as this is only performed on relatively small examples, the author is not able to present a fully satisfying solution for the heuristic decision problem (which heuristic should I use to get as fast as possible to a good optimal solution).

In the last chapters of his work the author tries to adapt this approach also on the more strategic decision of choosing an optimal timetable. But as his approach was only tested on some special cases, he is not able to show the applicability of the methods developed for operational decisions also on strategic issues.

Büker(2010) Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen

In this paper – also part of a dissertation- the author attempts to use an analytic approach to calculate the waiting times and the secondary delays for networks. He states, that in real world problems it is impossible to exclude delays from operation in a railway network. He therefore introduces distribution functions, which represent the variations in real world operation, in the calculation of the operation times into the analytic approaches to calculate the delays and the waiting times for a given network, which he found in older literature. He uses algebra to solve these functions, which also includes the randomness of process times. From these resulting equations he further develops recommendations regarding the robustness of timetables.

The author develops not only the equations; he applies these calculation schemes also on some example networks. He compares the results when applying this with the results of different simulation runs. The calculation of small problems reveals a good approximation, but some numerical problems had to be solved during the calculations. For example to get satisfactory results during simulation quad-double data types had to be used to achieve the necessary high precision. The author discovered a connection between precision and running-time of calculation, which seems quite trivial. If high precision is desired in a calculation, there is often a longer calculation time needed to achieve this. In the work, he was able to demonstrate that when simplifying assumptions were made, the calculations were fast but very far from the precise results. The mean error is reported to be between 1500% and 3200%. When more elaborated parameters were used for the calculations, the results are somewhat closer to the results of simulation, here 140% to 230%, but the running times increases by a factor of 800 to 2600. So unfortunately, the aim of the work to evaluate different timetables with respect to their robustness to delay propagation could not be solved satisfactorily. The author himself admits that further studies are necessary.

UIC (2010) Einfluß des ETCS auf die Leistungsfähigkeit von Knoten (version 1.1)

As the authors state on their second paper on ETCS, the European train control system, and its influence on railway capacity, the effects on train operation are an important aspect for the justification of introducing this train control system. In this paper the approach of the first paper (UIC 2008) about the effect of ETCS on the capacity of lines is enhanced to railway nodes. The authors state that in contrast to their paper on lines, the prediction of the capacity of nodes cannot be handled in an abstract way, so only the evaluation of existing nodes with their traffic and track layout is done. Two existing nodes are evaluated, one dead-end station, the main railway station of Munich in Germany, and one transiting station, the railway station of Bern in Switzerland. After an introduction into the methodology of calculating the capacity consumption of a railway network with a given timetable, the compression method, the authors present the results of applying the same rules on the timetable of two example nodes. The results of such calculations reveal that compared to the actual status of these two stations there is not much possibility for further capacity gain with the introduction of ETCS level 1. As the main reason the authors identify the already much elaborated timetables and track layouts of these stations, which have been optimized since decades by expertise gained from operational management. However a small increase in the number of possible train routes without overload seems possible with the introduction of ETCS level 2, as some of the block distances inside the nodes can be shortened and optimized to slightly increase the mean train speed during approach and departure. The authors mention, that there is a theoretical, notable advantage with the introduction of ETCS level 3 with respect to capacity of both nodes, up to 30% compared to the use of ETCS level 1 seems possible, but compared to the capacity increase in lines, as revealed in UIC 2008, it is not as obvious. As a result of their work the authors recommend to enhance the existing UIC code 406 for capacity of railway networks with chapters about the special aspects of capacity calculation in nodes.

Nils van Oort (2011) Service reliability and urban public transport design

Based on figures from European Union and other public sources the author of this paper realizes an actual major trend in transportation, there is increasing desire for personal mobility in America, Australia and Western-Europe. Due to the transportation systems already working at their capacity, he expects an increase also of the societal costs of this mobility, the loss of time due to congestion, the increasing possibility of getting injured during driving and last but not least the damage to the environment due to emission and noise. He then identifies the increase of the use of public transportation as one possible alternative amongst others. However he states that the attitude of potential customers towards public transport must change to increase its usage. One main factor for this attitude is according to the author the quality of service. Indicators for such quality are the price, the accessibility, the travel time, the comfort, the image and the service reliability. Under the last term he defines the ability of a service to meet the expected quality aspects such as waiting time and comfort and travel time. To further describe service reliability is one of the main topics of this paper, as well as the development of planning instrument especially for tram and light railway services to increase the networks ability for such service reliability. He concentrates on the service variability and (un-)reliability, stating the hypothesis that there are only some main reasons for poor service reliability and lack of acceptance of public transportation. He reasons his hypothesis by citing surveys and empirical data. On a practical example, he then presents different approaches to improve service reliability, especially with an improved timetable. After a discussion of some reasons for variability like driver´s behavior, schedule quality, vehicle design and passenger behavior he presents different operational instruments to improve the service. He then presents the results of service reliability in the urban transportation network of The Hague, a town in the Netherlands, before and after the implementation of such improvements. He is of the opinion that these results demonstrate that a significant reduction of variability can be achieved. By looking at the figures he does not only take into account the mean value of the different key performance indicators like punctuality but he realizes that also the variation of these values, mathematically expressed by their standard deviation, is very important and has to be taken into account. Based on findings in the example from The Hague he draws the conclusion that there may be a potential for saving in money by applying five planning instruments, which have been developed by the author. In the case of The Hague, he estimates that the savings alone add up to the sum of eight million Euros per year just by improving the service reliability. But he admits that there must be a trade-off with the cost of these planning instruments. In his example he estimated the additional cost for implementation of these instruments by around three million EURO per year. In his final conclusion the author states, that before adaptation of his improvement methods in other urban transportation networks, a case by case analysis has to be done, as there exist a lot of different operational strategies between the different systems.

SMA (07/2011) Audit zur Betriebsqualitätsüberprüfung Stuttgart 21

The authors of this paper deal with a special kind of simulation to verify the results as its outcome. The simulations were performed to verify the possibility of a planned new railway station to handle all the traffic, which will be estimated in the future, in a satisfactory manner. In Stuttgart, a town in southern part of Germany, the existing railway station is a dead-end station. The infrastructure management of the German railway as owners and operators is of the opinion, that the existing station and its layout which dates back to the twenties of the 19th century is the cause for many delays, due to overload. Therefore a new station has been planned and is currently built. But there was doubt between different parts of the public if the new concept of the current operator is really able to provide the necessary capacity for the traffic which will be handled in some years from now in a more reliable way. Therefore a simulation study was performed to demonstrate this ability. As the German railway together with some sub suppliers is currently the only company with the necessary resources to perform such a study, it was agreed by the opposing parties to perform an audit of the simulation to increase its acceptance. The results of this audit are published in this paper. Some of these results may be also valuable for railway simulation as a whole, especially with respect to those simulations dealing with the determination of the capacity increase due to changes in infrastructure.

In the first chapters of the audit report, the basic definitions and procedures are given, together with a description of the audited simulation method. The authors state, that even if there is an accepted definition for capacity (the UIC leaflet 406) in railway operation, there is still no standard for evaluating railway specific simulation studies. The authors give a short description on the methods, based on a three-step approach. In the first step the infrastructural input data is tested. In the second step the underlying timetable and the associated data is tested, whereas in a final third step the simulation methodology, the calculations and the results are tested.

The main problem of the authors was neither the input data nor the test of the simulated timetable. The main problem revealed in the audit was, that the main aim of the simulation study was quite poorly defined. The agreed target of the simulation was to proof the good quality of service of the new infrastructure. This definition seems not to be concrete enough to satisfy neither the performers of the simulation nor the critical views of the auditors. The people responsible for the simulation calculation used another key performance indicator instead of this poor quality of service. The performance indicator used was the delay propagation inside the simulated area. If the delays of the trains introduced into the simulation (only very few of them had predefined primary delay) were not increasing inside the simulation model during the simulation then the authors claimed a good performance. The auditors gave also the impulse to perform a sensitivity study, where additional simulation runs were performed with slightly changed input data or different infrastructure. As a result, the auditors saw also slightly different results out of the simulated runs. This led to the auditor´s conclusion that the behavior of the simulation is somewhat stable also with slightly other input data, so the behavior of the railway network is predictable and the results of the simulation are reasonable.

For the delays that were taken into account in the simulation, a negative exponential distribution was assumed, where the mean follows the following Table 4.

Type of train	Percentage delayed	Mean delay/ min	Max delay/ min
Passenger, long distance	50%	5.0	60.0
Passenger, regional	60%	4.5	30
Urban trains	25%	2.0	15
Freight trains	60%	10.0	60

Table 4: mean delay of trains in the simulation (SMA, 2011)

2.4 New Approach to simulate and evaluate different infrastructure of railway systems

The literature review in the proceeding chapters reveals that there are multiple attempts made to predict the capacity of a railway network. In fact, there are a lot more possibilities than just the approach with simulation, but also a lot of analytic methods. But a main obstacle for these methods is that they are not able to include the random as a factor.

But especially in the operation of a railway network with its interdependencies close to practical capacity it becomes obvious that the influences on normal operation have to be included to get reliable results. Therefor each new approach for calculating the capacity has to include the stochastic nature of operation. With a simulation model, this could be easily managed. But especially when using simulation to predict the operation in future configurations, there is one main obstacle: the environment is not known so the influences and interdependencies for the railway network could only be estimated. But how can one get reliable results in such a vague environment?

If we have a look in the literature (e.g. Law, 2007) there are already reliable concepts to compare different alternatives, in an environment full with randomness, by using the concepts of statistics, especially the concepts of analysis of variance. But unfortunately, these concepts are not used in simulation studies for railway operations, or their use is not documented.

Even if we are cannot be sure if the model used in a simulation study is able to predict the reality of railway infrastructure "totally correct", we use not the absolute value of the capacity and take this value as a given constant, but we compare different alternatives. (By the way, the assumption that there might be a correct model in a simulation model is wrong, since a simulation model is by definition a model, and not able to include all possible influences. Therefore a simulation model can only be correct under some constraints and assumptions and only with regard to a special purpose. For a deeper discussion of this fact, see for example Sargent (2010). So the influence of (possible) errors in the prediction may be reduced, as long as we use the similar models (with hopefully the same errors) to predict the future performance. But in contrast to the studies performed for UIC (Via, 2008 and Via, 2010), the new approach does not only compares the means of capacity results from simulation runs (display of the capacity changes in form of relative capacities), but also takes into account the variability of capacity calculations by using analysis of variation (ANOVA), too, to compare different alternatives. It is estimated that similar to other fields of application of ANOVA the results of such usage are more reliable and are able to predict the real behavior in a better way.

3 DETAILED PROBLEM STATEMENT

3.1 Simulation approach in railway systems

In the first moment, one has to consider the aim of the simulation. If a simulation is performed just for evaluating different types of infrastructure (not considering different timetables, signaling systems etc.) a very simplified model may be sufficient. However, for most of the railway simulations, it is essential to simulate the dynamics of travel to an appropriate level, as this dynamics influence not only the time of travel, but also the distances between trains for safe operation. If one looks into the literature of travel dynamics (see for example Wende, 2003) one could come to the conclusion that the calculation of dynamics in the railway business itself is complicated enough to develop software simulation packages for this reason alone. In fact, these dynamic calculations were one of the first applications of simulation in the railway business. So the full simulation of the travel dynamics of modern trains could be a potential area for a simulation paper.

But in the case of a study about different types of infrastructure, the full details of travel dynamics are not the main focus. In this work, the travel times are simulated as delay (process) times. The duration of delay is approximated via a function of travel distance and entrance speed, depending on the different train types (a high speed train has other accelerations than an urban train), the allowed maximum speed and the signals. If the signal is at halt, the train has to brake, whereas the signal´s aspect is at travel, the train may accelerate until the maximal speed of the track or the maximal speed of the train has been reached.

The coefficients of these functions have been derived via regression analysis from excel-calculations of travel dynamics using stepwise numerical integration (depending on a constant travel distance). For the proposal calculations were performed with a constant distance step of 50 m, in the dissertation it has to be demonstrated that this approach is appropriate. In the literature, at no point exact equations nor the way to describe them was given. This was true with one exemption. In the paper Influence of ETCS on line capacity, VIA 2008, it is stated on page 29, that the acceleration of the trains after a scheduled stop is calculated with an exact method using the parameters of the driving dynamics (delta-v-step-method). However, the method itself is neither described nor a literature is given. Therefore it was a main task to develop the calculation rules. In the Appendix to the final paper, these calculations and the mathematical basis behind the will be described in detail.

3.2 Influence of infrastructure on railway operation

It is obvious, that the infrastructure has a high influence on the operation of railway systems. Therefore, the infrastructure has also a high influence on handling and process times and in consequence on the capacity of the track. Some influences are quite obvious, like the increase of possible train numbers due to a double track layout compared to a single line layout. But some infrastructure measures are not so easy to evaluate in their effect on the capacity of a given track. For example, the release of a crossing station in a single line layout may be of no effect for normal undisturbed operation whereas in the event of a sudden failure it may have the result that further traffic is not possible because the railway (sub)-system is running in a deadlock which cannot be handled in an easy way.

3.3 Present approaches to predict the future behavior of railway systems

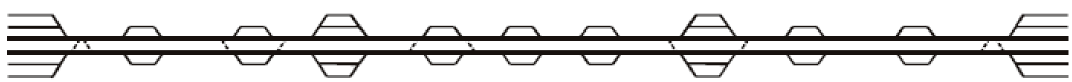
The paper (UIC, 2008, Influence of ETCS on line capacity) gives three different scenarios for the evaluation of the influence of ETCS, the European train control system. The first scenario consists out of a two-track configuration for high speed traffic. The maximal allowed speed of trains is 300 km/h. The distance of stations is 100 km; the length of a block section is given to 5 km. The whole free line is considered only. The entrance and exit speed of trains, leaving respective entering the station at the border of the scenario is 100 km/h only. Two different train types use this track layout: the premium train, like ICE or Thalys, which is capable of going at a speed of 300km/h, and a train type called in UIC 2008 Eurocity. The Eurocity is a train traveling at a maximal speed of 200 km/h. Each of the tracks is used for directed traffic, so no mixture between train directions is considered. For a graphical representation, see Figure 5.

Speed	300 km/h
Distance of overtaking stations	In the distance of published stops
Length of block section	5 km
Total length of the line	approx. 100 km
Station at the beginning and end of the line	Large station with discharging lines of different categories
no en route stations	
Entrance/Exit speed	100 km/h

Figure 5: Scenario 1: High speed line, track layout (VIA, 2008)

On the tracks, 80 trains per day are High-speed trains and 40 trains per day are Eurocities, so only long-distance traffic is considered.

The second scenario (see Figure 6) consists of a main line track layout, also a two-track layout. Here on the line of 100 km, there are two primary stations at the start and the end of the line, two secondary stations on the line and seven tertiary stations in between (for the whole line). The entrance and exit speed is fixed to 80 km/h, whereas the maximal speed on the open tracks is 160 km/h. In addition to the high-speed trains and the Eurocities, which make a halt only at the two primary stations, four different other train types are introduced. A fast regional train which stops at primary and secondary stations and a slow regional train which stops at every station on the line are added. In this scenario, also cargo trains are considered. There are two different types of freight trains, an inter-regional cargo train, going somewhat faster, and a slow regional freight train. The initial block distance for this scenario is 3 km.



Speed		160 km/h
Distance of overtaking stations		In the distance of published stops
Length of block section		3 km
Total length of the line		approx. 100 km
Station at the beginning and end of the line		Large station with discharging lines of different categories
Totally nine en route stations, thereof		
	Two	Large stations normally without discharging lines
	Seven	Overtaking stations without discharging lines
Entrance/Exit speed		80 km/h

Figure 6: Scenario 2: conventional main line (VIA, 2008)

The third scenario has a single line layout, where on the distance of 100km two main stations, at each end, and 4 intermediate (minor) stations with the possibility of crossing and overtaking are positioned. The full scenario is displayed in Figure 7. Such a scenario is characteristic for a regional line.

Speed	80 km/h
Distance of crossing stations	15 km
Length of block section	without block sections
Total length of the line	approx. 100 km (single-track line)
Station at the beginning and end of the line	Large station with discharging lines of different categories
Totally four en route stations	crossing stations
Entrance/Exit speed	50 or 40 km/h

Figure 7: Scenario 3: regional line (VIA, 2008)

For all scenarios different alternatives in the infrastructure are possible. At first, the reference situation is defined as the operation with ETCS according to Level 1. A second alternative for the high-speed line scenario as well as for the scenario with the conventional main line is the use of ETCS level 1, but with a second infill balise. As the alternative possible for all three scenarios it is defined to apply ETCS level 2 for all scenarios. For the high speed line and the conventional line the influence of application of ETCS level 2 but with 400m block sections is examined. The final alternative the introduction of ETCS level 3 (moving block system) is also investigated in (VIA 2008).

In the original study (VIA 2008), the results were evaluated by comparing the different mean values of the capacity. As these results were obtained using deterministic equations (either the STRELE formula or the compression method according to UIC 406), there was no variance in the received results, so this approach is the only method to compare them. With the new approach the differences of behavior of systems with different infrastructure could be evaluated in addition with their ability to handle variance. This should be based only on determination of statistical significance. So effects of infrastructure which are only minor compared to influence of other natural effects could be easily separated from major effects.

The new approach should be able to calculate similar results when no or only small variation is introduced into the railway network. But a major advantage of the new approach is the fact, that it is possible to introduce variation into the network based on statistical evaluation. With this approach it is expected that results of simulation runs should be different from those obtained before, resulting in a decrease of the (theoretical) capacity obtained without introduction of variation to a value close to the practical capacity. Actually the most important performance indicator – the practical capacity – is calculated by using a safety margin below the (theoretical) capacity in the area of 30 percent. This value, which can be found in the UIC leaflet 406 on capacity, was developed empirically upon the experience of several railway operators in Europe. With the new approach developed in this work it is expected, that this safety margin could be diminished, if not eliminated at all. This approach can therefore increase the confidence of the stakeholders into the results of simulation studies.

3.4 New approach for capacity calculation with general-purpose simulation packages

If we look into the existing literature we realize two different problems with the existing simulation-based solutions for capacity research on railway-networks:

Either the solutions depend strongly on the use of special simulation software with all the advantages and disadvantages of using such application-oriented simulation packages (examples are the calculations by Goverde and others using the max-plus-algebra and the railway simulation tool ANKE used by Vakhtel and others) and are mainly based on mathematical calculus. Or the authors use standard simulation software but do not make use of all the possibilities of statistics to further base their conclusions on (for example the use of rules of thumb to calculate the necessary numbers of different simulation runs as seen in SMA & Partner 2011). Additionally there is a lack of the inclusion of stochastic effects like variance to the results in the evaluation of simulation results. An example for this can be found in the Italian study from Confessore, Liotta et. al. (2009). Only lately also another main deficit of existing concepts to determine the capacity of a railway was described in Chu (2014): the existing concepts always calculate the capacity as upper limit of a steady-state simulation, whereas in practical operation such steady state is only quite seldom in railway operation (for example in some urban metro systems). In most cases the variation in traffic is quite high along the day. For example the times of high traffic demand in the morning, when people and goods need to be transported towards factories, differs very much from traffic amount in the night time. A factor of 10 to 1 is not uncommon. Also this paper (Chu 2014) reveals, that with the traditional simulation tools used in transportation research like the concept of time slices this behavior cannot be predicted. Therefore high mathematical and computational effort is needed to add additional coefficients in the existing model functions. On the other hand the application of simulation in this case loses much of its transparency. In this paper (Chu, 2014) it is demonstrated, that the effect of transient phase could not be neglected by contrasting average waiting times in steady state with those of calculations with limited preheating times. Especially when high utilization close to capacity limit is assumed, the average waiting times may differ by a factor of two and more.

Therefore we see an actual need to develop a new approach to evaluate more accurately the influences of infrastructure changes on the capacity of railway networks. This new approach shall make use of standard simulation software and simultaneously the principles of statistics for comparing different alternatives in infrastructure in a railway simulation study. By this novel approach such studies may improve in multiple ways.

First, the simulation itself will become more transparent, since standard simulation software, like ARENA of Rockwell, will be used. This eases the process of validation and verification since graphical methods like animation are already included in these simulation software packages.

Second the use of statistics will become much easier. This is because there are a lot of statistical features already available inside these standard software packages.

Third the comparison of railroad simulation with other applications of simulation will be much easier if the same software tools and basic approaches are used than just searching for an alternate solution. As the power of computer systems is increasing very fast, the disadvantages of using standard simulation software (longer simulation times, more complicated input procedures etc.) may no longer be so important.

As a fourth benefit of using standard simulation software one could regard the bigger ease of sharing good practice with other experts in simulation. Some of these benefits were already discovered in previous studies on railroad operation like in Firenze (2005) and Confessore (2009). But in contrast to these attempts the main reason for the introduction of this new approach is the introduction of variance into simulation.

It is expected that by applying the concepts of statistics in the evaluation of simulation results these results will be much more reliable. Also the accuracy will be improved, since re-modelling of existing scenarios could be used to adjust the simulation model.

4 SOLUTION METHODOLOGY

4.1 General

In this paper two different solution approaches are presented.

With the first approach, which is mainly based on the physics of train travel along a track (with acceleration, travel at constant speed and deceleration), it was expected to get a good representation of the actual train movement. It is demonstrated, that this approach is able to calculate the (theoretical) capacity of quite simple railway networks (like in the first scenario) with neglecting variance. However when variance was included in the model, this approach led to a dead-end as calculation times increased enormously. Additionally with this approach it was necessary to divide the real network into two different simulation models. One for the acceleration (including travel at constant speed) and one for braking. It will become quite complicated to determine the necessary connection points for changing from one to the other model, especially when more elaborated signaling concepts have to be simulated. But with increasing computational capabilities, this approach may lead to very accurate simulations. Therefore this first approach is included inside this paper, even if this approach was rejected.

In the second approach the possibilities of the ARENA software package were used to simulate train movement. Inside ARENA, there exists a model for automated guided vehicle

(AGV´s) called the guided transporter module. This module is capable of simulating acceleration, constant travel and even braking of a vehicle. Additionally this module is possible to handle the movement along a given network only. As a further advantage, simple blocking rules are also already included within this module. Unfortunately, this module has to be modified to be capable of some of the basic concepts in railway operation: scheduling and overtaking.

To overcome these obstacles, two new ideas were introduced into this module:

At first it was invented that the transporter was only a tool for transporting and not the train itself. The train is an entity, which is created in the first station of the model network, requesting for a free guided transporter (with matching characteristics). When the guided transporter has picked up this entity, it will follow the entities schedule along the network map until the final station. The entity (the train) will then be disposed, after the statistics of the travel have been recorded. To simulate different trains (with their specific characteristics in acceleration, top speed and braking capability) a matching number of different guided transporters have to be provided.

Secondly there was implemented a possibility for overtaking. Trains with lower priority (either freight trains or slow passenger trains) will always use the sidings in the stations as fixed in their entities schedule. If not, that only passenger trains make a regular halt in these stations for boarding and debarking of passengers. Freight trains will normally pass through the station without halting. This is introduced into the model with an additional entity variable, called waiting time. When such a guided transporter with its entity on top comes into a station, the position of all other guided transporters will be evaluated. If one having a higher priority is close to the corresponding station, the waiting time for that entity (including the matching guided transporter) is increased to allow safe overtaking. The top priority trains will use the direct tracks running through the station. After passage of the faster train, the slower train will then move back on the main track.

Additionally the rules of blocking were introduced into the simulation by dividing the tracks between stations into several sub-tracks, which have the length of the blocks in the signaling infrastructure.

Simulations with this second approach demonstrated the capabilities even in more complex scenarios. Additionally it was possible to demonstrate the use of visualization as an important tool in simulation. Additional findings, like the possibility of automated search for an optimal timetable, were made with this second approach.

Simulation with variation may be used to handle two different operation problems within railway networks. First a simulation study can be performed with routine delays, which may be handled without the need to establish new timetables and train routes. Otherwise simulation can also be used to handle rare, major delays such as a complete train breakdown resulting in new train routes, a new timetable and cancellation of trains.

In this work only the first operation problem is tackled, however the basic concepts developed in this work may be also applied to find a solution for major traffic disruptions. But some of the concepts necessary for that are quite complicated, like solving a blockade and rescheduling timetables while running a simulation.

Additionally it has to be noted that the calculation of the (theoretical) capacity of a railway network is required to reach steady state. In a system with variation, it may be hard to determine when such steady state has been reached. In this work, a warm up period of six hours was generally used, based on evaluation of model´s behavior. The steady state conditions were then calculated from the system´s behavior during 24 hours of simulation (using the same timetable again and again until one full simulation day has passed).

Therefore one has to distinguish between two ideas: Finding the theoretical capacity does not require variation, but for the calculation of the practical capacity of a railway network, the variation is essential.

As in real life, the network is almost never running at full capacity for the whole day, the transient phase before reaching the capacity maximum from an (almost) empty system is also worth to be considered (see also Chu (2014)). One main advantage of using standard simulation software (discrete event modelling) is, that even this transitioning phase can be modelled in an accurate manner.

4.2 Main ideas valid for both approaches

To diminish the risk of getting not statistically valid results, the author uses the procedures of simulation techniques for his work. With this approach, he attempts to assure that the concepts for simulation correctness from a mathematical standpoint are strictly followed.

For the determination of the capacity of a railway network, the UIC definition for capacity is strictly followed. That means that in a first step a railway network is described, either an already existing network with the necessary procedures and underlying safety and handling rules. This may be an already existing network or a fictive one, the latter necessary to compare such a different scenario with the original one. Then it is necessary to develop a timetable for the (real or expected) traffic on this railway network. In a third step the possible influencing factors on the traffic in the network are collected, described and evaluated. Such evaluation can either be based on data collected in the real world of railway operation, or a value derived from expert knowledge or from a combination of these methods. These three steps: description of the network, development of a timetable and evaluation of influencing factors are necessary to develop in the next step a simulation model, which should be able to describe the network´s behavior in a realistic way. After the development of a deterministic model for the undisturbed operation of the railway network in a first sub-step, the influencing factors are introduced by adding elements of random into this model in a second sub-step. The necessary validation and verification can be either made in a single step approach at the end of the development phase or (more easy and more useful) in a two-step approach at the end of each sub-step separately. For verification and validation it may be helpful to compare the results of the model with real data or with data gained from the simulation studies with the same or similar railway networks. For calculation of the capacity of the simulation model, the timetable has to be compressed until the system reaches the theoretical capacity, resulting in a sharp increase of waiting times in the system and/ or a slowing down of average train speed. For determination of this point, several simulation runs have to be performed, all of them based on the same timetable but more and more compressed. For the reaching of the theoretical capacity several different criteria have been described in literature. The author used the criterion of (dramatically) increasing waiting times for the trains inside the system to calculate the theoretical capacity. In the work of C. Schmidt (2009) it has been demonstrated that the waiting time is a good indicator for railway operation close to the theoretical capacity of the existing infrastructure. Then a very important step has to be performed: the design of a statistical experiment for this model to derive the necessities for a valid recommendation for system development and changes in infrastructure. This is performed to demonstrate that the results are of statistical evidence. As often performed in statistics, this will be done by developing a hypothesis and its counterpart. Such a hypothesis may be for example that the capacity of the existing railway network will be increased by more than ten percent compared with the existing status after the introduction of ETCS level 2. The contrary hypothesis may then be that the increase of capacity after the change is lower than that. To test this hypothesis the statistics provides different possibilities, like the t-test. The simulation runs necessary for this designed experiment has then to be performed. The results shall then be collected and evaluated with the t-test. If the hypothesis cannot be rejected with a certain confidence, it is very likely that the hypothesis holds. The problem must be solved in this way, as it is not possible to derive statistically valid results just by comparing the means of different scenarios. The statisticians have shown that for evaluation of different scenarios, especially those with large effects of random, it is necessary to use such approach. So far this approach has not been seen inside railway simulation community, perhaps since most of railway simulation was performed with specially developed simulation software packages, not designed for inclusion of random factors and for evaluation with hypothesis tests. But when the simulation is performed with a standard simulation tool like ARENA, it is possible to use the standard tools provided together with this software packages to perform such hypothesis testing very easily.

4.3 Model development, first approach

The development of the model for the railway network is split into several steps. The basic idea for the development is the decomposition of the infrastructure into a track subpart and a signal subpart. This procedure is illustrated in picture 8.

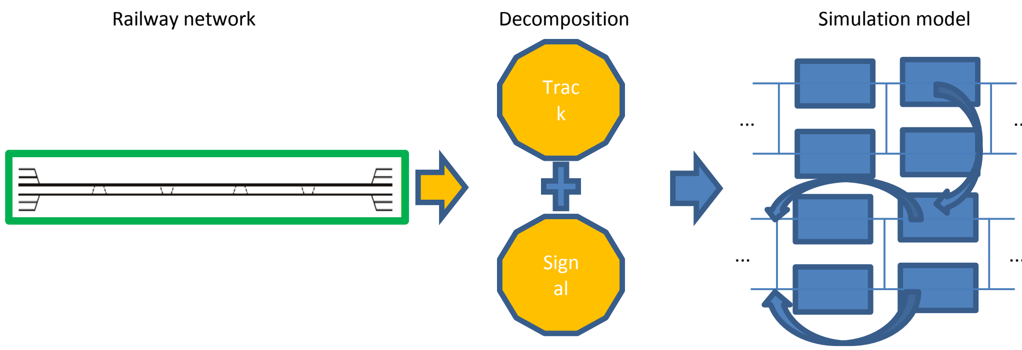


Figure 8: Basic concept for development of model

For the modeling of the track, the continuous process of a train travelling along a track is transformed into a discrete event model. The idea is, to simulate just the time delay of the train and to change the train´s position from the entrance of the module to its exit right after the delay has passed. This procedure is further illustrated in figure 9.

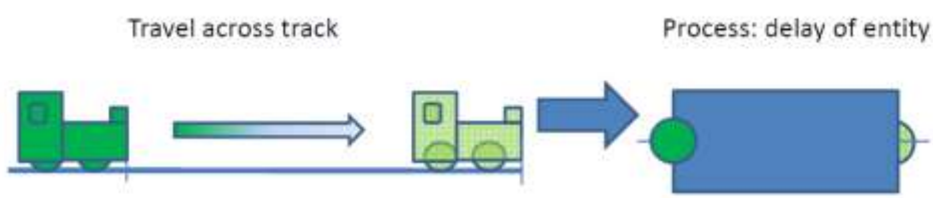


Figure 9: Development of delay model, simplified

For the simulation, the delay of the entity (here the train) depends of some variables like the type of train, the distance of the track to be simulated, the position of the throttle and so on.

In addition to this track model there has to be established a basic concept for introducing the signaling into the model. This is done by adding decision gates where the time of delay is changed according to the actual status of the signaling infrastructure. This idea is shown in figure 10.

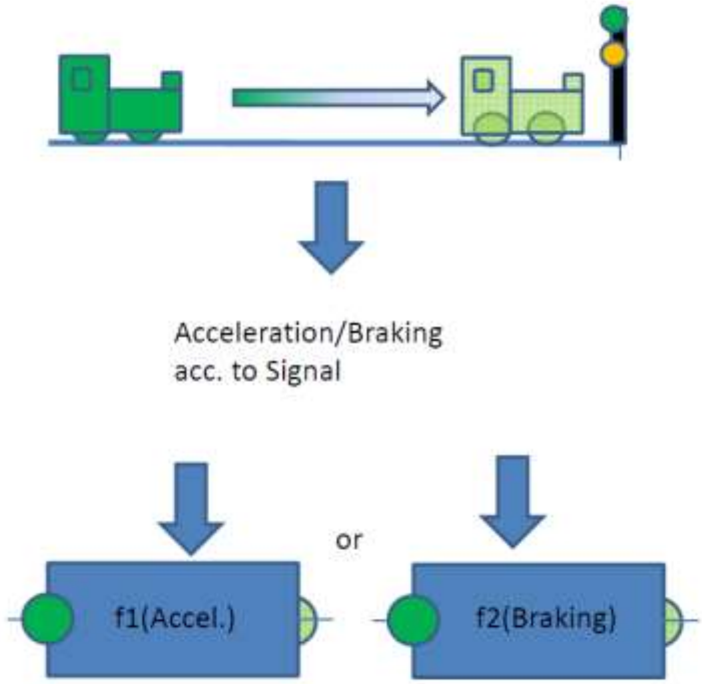


Figure 10: Introduction of signal status into the model, simplified

One has to have in mind, that with this concept, the status of the infrastructure is influencing the models behavior only at discrete points of time, only when an entity (here the train) is entering or leaving a delay module (which represents a certain distance of the track). Also, the signaling status has to be updated at each of these discrete points of time, as the signals change always, when the connected area (for example the block) changes its status from free to occupied and back to free. The concept gets even more complicated, when we take into account, that the braking and acceleration of each train may be different according to the train´s performance. So it is necessary to simplify the model, for example by allowing only a small number of so called train types (like in example 1 there are only two different types of trains). Also the signaling times are neglected, since they can be assumed to be very short compared to the travel times. For further explanations for the basic ideas to realize this concept within ARENA, the reader is guided to look in Appendix A and Appendix B. Also in 4.1 the basic idea is described.

4.3.1 Two different methods to approximate the motion of trains

As demonstrated in the Chapter 4.3 it is essential for a correct model of a railway system to predict the actual position and speed of the modelled trains, at least at decision points like positions of the signals. This model is then able to predict the behavior of the network when several trains are approaching a railway network at one connection at one time. Then there is a development of a queue at the connection, which then can be called the network´s entrance. With this model, even the transient phase during the development of the queue and the possible disbanding of the queue can be predicted. Furthermore this model is able to predict the reality when a faster train is running after a slower train. In this case the faster train is slowed down when he is approaching a block section, which has not been released by the proceeding slower train.

At first it was tested if a simple model consisting of the combination of a delay module (see figure 9) and an assign module (to set an attribute of the train entity to the new calculated speed) according to the block length. As there is the possibility of either acceleration or during the travel, both possibilities have to be considered by adding two of them in parallel. In this approach the delay time and the new speed is calculated by just by predicting the needed values. This prediction is done using regression analysis of the equations of motion. First tests showed evidence that this approach is able to predict the actual speed and the necessary delay with acceptable accuracy, but especially in the case of more complicated signaling methods like ETCS 2 and ETCS 3 and more types of train existing on the same track it was obvious that this concept was not achieving the desired precision.

However, there is a drawback in this attempt to use this approach. As demonstrated in Appendix A and Appendix B, a stepwise approach to calculate the actual speeds, the accelerations or the decelerations have to be followed. It has been discovered during the work with this modelling approach, that the accuracy of the model is reciprocal dependent to a major extent on the length of the step length, either measured in time steps or travel distances between calculations. This is based on the fact, that the approximation of a curve, here a curve with the highest coefficient linked with a power of two, by a line is satisfactory, as long as the steps between approximations are quite small. But especially when the train is reaching the borders of speed- either maximal speed or minimal speed- the stepwise calculation may lead to unwanted effects. It is expected that the train is maintaining its speed when it reaches the borders of speed, which are defined by the combination of the track including the occupancy and the train´s abilities. But with a stepwise approach two effects are possible when the borders are reached. On one hand there may be an unwanted "oscillation" in the train´s speed, as the speed in the first calculation step may be too slow leading to an acceleration; however then in the next too fast, which then results in a deceleration. But when calculating the speed in a third step, the speed may be again too slow, which then makes it necessary to accelerate the train again. The first of the "oscillations" has happened. These oscillation may continue for a long time, either stopped by accidental hitting of the correct value or by external intervention. With this approach, it may be possible to match the mean speed quite good, but a lot of calculation effort is necessary. On the other hand the simulation algebra could tolerate a minor deviation of the actual speed compared to the maximal allowed speed, which the leads to a sailing mode of the train, meaning that neither acceleration nor deceleration is applied. This approach leads to more even speeds at definition borders. But more important this may result in relatively incorrect travel times, as travel times are calculated as the integral of the speed over the time. But in both, the influence of simulation steps cannot be neglected. Especially when short distances within the railway network had to be simulated, like in the case of small block distances, these effects could only be handled with step distances in the range of 10 m. When a network length of 100 km is considered, this will lead to at least 2 x 10000 = 20000 small simulation blocks needed for the network model. Therefore already in the deterministic approach the simulation effort may become enormous.

4.3.2 Introduction of variation in the models

The simulation model, derived from the first model approach as described in paragraph 4.2.1, is based on the assumption that the travel time can be exactly calculated and predicted. These are the main characteristics of a deterministic simulation. But as demonstrated in the literature review in the second chapter the variance of trains is a main characteristic of railway operation. This variance has to be introduced in the simulation model to correctly simulate the network´s reaction, which is especially important for evaluation of signaling and infrastructural alternatives. This reaction is mainly the development of secondary delays due to the variation as primary delay introduced in the network. The variance could be introduced in the simulation model by several methods. First such variance could be introduced by adding random effects to travel times. Second the variance could be introduced at the borders of the simulated network by changing the dwell times according to random distributions. In the first method, the travel time has to be changed according to some measured random distribution. Unfortunately, the different travel times are not easily accessible, since travel times are complicated to calculate, not the only source of primary delays as we have seen in the first chapter and not published by the network operators. In contrast to this the delay of a train at a station is easy to measure and is even sometimes published by railway operators, since this time is the value the customer of a railway is interested in. Additionally this delay is not dependent on the possible source of the delay, so it is very convenient to use the second approach for introducing variance into a simulation model. However for the simulation it is important to introduce this delay at the correct point in the model as well with a realistic random based distribution. As such variance is trying to imitate the primary delay, it is quite logical to introduce the variance at the borders of the simulated network. By using this approach the reaction of the whole network can be analyzed by looking at the results of the simulation. In this case the delay has to be introduced in a manner, which mimics the realistic variation of measured train delays.

These delays have been in the center of quite a number of research studies, for example Goverde (2001), Yuan (2002) and Yuan (2006). To derive a good approximation, the delay has to be measured in train operation. Goverde (2005) has used the data available from network operation to measure the train delays. For the approximation of the statistical variation at first a look at the data is needed. A characteristic of railway operation is that the trains are either punctual or late. A train that is leaving earlier than noted in the published timetable is not allowed by operational restrictions, due to customer need. Therefore a lot of theoretical possible mathematical distribution functions are not able to match the given data, like the normal (Gaussian) distribution, because the data for the delay is always skewed to the right. But still several different distributions have to be evaluated. Yuan (2006) has evaluated the different models, including Beta-, Weibull-, Gamma- and Lognormal-distributions in relation to more than 10,000 records of delays measured at the railway station The Hague HS. Yuan found that the lognormal distribution is suited most to simulate the real behavior. It has to be noted, that the Yuan studied only the normal delay events. Outlying delays due to major events, like accidents which may lead to closure of lines for a certain time were not taken into full account. This leads to a major restriction of the actual work. Simulation with variation may be used to handle two different operation problems within railway networks: a) Simulation with smaller amount of delays, which may be handled without the need to establish new timetables and train routes. Or b) Simulation performed to handle a major event resulting in new train routes, a new timetable and cancellation of trains.

In this work - as described before in 4.1 - only the case a) is modeled, however the basic concepts developed in this work may be also applied to find a solution for cases b).

It is important to mention that all such approximation has to be followed by a sensitivity analysis since the estimation of parameters is one thing, but the results of a simulation study may depend strongly on these parameters. An easy way to illustrate such dependency and to remind the user to be careful in drawing results from simulation runs is the sensitivity analysis by changing the parameters by little amount and looking at the results. If results vary a lot if only one input parameter is varied by a very small amount, one has to be very careful in evaluation.

4.3.3 First attempts for verification and validation of the first model approach

At first, travel times and arrivals were held fixed (deterministic) to verify the basic model approach.

For a first step, a first and simple verification was performed just by looking at the animation available during the first runs.

In a second step, the time needed for the travel along the model (representing a distance of 100 km) by a high speed train without hindrance was compared with the time the travel would have taken in real world. Simulation came up with a travel time of 1201.96 seconds for the high speed train model. The calculation solving the differential equations comes up with a travel time of 1190 seconds for the high speed train: This is a good correlation since it has to be taken into account that simulated travel must be somewhat longer as the train enters the network with a speed of only 100 km/h equal to 27.78 m/s. Additionally looking at the animated run it was seen, that the travel times for the different delay blocks were all in a reasonable area.

Then a third step in verification was performed by checking the travel time for the second type of train, the intercity train. This train travels only at a maximal speed of 200 km/h respective 55.56 m/s, so travel time is longer. The simulation revealed that travel time for this train was 1805.53 seconds (without hindrance) which is quite close to the expected value. By this step, the basic concept in the simulation model, the usage of the same delay modules with different delay times to simulate the train movements, was verified.

In a fourth step, the high speed train and the intercity train travelled along the network in a short time distance. Here by looking at the animation it became obvious, that the model behaved as expected: no overtaking of trains was noted (like defined in the basic network) and the second train had to wait for release of blocks, if the first train has not left the blocks needed for hindrance free travel of the second train. Also the travel times of the second train increased when this occurred as well as the speed of the second train decreased. It could be also shown, that this happens only, of the travel schedule of trains is chosen very badly, as in most cases a hindrance free travel of both types of trains (40 IC-train and additionally 80 high speed trains on a single day) could be handled in a lot of cases. A case in this sense is a fixed timetable for the trains.

Then the arrivals were changed from deterministic into stochastic, to demonstrate the influence of variation for decreasing travel speeds and increasing travel times.

4.3.4 Results of the first model approach

Several calculations with ARENA were performed using the first model approach. For simple problems (like scenario 1 with signaling system ETCS 1) neglecting variation calculation times were within minutes using a simple laptop plus ARENA using the release 13.90.00. For calculation of more complicated problems (like the scenario 2 and networks with signaling concepts according to ETCS 2) the performance was poor using this computer configuration.

To get a reliable value for the (maximal) capacity of a system, the steady state of this system has to be reached. For a system without variation it can be assumed that this point is reached, when the number of entities (here trains) added to the system equals the number of entities leaving it. For a system having variance in arriving entities it is not as simple to distinguish the steady state. After evaluating the simulation, the warming up period was fixed to six hours (see also chapter 4.1).

4.3.5 ETCS Level 1 without infill

The test run was performed for a test period of 24 hours (with an additional warm-up period of 6000s). Ten replications were used.

The result was an amount of 63 Trains per test period in average for the effective capacity of the network. In the literature a value of the line capacity (theoretical capacity) was calculated to 129.1 respective 127.4 Trains per day. The two different results have been obtained with different calculation methods. When the results are compared, one has to take into account the Standard values of capacity consumption for congested infrastructure according to table 2 (here 0.6), so the value of 63 has to be compared with 77.5 respective 76.4. This shows that the calculation gives a somewhat conservative estimate of the effective capacity of the network. Further investigation showed, that this is mainly based on the fact, that the variation of train delays was somewhat chosen artificially to MIN(6, logn(4,0.75)) in minutes based on literature Yuan (2006). The function minimum(6,f()) was used, to suppress the effects of the tail in the function. These represent the events quite seldom in occurrence but heavy in their consequences. However in this work it is not the main topic to develop strategies to handle with such major events (due to accidents etc.) but to find possibilities to improve the networks capability to cope with the normal irregularities occurring in railway operation. Unfortunately it was not possible to verify this function, since there is a lack of public available data for distribution of delays in railway systems. But the results obtained in 4.2.5 demonstrate, that based on the assumptions made a very good approximation of in one test case compared with the experience contained in Table 3 was achieved. In the work a main focus is put on the difference between different infrastructure measures. In VIA (Via, 2008) the same concept was also applied, since the main interest of this paper was to compare the outcome relative to different infrastructure, not the absolute value.

4.4 Model development, second approach

In contrast to the first approach, in the second approach makes more use of the capabilities of ARENA. This software tool provides a feature named guided transporter in the Advanced Transfer Project Bar. The movement of such transporters uses a set of links and intersections which define the possible paths a transporter can take in a network. The set of links and intersections define a system map. If this system map can be considered as a representation of the actual railway network. The guided transporters with an entity on board as load would then represent the trains. The intersections could then be viewed as tracks and lines, whereas links can represent sidings. The actual representation of such model approach is shown in figure 11 and figure 12.

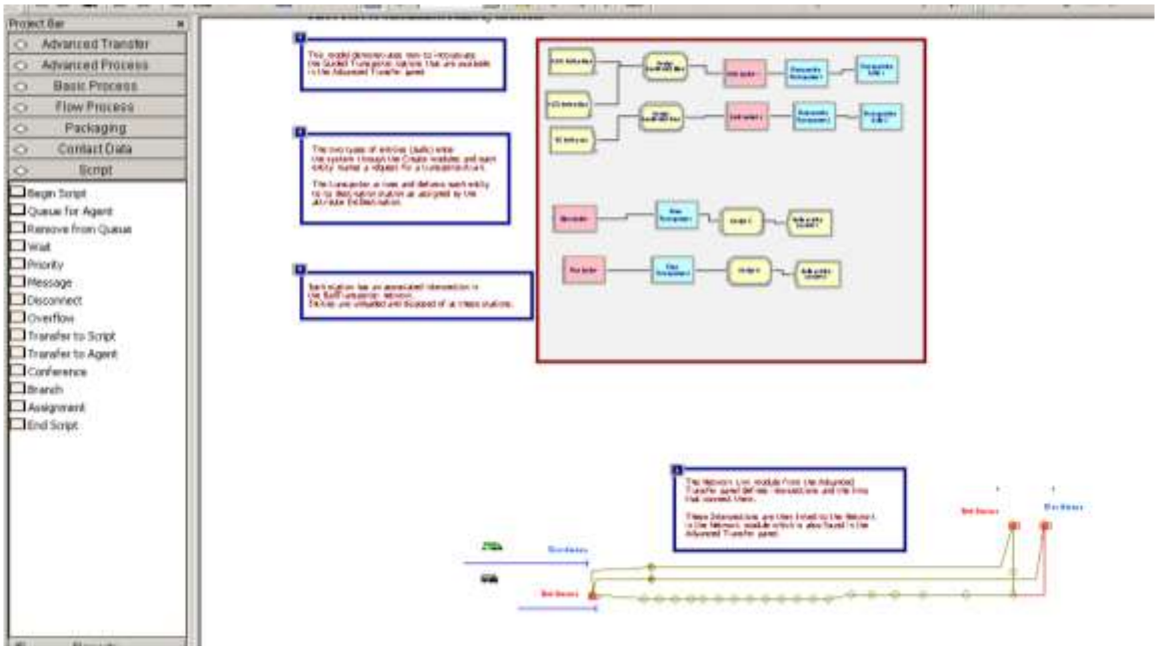


Figure 11: Model approach using a guided transporter network

Network Link - Advanced Transfer

	Name	Type	Beginning Intersection Name	Ending Intersection Name	Beginning Direction	Ending Direction	Number Of Zones	Zone Length	Velocity Change Factor
1	Network Link 1	Unidirectional	Intersection 1	Intersection 2	0		5	1000	1.0
2	Network Link 2	Unidirectional	Intersection 2	Intersection 3	0		1	5000	1.0
3	Network Link 3	Unidirectional	Intersection 3	Intersection 4	0		1	5000	1.0
4	Network Link 4	Unidirectional	Intersection 4	Intersection 5	0		1	5000	1.0
5	Network Link 5	Unidirectional	Intersection 5	Intersection 6	0		1	5000	1.0
6	Network Link 6	Unidirectional	Intersection 6	Intersection 7	0		1	5000	1.0
7	Network Link 7	Unidirectional	Intersection 7	Intersection 8	0		1	5000	1.0
8	Network Link 8	Unidirectional	Intersection 8	Intersection 9	0		1	5000	1.0
9	Network Link 9	Unidirectional	Intersection 9	Intersection 10	0		1	5000	1.0
10	Network Link 10	Unidirectional	Intersection 10	Intersection 11	0		1	5000	1.0
11	Network Link 11	Unidirectional	Intersection 11	Intersection 12	0		1	5000	1.0
12	Network Link 12	Unidirectional	Intersection 12	Intersection 13	0		1	5000	1.0
13	Network Link 13	Unidirectional	Intersection 13	Intersection 14	0		1	5000	1.0
14	Network Link 14	Unidirectional	Intersection 14	Intersection 15	0		1	5000	1.0
15	Network Link 15	Unidirectional	Intersection 15	Intersection 16	0		1	5000	1.0
16	Network Link 16	Unidirectional	Intersection 16	Intersection 17	0		1	5000	1.0
17	Network Link 17	Unidirectional	Intersection 17	Intersection 18	0		1	5000	1.0
18	Network Link 18	Unidirectional	Intersection 18	Intersection 19	0		1	5000	1.0
19	Network Link 19	Unidirectional	Intersection 19	Intersection 20	0		1	5000	1.0
20	Network Link 20	Unidirectional	Intersection 20	Intersection 21	0		1	400	1.0
21	Network Link 21	Unidirectional	Intersection 20	Intersection 22	0		5	400	1.0
22	Network Link 22	Unidirectional	Intersection 22	Intersection 23	0		1	500	1.0
23	Network Link 23	Unidirectional	Intersection 23	Intersection 22	0		20	500	100
24	Network Link 24	Unidirectional	Intersection 21	Intersection 24	0		20	500	100
25	Network Link 25	Unidirectional	Intersection 24	Intersection 1	0		5	500	100
26	Network Link 26	Unidirectional	Intersection 25	Intersection 1	0		5	500	100

Double-click here to add a new row

Figure 12: Network list of the model shown in Figure 11

But in contrast to the model 1, where trains are generated as separate entities using a create element and are disposed after the travel through the simulation network, all guided transporters already exist at the very first instance of the simulation. Therefore the guided transporters have to be allocated, moved and requested. Also the storage of unused guided transporters have to be managed. This behavior is much more like that of trains in reality, this also allows the simulation model to be used for development and verification of dispatching schemes. But for the case of simulating a railway network for calculation of capacities, this implies a lot of extra work to be able to use this ARENA feature. For example the process of storage of unused transporters has to be solved. In the present work it was solved by using a shadow station, which is situated close to the main station, where all trains routes star. In this shadow station, there is a different track for all different guided transporter types, one for each different train type used in the simulation. This is necessary, as no type of transporter should be blocked on its track in the shadow station by another type of transporter, which is actually unused. For representing the stations, an element called station is provided in ARENA. In these ARENA stations normally the entity transported would undergo some kind of manipulation. In our case the entity represents the railway transport, necessary according the timetable. Therefor a separate type of entities is needed for each type of train, high speed trains, local trains and goods trains. The maximum allowed speed, the acceleration rate and the deceleration rate of a guided transporter can be defined in ARENA by adding these values to the properties of the transporter in the appropriate fields. This allocation of properties to the guided transporters is also the reason, why there have to be two different elements for each train: an entity representing the train from the timetable and the guided transporter necessary for the movement between stations on the tracks. In addition there is another inbuilt feature of transporters: in ARENA they have the property of length. With this property, they are able to represent the actual length of the train, but they also can prevent other trains from running through the occupied intersections. Additionally, the feature of length allows to use a blocking scheme to represent the blocking procedures typical for railway networks. Further information about blocking is provided in chapter 2.2.1 of this paper.

In railway networks it is typical, that some types of trains stop at every station, others stop at some and some train types like the high speed trains stop only at major stations. This concept is also possible in ARENA, as there is the possibility of sequences. Such a sequence can be defined in ARENA for all different entity types, in this case all kinds of train types. Additionally, there are different priorities assigned to different train types: for example high speed trains have the highest priority, whereas local trains and goods trains have lower priority, resulting in additional waiting time until approaching trains with higher priority have passed at certain overtaking places, like stations or sidings with additional parallel tracks.

For representing a scheduled railway movement, the following scheme has to be followed:

1. Create an entity according to the train type necessary (train types are high speed trains, Eurocity trains, Intra-Regional cargo trains etc.).
2. Move this entity at the main station.
3. Request a matching empty guided transporter and move it, if available to the main station.
4. Load the transporter with the entity and move it along its schedule along the network, with respect to the priority assigned to this entity. This may result in additional waiting times.
5. Halt the entity at the designated stations for the predetermined waiting times.

6. Dispose the entity after reaching its final destination, after having recorded all times associated with its travel.
7. Move the empty guided transporter to the corresponding waiting area, ready for the next request.

Further information about the features of guided transporters can be found in (Pegden 1995) in Chapter nine Advanced Manufacturing Features.

To simulate the process of overtaking correct, it is necessary to move all trains with lower priority always through the overtaking sidings, even when no train is ready for overtake. Only within these parallel tracks it is possible due to the underlying SIMAN code to look for the positions of all entities with higher priority. If there is one approaching in the track behind the last siding, the train with the lower priority will have an additional waiting time. It has to be noted, that for scheduled stops the waiting time is enlarged, whereas for unscheduled stops, like in the case of goods trains, a waiting time other than zero has to be inserted.

For developing the model, the Guided Transporter Model File of the Smart files Library provided with ARENA was used. Within this file also an example for visualization of transporter movement was included. This example was also very useful when developing the visualization for validation and verification.

In the following, some of the pictures for transporters developed during this work used in the conventional main line in scenario 2 are displayed.

Figure 13: Drawing representing the high speed train

As most of the high speed trains actually in use worldwide are using electric traction force, the picture shows streamlined traction vehicle with a pantograph.

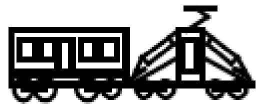


Figure 14: Drawing representing the Eurotrain

As the Eurotrain represents a train type slower than the high speed train type, but also representing the top priority trains, its picture shows a streamlined electric locomotive followed by a passenger wagon.

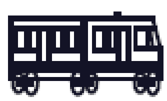


Figure 15: Drawing representing the Regional train

The Regional train is the slowest passenger train, making a halt at every station. In Europe, in most countries it is a diesel hauled passenger train. As its maximal speed is only at 100 km/h, the design is oriented at efficiency of space and weight, not towards minimization of aerodynamic resistance. The fast regional train, which is not displayed here, is similar to the Regional train, but has a higher maximum speed and does stop a only some of the stations.

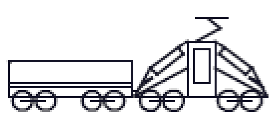


Figure 16: Drawing representing the Intra-Regional Cargo Train

The Intra-Regional Cargo Train is composed of an electric hauled locomotive followed by a freight wagon. Its maximum speed is 100 km/h, so it is somewhat faster than the regional Cargo train (here not displayed, similar to the Intra-Regional Cargo Train but with a diesel hauled locomotive). All cargo trains have lowest priority have lowest priority during the operation. So these trains will allow passage for faster trains if necessary.

During the simulation run, the graphical representations of the different transporter types move along the tracks, using sidings and stations. In addition to the pictures shown above the busy transporter has a ball on board, which represent the entity. The color of the ball is chosen to be an additional sign of the different type of train.

4.4.1 Variation is introduced in the second model

As described in the previous chapters, the quality of railway service for the users depend very much on the punctuality of trains. As demonstrated before, there are external causes, which lead to variation in train handling. The capability of a railway network infrastructure together with the associated timetable to cope with such input is of major interest in research. This variation can be inserted in this second model at certain positions. It is quite easy to represent the delay of a train arriving at the networks border. This can easily be done by introducing an additional random delay in the creation process. Also, it is simple to insert a random effect at the waiting times within stations by adding additional delay elements from the advanced process bar. But for the movement between stations such an inclusion of random effects is not easy. But there is a work around this obstacle: all additional variation in travel time will be inserted additionally at the stations.

It is worth to note, that the concept of adding variability depends very much on the availability of real (measured) data to adjust the models. Unfortunately only very small amount of data is accessible in literature. In the last chapter, this problem will be restated.

4.4.2 Verification and validation

As described in paragraph 2.1.3 the verification is an important task in a simulation study.

In this study, there was always a first step for verification applied. This test is simple the check, if a single train running at maximum speed through the simulated network will reach the correct upper limit of speed and if the time the train needs to travel from entrance station to end station is simulated correctly with respect to the calculated value. This should also include to view at the correct routing, where the routing is of influence like in the scenario 2.

It has been tested, that all maximum allowed speeds are reached and also the travel time are within small tolerance close to calculated travel times.

In a second step the behavior of different trains on the simulation network is tested. For example in the scenario 2 (conventional main line) eurocity trains and goods trains are send on the track. Then it has to be checked, if the goods train gives way to the trains with higher priority.

The following four pictures illustrate the process of overtaking in scenario 2.

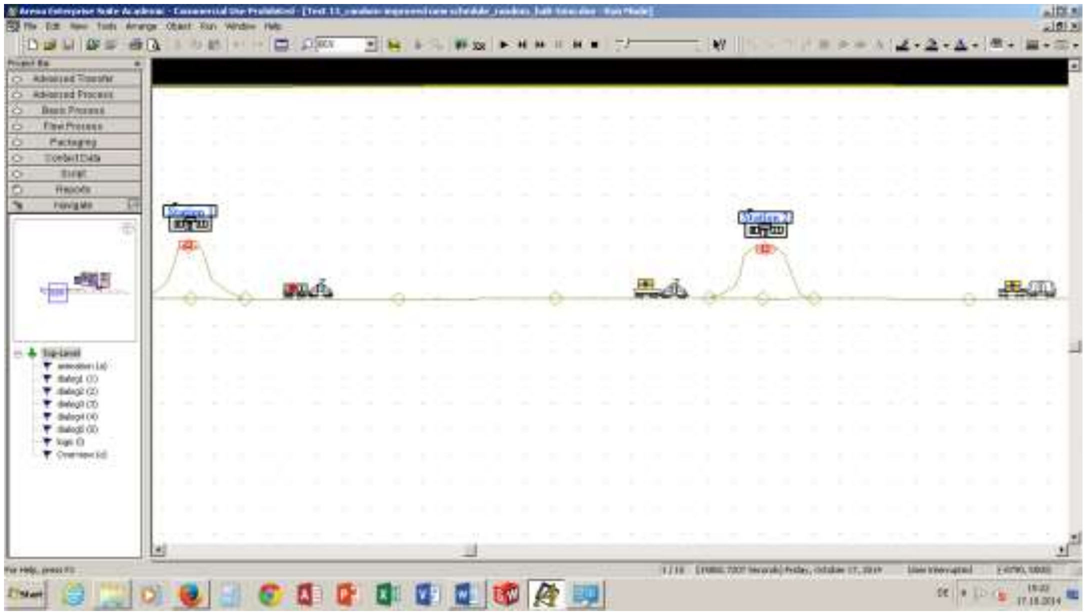


Figure 17: Eurocity train is following after the goods train,

which is approaching station 2 in scenario 2

Figure 18: The goods train is waiting in Station 2,

while the Eurocity is approaching in scenario 2

Figure 19: Eurocity train is overtaking the goods train at station 2,

which is still waiting on the parallel track in scenario 2

Figure 20: Eurocity train is gaining speed after leaving station 2,

while the goods train starts again going back to the main track in scenario 2

Also it has to be checked by viewing at the visualization if the correct distance between trains is kept. An example is shown in the following three pictures also taken from the scenario 2.

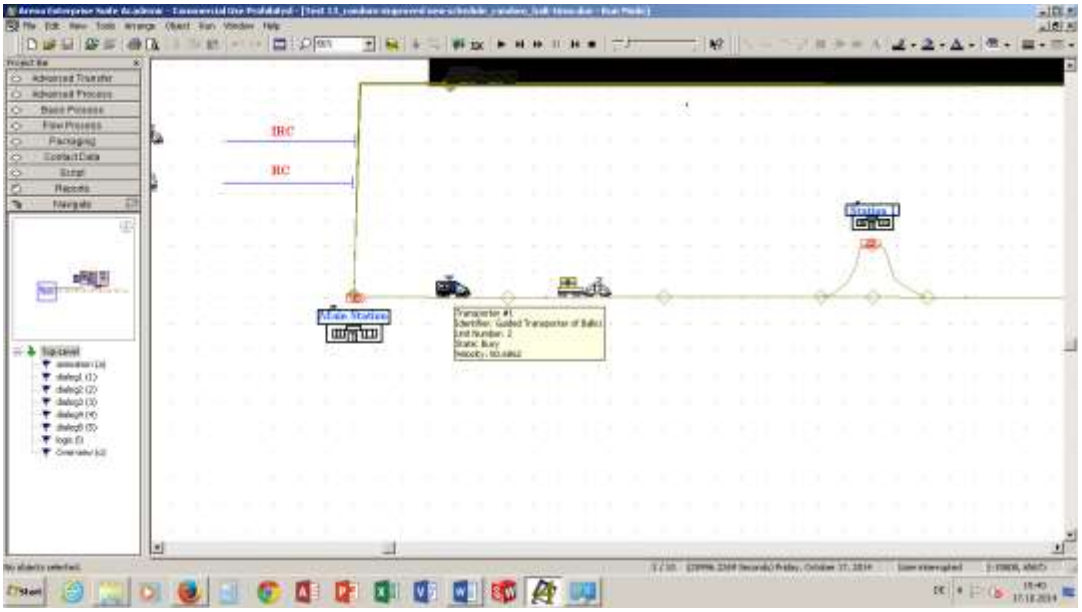


Figure 21: Goods train is gaining speed after leaving the main station,

while a high speed train has just started its journey at the main station in scenario 2,

having an actual speed of approx. 50.7 m/s

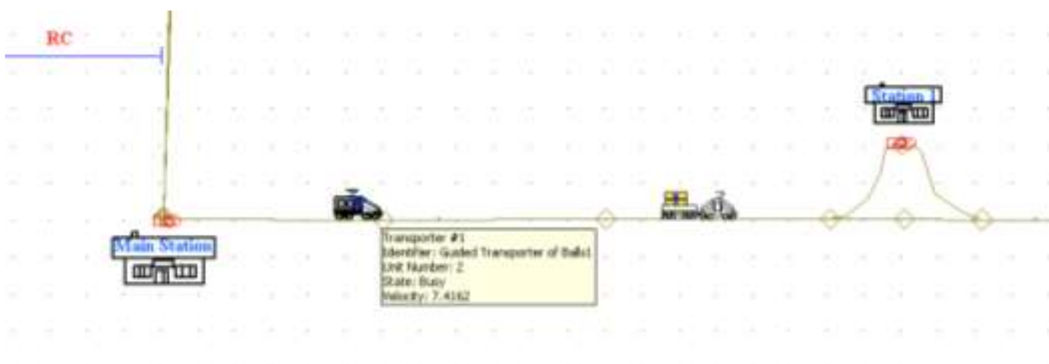


Figure 22: Goods train is approaching station 2,

while a high speed train has slowed down to hold the block distance in scenario 2,

having now an actual speed of approx. 7 m/s

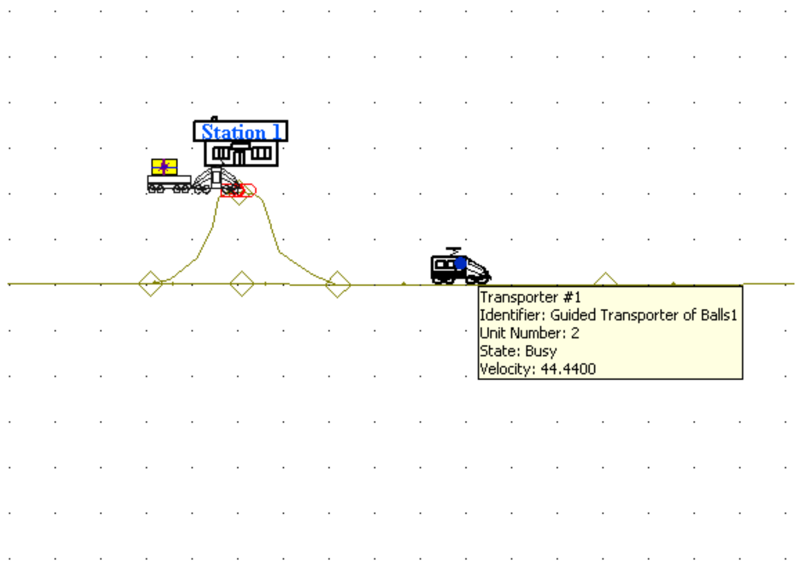


Figure 23: Goods train is waiting in station 2,

while the high speed train is speeding up in scenario 2

having now an actual speed of approx. 44 m/s

4.5 Results of the second model approach for two of the scenarios

In the following chapters the results for two different scenarios are displayed, the High speed line (scenario 1) as well a conventional main line (scenario 2). Also the basic idea of discovering statistically significant differences between these results.

4.6 Results of the second model approach for scenario1: High speed line

The scenario 1, the high speed line, is described in more detail in 3.3. Here only a short description is given: only two different train types exist in this railway network, only a high speed train and a eurocity train. Both train types do not stop in a scheduled way when running on the track, which has a length of 100km.Additionally, there is no possibility of overtaking. There is a difference in maximum speeds and possible accelerations between both train types. As there are two times more high speed trains than eurocity trains running per day, a very simple timetable can be used: The timetable is devided into time slots. In each time slot tree trains are entering the system. The first train in such scheme is always a high speed train followed by a eurocity train and then a high speed train again. As there is no overtaking possible in the network, this scheme of entering the system is also the scheme of leaving it. After some time after the second high speed train has entered the system, a third high speed train is entering, which can be assumed as the first train of the following time slot. In some literature (Schmidt 2009; Chu 2014) these time slots are called time slices and are a major characteristic of the simulation system used. It is obvious that the trains of the first time slot are not influenced by those entering the system in the second slot. If the time difference between the last train of first slot 1 and the first train of the second slot is big enough, this is also true for the trains of slot two with respect to those of the first time slot. If we assume an undisturbed train operation (the ideal case) this time difference must be at last the time the high speed train needs for travelling along the blocking distance.

4.6.1 No variation in the model

4.6.1.1 No variation in the model, ETCS level 1

ETCS level 1 has been simulated by using the standard blocking rule of release at the end. Then the guided transporter frees the block resp. the intersection at the point when it leaves it. The timetable is introducing a high speed train first, then followed by another high speed train and then finally comes a Eurocity train. Then the next cycle begins with a high speed train, followed by another high speed train and then a Eurocity train. The point of optimal utilization is reached, when the number of trains put through the network is maximized while the waiting times of the trains is still zero. For this calculation the process analyzer of ARENA has been proven as a very useful tool. The calculation shows that during the simulation run of 24 h, after a warm up period of 14400 seconds, at a maximum 159 trains will pass through a time of 24 hours. Note: this is the value for the steady state without variation resulting in a theoretical capacity, therefore it has to be reduced by the factor of 0.6 to get to the practical capacity. The practical capacity is then 95.4 trains per day. Compared to literature (Via 2008) which gives a value of 129 resp. 127 (two different calculation methods) this is quite low. This may be due to the fact that in the simulation system, the guided transporter need to return to the starting point, whereas in the capacity calculation from literature ideal conditions are used. In the simulation it was assumed, that the distance between the endpoint and the starting point is around 12.5 km (unwanted travel), whereas the travel distance from starting point to end point (the wanted travel) is 100 km. From that it can be estimated, that simulation results will be around 12.5 % too low. However this correction factor is still not enough.

4.6.1.2 No variation in the model, ETCS level 2

ETCS level 2 is quite easy to calculate, there is no new simulation model needed. The zone control rule of the transporters in the model for ETCS level 1 has just to be changed from END to START. Then all transporters will follow the proceeding transporter with a minimal distance (according to their actual speed and that of their predecessor). This is the behavior we would expect for the case if ETCS 2 is introduced.

In this case the result for the theoretical capacity is 162, which is a very small increase of only 3 trains per day. In the literature (VIA 2008) however, an increase of 18 respective 19 (depending on the calculation method) was obtained. So perhaps the expectation, that the second model is capable of representing the different signaling procedures connected with ETCS level 2 is not justified. Therefor no additional evaluation was performed. No variation in the model, ETCS level 3

Unfortunately the author actually sees no possibility to simulate the performance of the railway network with ETCS level 3.

4.6.2 Variation in the model

4.6.2.1 Variation in the model, ETCS level 1

In a first attempt, the same data has been used as in subcase 1 (no random), but with additional decide module (10%) and delay module MIN(6,LOGN(4, 0.75)) in minutes added for all train entries. The results of this test 1 are the following:

For ten replications, each with a running length of 24 hours with additional warm up phase the number out in is 155 in average, so lower than before (159). However we see an increase of waiting times (from 0.0) and for travel times.

So it is obvious, that the additional variance results in an additional delay of trains equal to longer travel times, however some trains just have the same travel time as before.

Now the timetable has to be changed, until the simulated system can cope with the variance introduced.

For this tasks, the tool process analyzer is used, which allows for easy calculation with variation of input values. An example of the procedure is shown in the following picture 24.

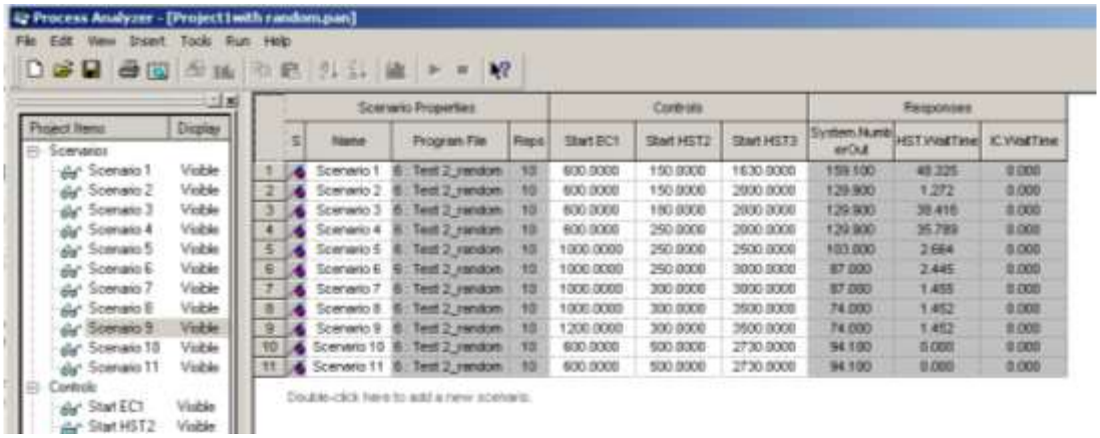


Figure 24: Finding the near optimal timetable for trains using ARENA PAN

With PAN, the process analyzer of ARENA, we come to the conclusion that acceptable performance could be obtained with a throughput of 94 trains per day. This is not exactly the expected Factor of 0.6 (from VIA 2008), a rule of thumb. But it is in fact a value of 0,5911… . For a value obtained via rule of thumb this is quite a good approximation. This is therefore an indirect test showing that the assumption of the distribution of variation as lognormal is close to reality.

4.6.2.2 Variation in the model, ETCS level 2

The results of the simulation without variation give unfortunately evidence that the simulation of ETCS level 2 is not correct. Therefore no additional calculations with random effects were performed.

4.6.3 Comparison of results from 4.5.1.1 and 4.5.2.1

For a full comparison of different infrastructure simulations hypothesis testing is used, since both results (that for ETCS level 1 and in comparison to that results for ETCS level 2) contain variance. In this specific case this can be done using a hypothesis test on the difference of means with variance known. Such search for statistically significance is explained for example in Montgomery (2011) in chapter 10 as two-sample t-test. To get reliable results the number of replications has to be chosen large enough, so that the assumption of distribution of test results close to normality could be assumed. In Kleijnen (2008) it is shown, that in many cases of simulations the normality can be assumed, especially when large sample sizes are evaluated (page 78 and page 79). Otherwise a Wilcoxon Rank-sum test can be performed instead to demonstrate that the assumption of statistically significant difference between the different infrastructures really does exist.

4.7 Results of the second model approach for scenario 2: Conventional main line

The network for the conventional main line is described in 3.3 in detail as scenario 2. Here comes only a very short description. Ten different stations are situated along a 100 km track. Trains of 6 different train types are running on the linear railway network, originating from a main station and running to an end station. There are these two stations where all train stop, on two additional stations only all regional trains stop, whereas on seven more stations only slow regional trains stop. In all stations overtaking is possible. The maximal allowed speed is 160 km/h.

4.7.1 No variation in the model, ETCS level 1

ETCS level 1 has been simulated by using the standard blocking rule of release at the end. Then the guided transporter frees the block resp. the intersection at the point when it leaves it. As it was obvious in the model for the scenario 1 that the process of empty transporters returning to the entry point needs improvement. That’s why in the model for the conventional main line there were two improvements implemented. First the process of returning the emptied transporters was improved by dramatically shortening the virtual distance between exit and entry station. Additionally the process of recording the travel time was improved by adding additional record modules from the basic process panel. This also improves the possibilities to get more information about the output values.

A main problem for a simulation is the development of a feasible timetable. Feasible means here, that the number of trains leaving the simulation network meets the requirements of the given infrastructure utilization: fifty long-distance passenger trains, forty short-distance passenger trains and additional 60 goods trains.

A first idea of a timetable was found from trial and error. However this timetable does not fit the criterion of requested trains. As there are too many different possibilities to change the existing timetable (in this example 6 controls), it is not an easy task to find a better one. Therefore the support program OPTQUEST engine 6.4.1.1 was used to determine better timetables, since it allows to do such search in an automated manner. It has to be noted, as OPTQUEST is using a heuristic method to come up with better solution, so it is not guaranteed if the solution vector OPTQUEST is providing is really the best possible solution. However for such complex input vectors, OPTQUEST provides valuable assistance and has showed acceptable performance.

As input vector the different starting times of the different train types are used. The simulation is then run for a certain time plus extra warm-up period. The objective is to maximize the number of trains that leave the system within the given time. The constraints are the minimal travel times of the different train types. This type of constraints is very realistic since it allows demanding limits for travel times for high speed trains and slack limits for goods trains. As overtaking and stopping is a normal way of operation in this network, almost none of the trains is running at full speed all the time. This leads to the observation that almost none of the train´s travel times are reaching the maximum, but is travelling for a longer time through the simulation network. The definition of a upper limit in travel time for each train type separately as a constraint is realistic, as for example in real situation a traveler using a high speed train is expecting, that he or she reaches his final destination in quite short time interval, whereas the travel of a goods train is not as time critical and delay is allowed for this low priority train. This delay will occur when additional halts are introduced for overtaking and due to possible blocking.

At first it was tried to develop the timetable manually, just as that used in scenario 1. The numbers of trains put through the network is 75. The input vector for this timetable is used as starting point. An example of such optimization is shown in the following three figures. In Figure 25 you see the output of OPTQUEST at step 15. In steps two, three and seven opt quest found input vectors with improved performance, which also fulfill the constraints.

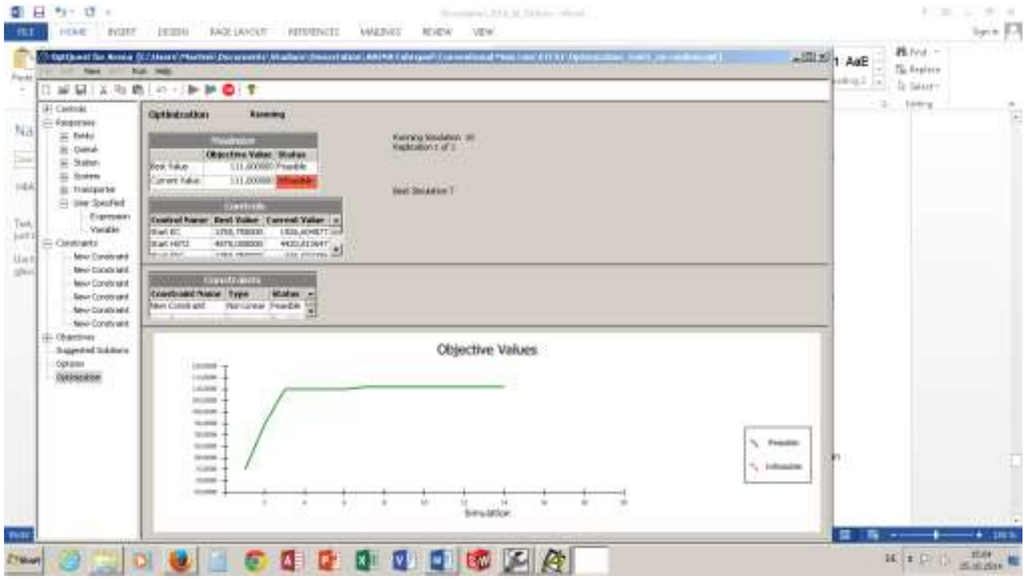


Figure 25: Finding a timetable with larger output than 75, part 1 step 15

In Figure 26 we see the output of OPTQUEST at step 351. No better input vector (which also fulfills all constraints) than that found in step 7 was found.

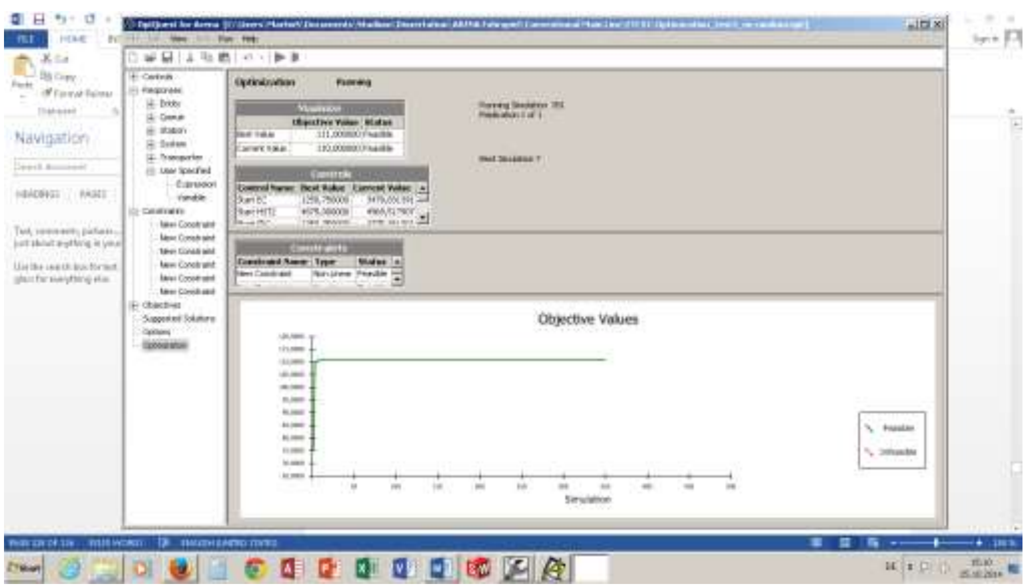


Figure 26: Finding a timetable with larger output than 75, part 2 step 351

The result of OPTQUEST are displayed in Figure 27. It is notable that a lot of input vectors were found that have the same objective value of 111.

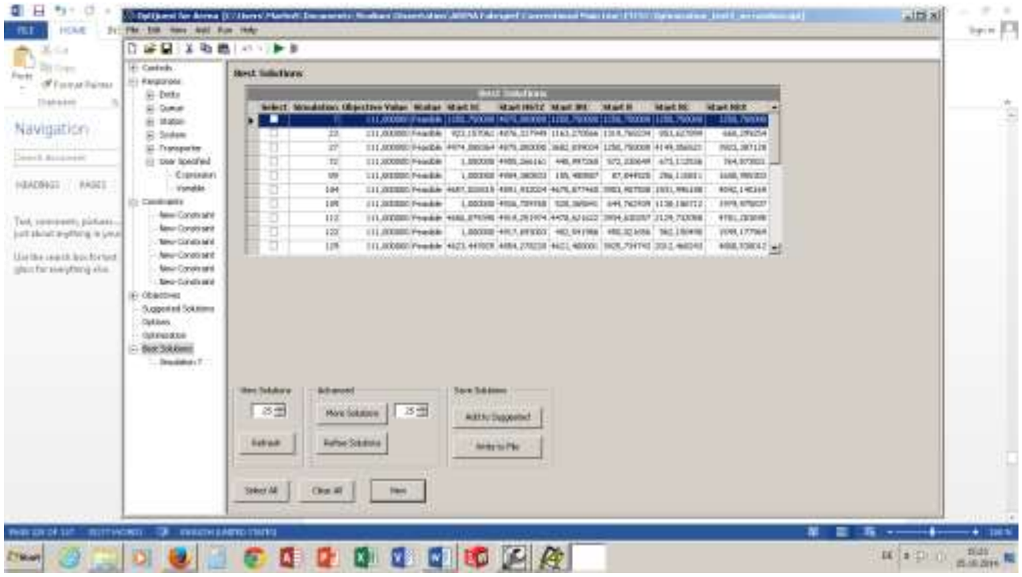


Figure 27: Results of OPTQUEST run after termination in step 500

This input vector is then used as improved timetable. The point of optimal utilization is reached, when the number of trains put through the network is maximized while the waiting times of the trains is still within limits. Further information about using OptQuest may be found in Appendix C.

The calculation which is made with the improved timetable as basis shows that during the simulation run of 24 h, after a warm up period of 14400 seconds, at a maximum 224 trains will pass through a time of 24 hours without random input. Note: this is the value for the steady state without variation resulting in a theoretical capacity, therefore it has to be reduced by the factor of 0.6 to get to the practical capacity. The practical capacity is then 134.4 trains per day. Compared to literature (Via 2008) which give values of 147.6 resp. 112.6 (two different calculation methods) this is in the same range. So this calculation shows, that the simulation (without random effects) is also an option to calculate capacities of railway networks. Depending on the railway network, values may differ from other calculation methods. Further work may provide further inside, why for example in scenario 1 simulation leads to low capacity values.

4.7.2 No variation in the model, ETCS level 2

As for scenario 1 – the high seep line – no acceptable prediction with the second modelling approach (using guided transporter features) was achieved, this was not tested with scenario 2, the conventional main line.

4.7.3 Variation in the model, ETCS level 1

In a first attempt, the same data has been used as in subcase 1 (no random), but with additional decide module (30%) and delay module MIN(6,LOGN(4, 1)) in minutes added for all train entries. Note that this value is different from the high speed line scenario, since it is very likely that probability of train delay increases in real traffic with inhomogeneity of traffic. The results of this test 1 are the following:

For ten replications, each with a running length of 24 hours with additional warm up phase the number out in is 154 in average. At first glance this is a little bit surprising, since we would expect significantly lower value of output (in the range of 134). But compared to differences between calculation methods in literature (Via 2008) which gave values of 147.6 resp. 112.6 (two different calculation methods) the result is quite close. It has to be noted, that the variance in input values has been taken from one publication based on empirical values whereas the scenario is not a real network, but only a scenario. With this fact in mind the results demonstrate that there is nevertheless some correlation.

4.7.4 Sensitivity Analysis

The test scenario for the ETCS level 1 with variation is used to perform several sensitivity analyses. As stated in Kleijnen (2008, p. 64) a sensitivity analysis can be performed during a simulation study to better understand the simulated system. By taking into account not only the main factors but also the interactions between factors, the combination of different factors could also be discovered. However in the actual case of the simulation of a railway network, we perform such sensitivity analysis only to discover, if changes in one factor at a time is of influence on the performance of the simulated system

A) Sensitivity analysis of the tail ending time in the variation delay module.

To ease calculations, the process analyzer of ARENA was used.

A1) min (6; logn(4,1)) as the standard value

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2544.85	10,23	2521.15	2564.49
Record HST travel time	2561.14	13,95	2538.82	2594.41
Record IRC travel time	4593.98	24,90	4544.44	4650.89
Record R travel time	4783.74	26,80	4735.24	4861.81
Record RC travel time	4876.85	33,84	4773.83	4936.69
Record REX travel time	2913.68	10,53	2884.56	2932.73

Figure 28: Results for travel times after 10 replications with standard value

A2) min (5; logn(4,1))

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2541.39	10,49	2512.24	2567.32
Record HST travel time	2560.38	10,20	2539.13	2580.64
Record IRC travel time	4598.92	30,42	4542.46	4651.79
Record R travel time	4779.48	22,70	4729.38	4827.17
Record RC travel time	4876.68	36,65	4764.13	4936.69
Record REX travel time	2910.05	11,60	2889.35	2945.10

Figure 29: Results for travel times after 10 replications with min (5; logn(4,1))

Note that for some train types, the average is decreasing (like EC and HST) as expected, but for some (IRC) they are increasing instead.

A3) min (5.5; logn(4,1))

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2548.40	8,40	2528.02	2563.96
Record HST travel time	2559.77	13,04	2540.64	2594.20
Record IRC travel time	4593.13	24,02	4533.05	4651.07
Record R travel time	4783.68	24,99	4734.12	4853.99
Record RC travel time	4876.07	34,30	4773.66	4936.69
Record REX travel time	2909.89	9,96	2885.04	2929.34

Figure 30: Results for travel times after 10 replications with min (5.5; logn(4,1))

Note, that in this case, for some types of trains the half width is changing notable compared to the other calculation results.

A4) min (6.5; logn(4,1))

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2546.63	12,71	2517.82	2577.93
Record HST travel time	2560.77	14,35	2536.53	2594.57
Record IRC travel time	4591.41	26,68	4544.44	4650.89
Record R travel time	4783.53	26,85	4735.24	4862.00
Record RC travel time	4876.32	34,23	4772.40	4936.69
Record REX travel time	2910.68	12,16	2880.25	2933.56

Figure 31: Results for travel times after 10 replications with min (6.5; logn(4,1))

A5) min (7; logn(4,1))

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2545.89	12,08	2516.91	2572.35
Record HST travel time	2560.74	14,33	2536.53	2594.57
Record IRC travel time	4592.92	26,73	4544.44	4650.89
Record R travel time	4783.58	26,85	4735.24	4862.00
Record RC travel time	4876.49	34,39	4770.78	4936.69
Record REX travel time	2912.64	11,50	2885.04	2934.92

Figure 32: Results for travel times after 10 replications with min (7; logn(4,1))

B) Sensitivity analysis of the timetable: change of cycle time

The cycle time (the time between the starts of different, adjacent time slots has been varied using process analyzer from ARENA.

It is expected, that the travel times will change, especially that travel times will be reduced as cycle time is elongated.

	Scenario Properties				Control	Responses						
	S	Name	Program File	Reps	Cycle 1	System NumberOut	Record EC travel time	Record HST travel time	Record IRC travel time	Record R travel time	Record RC travel time	Record REX travel time
1		Scenario 1	1 : Sensitivity	10	6000.0000	217.5000	2716.0729	2534.0342	4863.4665	5141.9232	5162.1909	3297.7209
2		Scenario 2	1 : Sensitivity	10	7000.0000	186.6000	2820.8440	2587.8478	4866.8502	4962.5481	4965.0376	3062.0032
3		Scenario 3	1 : Sensitivity	10	7500.0000	173.0000	2586.3239	2607.2256	4629.2878	4910.2289	4858.7917	2965.5123
4		Scenario 4	1 : Sensitivity	10	7750.0000	188.1000	2569.2307	2572.8489	4819.6616	4845.7950	4917.8941	2942.5087
5		Scenario 5	1 : Sensitivity	10	8000.0000	161.9000	2544.9505	2561.1439	4593.9756	4783.7425	4876.8498	2913.6811
6		Scenario 6	1 : Sensitivity	10	8250.0000	157.7000	2518.0641	2534.0701	4557.7702	4728.1568	4820.5101	2913.0049
7		Scenario 7	1 : Sensitivity	10	8500.0000	153.9000	2487.5896	2480.6283	4532.4047	4718.7729	4722.9576	2909.7459
8		Scenario 8	1 : Sensitivity	10	9000.0000	144.2000	2481.2842	2418.8443	4515.5574	4713.8652	4836.1356	2909.2342
9		Scenario 9	1 : Sensitivity	10	10000.0000	130.0000	2481.1949	2411.7851	4514.3109	4715.0216	4630.3328	2910.9664
10		Scenario 10	1 : Sensitivity	10	12000.0000	109.0000	2481.9033	2414.0442	4506.8226	4709.3787	4627.2016	2908.6181

Double-click here to add a new scenario.

Figure 33: Results for travel times with changing cycle times (from PAN)

B1) Cycle = 6000

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2716.07	19,99	2687.19	2771.02
Record HST travel time	2534.03	24,20	2488.49	2587.38
Record IRC travel time	4863.47	9,28	4844.19	4887.88
Record R travel time	5141.92	25,37	5095.67	5204.86
Record RC travel time	5162.18	20,32	5100.67	5195.67
Record REX travel time	3297.72	32,68	3212.33	3351.21

Figure 34: Results for travel times after 10 replications with cycle = 6000

B2) Cycle = 7000

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2820.84	21,67	2765.93	2866.06
Record HST travel time	2587.85	19,18	2551.64	2640.52
Record IRC travel time	4668.85	16,77	4624.35	4698.53
Record R travel time	4962.55	18,25	4930.50	5011.40
Record RC travel time	4965.04	30,86	4863.70	5012.84
Record REX travel time	3062.00	18,34	3034.29	3104.22

Figure 35: Results for travel times after 10 replications with cycle = 7000

B3) Cycle = 7500

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2586.32	16,00	2557.27	2616.97
Record HST travel time	2607.23	20,21	2568.56	2673.89
Record IRC travel time	4629.29	18,55	4580.80	4662.07
Record R travel time	4910.23	27,67	4848.22	4973.29
Record RC travel time	4958.79	32,75	4865.45	5018.32
Record REX travel time	2955.51	12,15	2928.26	2981.99

Figure 36: Results for travel times after 10 replications with cycle = 7500

B4) Cycle = 7750

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2569.23	10,99	2537.04	2587.39
Record HST travel time	2572.65	17,98	2531.03	2620.36
Record IRC travel time	4619.66	25,19	4563.97	4658.43
Record R travel time	4845.79	26,82	4787.81	4904.65
Record RC travel time	4917.89	29,60	4838.86	4976.94
Record REX travel time	2942.51	7,15	2926.66	2955.07

Figure 37: Results for travel times after 10 replications with cycle = 7750

B5) Cycle = 8000

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2544.85	10,23	2521.15	2564.49
Record HST travel time	2561.14	13,95	2538.82	2594.41
Record IRC travel time	4593.98	24,90	4544.44	4650.89
Record R travel time	4783.74	26,80	4735.24	4861.81
Record RC travel time	4876.85	33,84	4773.83	4936.69
Record REX travel time	2913.68	10,53	2884.56	2932.73

Figure 38: Results for travel times after 10 replications with cycle = 8000

B6) Cycle = 8250

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2518.06	9,56	2499.40	2536.87
Record HST travel time	2534.07	15,89	2512.96	2587.44
Record IRC travel time	4557.77	22,61	4492.29	4595.28
Record R travel time	4728.16	18,68	4704.74	4776.00
Record RC travel time	4820.51	44,40	4699.43	4907.16
Record REX travel time	2913.00	9,76	2885.73	2933.80

Figure 39: Results for travel times after 10 replications with cycle = 8250

B7) Cycle = 8500

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2487.57	7,32	2474.00	2504.14
Record HST travel time	2480.63	12,53	2453.62	2514.91
Record IRC travel time	4532.40	18,92	4479.98	4565.72
Record R travel time	4718.77	13,61	4701.58	4764.57
Record RC travel time	4722.96	47,70	4593.74	4802.96
Record REX travel time	2909.75	10,20	2882.56	2927.46

Figure 40: Results for travel times after 10 replications with cycle = 8500

B8) Cycle = 9000

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2461.26	10,25	2440.52	2483.89
Record HST travel time	2418.84	11,82	2398.96	2450.02
Record IRC travel time	4515.56	14,59	4478.69	4542.43
Record R travel time	4713.67	11,03	4690.37	4739.21
Record RC travel time	4636.14	22,70	4578.59	4671.70
Record REX travel time	2909.23	14,60	2883.75	2944.47

Figure 41: Results for travel times after 10 replications with cycle = 9000

B9) Cycle = 10000

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2461.19	10,13	2442.99	2487.11
Record HST travel time	2411.79	8,35	2396.20	2432.21
Record IRC travel time	4514.31	17,88	4473.20	4561.04
Record R travel time	4715.82	12,06	4689.66	4738.54
Record RC travel time	4638.33	23,09	4586.17	4676.50
Record REX travel time	2910.97	14,97	2875.73	2947.25

Figure 42: Results for travel times after 10 replications with cycle = 10000

B10) Cycle = 12000

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2461.90	14,45	2431.82	2498.40
Record HST travel time	2414.04	7,44	2399.78	2433.32
Record IRC travel time	4508.62	16,82	4470.18	4552.15
Record R travel time	4709.38	12,79	4687.52	4743.91
Record RC travel time	4627.20	25,77	4572.44	4680.06
Record REX travel time	2908.62	13,45	2877.82	2946.32

Figure 43: Results for travel times after 10 replications with cycle = 12000

C) Conclusions from the two sensitivity analyses

Based on the findings in A) and B) we realize, that small changes to the input values are of effect to the systems performance. However it is not a simple task of linear relation. From the tables it is not easy to discover a trend.

Therefore an additional chart was made, see Figure 44.

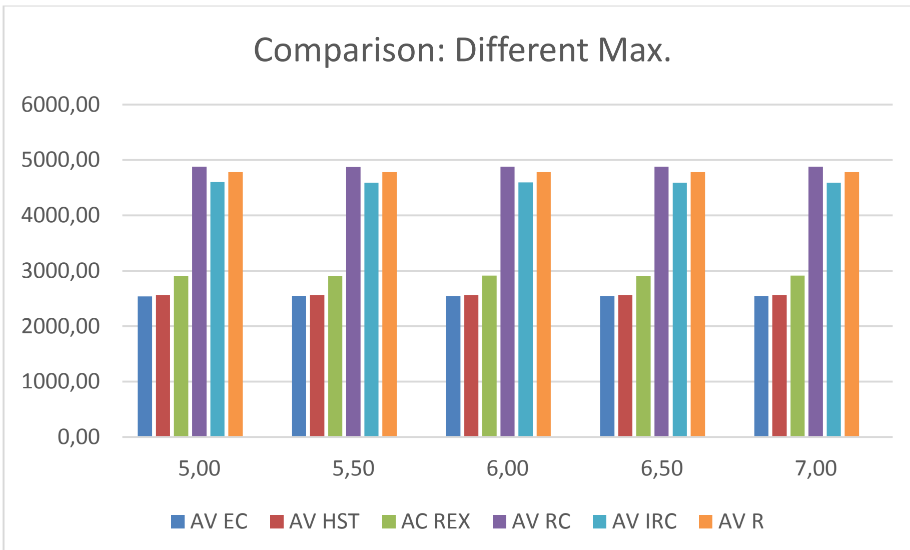


Figure 44: Results for travel times after 10 replications with different max-values from 5 to 7 min

There we see, that there is in fact no trend visible. So the factor chosen her has (at least in the area of values chosen) almost no effect.

The change of cycle times however has an enormous effect on travel-times. To illustrate the Figure 45 shows the effect.

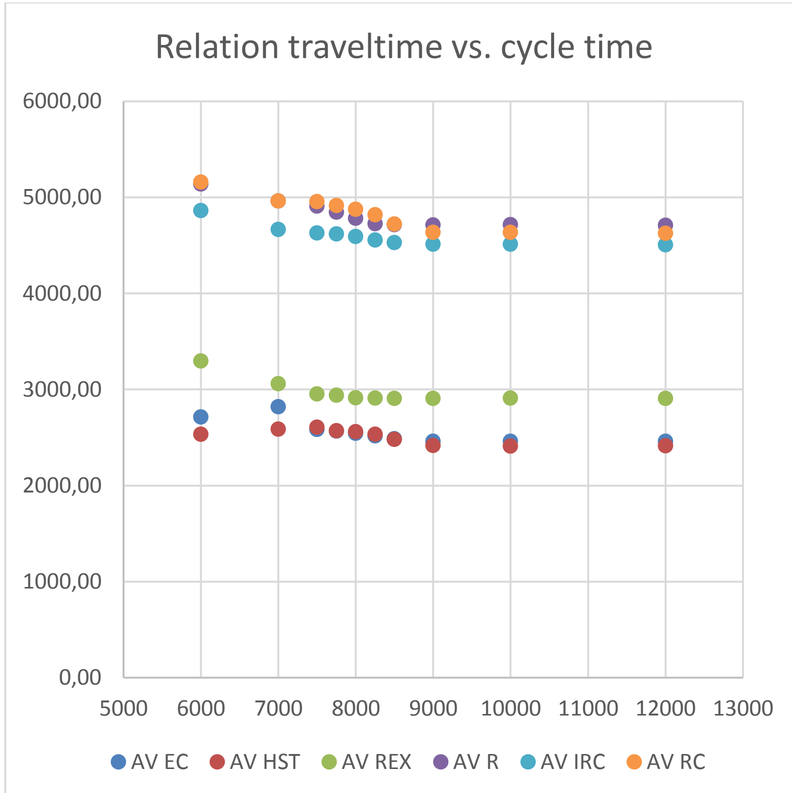


Figure 45: Results for average travel times in s after 10 replications with different cycle times from 6000 to 12000 s

Here we realize, that average travel times are decreasing with increasing travel times. After reaching the near optimal value of 8000 s for the cycle time, additional increase in travel time does add only small decrease.

4.7.5 Test for correctness of normality assumption

The test scenario for the ETCS level 1 with variation is used to test the normality assumption. As stated in Kleijnen (2008, p. 64) it may be justified to assume normality for large numbers of replications. To test, if – even with the number of replications is only ten - the properties of the distribution still remains the same, we simulate with different number of replications, changing none of the other properties.

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2544.85	10,23	2521.15	2564.49
Record HST travel time	2561.14	13,95	2538.82	2594.41
Record IRC travel time	4593.98	24,90	4544.44	4650.89
Record R travel time	4783.74	26,80	4735.24	4861.81
Record RC travel time	4876.85	33,84	4773.83	4936.69
Record REX travel time	2913.68	10,53	2884.56	2932.73

Figure 46: Results for travel times after 10 replications with cycle time of 8000 s

Expression	Average	Half Width	Minimum Average	Maximum Average
Record EC travel time	2547.09	4,30	2495.52	2616.82
Record HST travel time	2562.45	3,76	2519.83	2612.92
Record IRC travel time	4582.05	5,96	4505.66	4650.89
Record R travel time	4778.11	6,76	4696.40	4880.24
Record RC travel time	4882.95	6,13	4773.83	4946.31
Record REX travel time	2913.23	2,87	2884.56	2941.31

Figure 47: Results for travel times after 100 replications with cycle time of 8000 s

Here we see, that the half width is decreasing, whereas the averages for the extremes are going apart. So it is obvious, that the distribution of values changes with number of replications: it becomes much steeper and slimmer, whereas the tails are getting longer on both sides. This illustrated in Figure 48.

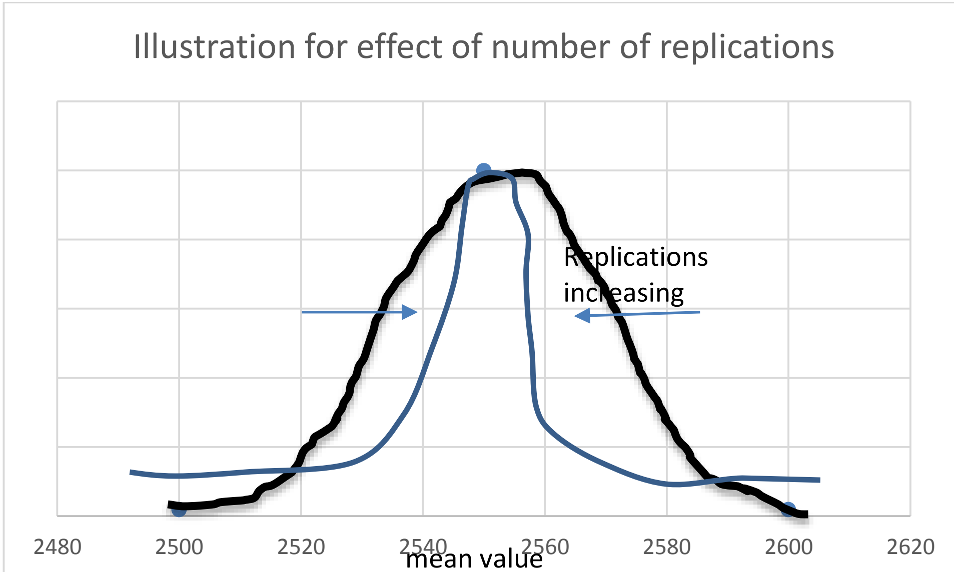


Figure 48: Effect of increasing number of replications: curve becomes steeper

Therefore it is not justified to assume normality for a number of just ten replications. Further calculations are needed to find out, if the normality assumption can be achieved when increasing the number of replications.

4.8 Results of the work and further findings

4.8.1 Result 1: Simulation with a standard simulation software is possible, but some restrictions exist

It has been demonstrated, that it is possible to develop simulation networks for railway networks with standard simulation software like ARENA. However, we have also seen that there are still some unexpected difficulties. For example when using a simulation which is quite close to the physics of train motion as in the first model approach the models will become very complicated with increasing network complexity. Also calculation times may become inacceptable when searching for very accurate calculation by shortening simulation intervals. When using the possibilities of the software (like the guided transporter feature of

ARENA) however other restrictions exist, which do not allow to simulate different infrastructure very accurately.

4.8.2 Result 2: Simulation without random effects as tool to calculate network capacity

Simulation (without random effects) is an option to calculate capacities of railway networks. Depending on the railway network, values may differ from other calculation methods. Further work may provide further inside, why for example in scenario 1 simulation leads to low capacity values.

4.8.3 Result 3: Simulation with random effects as tool to predict practical network capacity

Simulation with random effects is an option to predict practical capacities of railway networks. Depending on the railway network, values may differ from other calculation methods. Further work may provide further inside, why for example in scenario 1 simulation leads to low capacity values.

4.8.4 Result 4: Simulation with random effects is a tool to predict practical network capacity in the transition phase

In practical traffic, the steady state is not realistic. In fact, the transition between different usages is much more often the case in real railway systems. For the simulation witd.Simulation (without random effects) is also an option to calculatcapacities of railway networks. Depending on the railway network, values may differ from other calculation methods. Further work may provide further inside, why for example in scenario 1 simulation leads to low capacity values.

During the simulation procedures some additional findings were made.

4.8.5 Finding 1: transition phase and steady state

In practical traffic, the steady state is not realistic. In fact, the transition between different usages is much more often the case in real railway systems. It has been demonstrated in this work that simulation in ARENA is a suitable tool to simulate the steady state. In addition, it is also a tool to simulate the transition phase. But additional work seems necessary.

To calculate the capacity of a steady state, the trains have to be introduced into the simulation network in the predefined timetable over a longer time. Than the behavior of the system is viewed. The number of guided transporters is adjusted so that no major queueing up of entities is viewed in the main station, the entrance place. The simulation is then run for a longer time, here a whole day, and the results taken with neglecting the values from the warmup phase.

To calculate the transition phase, the warmup phase is taken as time of interest when the ramp up is in the focus of research. But also the emptying process can be evaluated, when the full system is emptied just by stopping insertion of additional entities.

4.8.6 Finding 2: OPTQUEST engine for development of a timetable

It is possible to find a feasible timetable for a given scenario using OptQuest for ARENA. This approach has some advantages over random choose or other manual methods. Most important, this approach is an automated search. So the person performing a simulation study is unburdened. Additional features of the desired timetable can be chosen. At first the desired amount of traffic can be directly input. Additional the goals for certain performance criteria in the network can be chosen as constraints. However, there is also a drawback, since through the heuristic search approach it is not generally possible to find the best timetable. Here also further scientific effort is necessary.

4.8.7 Finding 3: Process analyzer from ARENA

During the calculation during the work on this paper the process analyzer was used with success for the capacity of simulated networks. Unfortunately, a lot of manual input is necessary to perform such search. Additionally this approach seems not appropriate for complicated situations with lots of different variables.

5 PLANS FOR FURTHER, ADDITIONAL WORK

5.1 Improve the simulation model

In this paper two different approaches for development of a simulation mode were used.

In the first approach it was attempted to simulate the physics of railway transportation as close to reality as possible. This gave a very complicated model which was very complicated. In the end this model could not be used for complicated scenarios since computational resources were not actually available. Perhaps in the future this model could be used, if more computer capacity is available.

The second approach was using the capabilities of the simulation software used. Unfortunately it was not possible with this model approach to simulate different signaling regimes to full satisfaction. So additional improvement for this model using guided transporters is advised.

5.2 Develop a test method for normality for simulation results

As described in paragraph 4.3.3.1 for comparison of different scenarios of simulation models for statistical significance, the t-test of hypotheses could be a possible approach. However it is necessary to demonstrate, that simulation results are distributed close to normality. As demonstrated in 4.3.3.2 this assumption cannot be justified for a small number of replications. It would be of interest if the normality assumption for large numbers of replications is justified when simulating railway networks. Additionally it would be important for a future simulation study to develop a criteria for the minimal number of replications necessary to justify normality assumption.

5.3 Additional improvements of the simulation

In this paper results of literature were used to develop models for variances due to external factors. It was demonstrated that with this data results could be obtained via simulation which come close to observed performance. However it was not possible to get such good results for all scenarios.

As an improvement, in the future a real railway network should be chosen as scenario. Then performance values could be measured in the real world. These values could be compared with the results of a simulation study of this network. Then additional adjustments on the models for the network as well as on the models for the variances could be made. Also then it could be observed if changes to the real network could also be reproduced by simulation.

5.4 Apply the results also in other fields of research

In this paper the simulation was performed to discover the effects changes in infrastructure have on railway operation. As indicated in the abstract, the concepts of this simulation could be also used to analyze other automated transport systems. For example in modern factories transportation of semi-finished goods between workstations is fully automated. To prevent from accidents, the same concepts as in railway operation could be applied. In this case it would be of major benefit, to do some research on different infrastructure before introducing these changes. One possible application may be also in road traffic, where there is also a trend towards automation to reduce number and severity of accidents.

REFERENCES

Abril, M.; Barber, F.; Ingolotti, L.; Salido, M.A.; Tormos, P.; Lova, A. (2007). An assessment of railway capacity. In Transportation Research Part E, Vol. 44 (pp.774-806).

Adtranz; Alcatel; Alstom; Ansaldo signal; Invensys rail; Siemens (Editors) (2000). Glossary of UNISIG Terms and Abbreviations Subset 023 v200

Ahuja, Ravindra K. ; Jha, Krishna C.; Liu, Jian (2007). Solving Real-Life Railroad Blocking Problems. In Interfaces Vol 37, No. 5 (pp. 404-419).

Anderson, Nicholas P.; Evans, Gerald W. (2008). Determination of operating policies for a barge transportation system through simulation and optimization modeling. In Proceedings of the 2008 Winter Simulation Conference (pp. 2756-2760).

Barbu, G. (2010). E-Train – Broadband Communication with moving trains Technical Report Technology state of the art. UIC International Union of Railways ISBN 978-2-7461-1849-2 (English Version)

Barnhart, Cynthia; Jin, Hong (1998). Railroad Blocking: A Network Design Application. In Operations Research Vol. 48, No. 4 (pp. 603-614).

Beck, Katja (2005). Methodik zur Gestaltung eines migrationsfähigen Systemdesigns am Veispiel der Zugbeeinflussung. (Diplomal Dissertation) Received from Technische Universität (TU) Ilmenau, Fachbereich Informatik und Automatisierung (in German).

Büker, Thorsten (1010). Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen. (Doctoral Dissertation) Received from Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen (in German).

Caprara, Alberto; Fischetti, Matteo; Toth, Paolo (2001). Modeling and Solving the Train Timetable Problem. In Operations Research Vol. 50, No. 5 (pp. 851-861).

Chang, C.S.; Ho, S.L.; Chan, T.T.; Lee, K.K. (1993). Fast AC train emergency rescheduling using an event driven approach. IEEE proceedings-B, Vol. 140, No.4 (pp.281-288).#

Chu, Zifu (2014). Modellierung der Wartezeitfunktion bei Leistungsuntersuchungen im Schienenverkehr unter Berücksichtigung der transienten Phase. (Doctoral Dissertation) Received from Universität Stuttgart, Institut für Eisenbahn- und Verkehrswesen (in German).

Clymer, John R. (1995). System design and evaluation using discrete event simulation with AI. European Journal of Operational Research Vol. 84 (1995) (pp. 213-225).

Confessore, G.; Liotta, G.; Cicini, P.; Rondinione, F.; De Luca, P. (2009). A Simulation-Based Approach for Estimating the Commercial Capacity of Railways. In Proceedings of the 2009 Winter Simulation Conference (pp. 2542-2552).

Craven, Paul Vincent (2009). Simulating and Analyzing Wireless Railway Control Networks Using NS-2. (Doctoral Dissertation) Retrieved via ProQuest UMI Number 3360993.

Di Febbraro, Angela; Recagno, Valerio; Sacone, Simona (1994). Synchronization and Control of a Multimodal Transportation System. Retrieved from IEEE as 0-7803-1772-6/94 (pp. 750-753).

Ekren, Banu Yetkin (2009). Simulation and Analytical Models for Autonomous Vehicle Storage and Retrievel System. (Doctoral Dissertation Proposal) Retrieved from University of Louisville, Ky, Department of Industrial Engineering.

ERA, 2012, Railway safety performance in the european union 2011, ISBN 978-92-9205-014-6

ERTMS Platform (2008). ETCS Implementation Handbook. UIC International Union of Railways ISBN 2-7461-1499-2 (English Version).

Eurostat (2009); Illustrated glossary for transportation statistice

Fiorioni, M. M.; Franzese, L. A. G; Pereira, N. Y. H. ; Machado, M. N. (2005). Railroad Infrastructure Simulator. In Proceedings of the 2005 Winter Simulation Conference (pp. 2581-2584).

Franzese, Luiz Augusto Gago; Fioroni, Marcelo Moretti; Botter, Rui Carlos (2003). Railroad simulator on closed loop. In Proceedings of the 2003 Winter Simulation Conference (pp. 1602-1606).

Goverde, R.M.P; Odijk, M.A. (2002). Performance evaluation of network timetables using PETER. In Allan, J., Andersson, E., Brebbia, C.A., Hill, R.J., Sciutto, G., Sone, S., Computers in Railways VIII, WIT Press, Southampton (10 pages).

Goverde, Rob M.P. (2005). Punctuality of Railway Operations and Timetable Stability Analysis. (Doctoral Dissertation) Received from Technische Universiteit Delft, ISBN 90-5584-068-8.

Goverde, Rob M.P.; Daamen, W.; Hansen, I.A. (2008), Automatic identification of route conflict occurences and their consequences, Computers in Railways XI, (pp. 473-482)

Hansen, Igo Arne (2010). Railway Network Timetabling and Dynamic Traffic Management. In International Journal of Civil Engineering, Vol 8, No.1 (pp. 19-32).

Hansen, Igo Arne; Pachl, Jörn (editors). (2008) Railway Timetable & Traffic – Analysis-Modelling-Simulation. Eurailpress ISBN 978-3-7771-0371-6

Hill, R. John; Bond, Lpouisa J. (1995). Modelling moving-block railway signalling systems using discrete-event simulation. Retrieved from IEEE as 0-7803-2556-7/95 (pp. 105-111).

Hillier, Frederick S.; Lieberman, Gerald J. (9th Edition 2010) Introduction to Operations Research. McGraw Hill ISBN 978-0-07-337629-5

Huang, Yilin; Verbraeck, Alexander (2009). A Dynamic Data-Driven Approach for Rail Transport System Simulation. In Proceedings of the 2009 Winter Simulation Conference (pp. 2553-2562).

Jha, Krishna C.; Ahuja, Ravindra K.; Sahin, Güvenc (2007). New Approaches for Solving the Block-to-Train Assignment Problem. In Wiley InterScience. DOI 10.1002/net.2D195 (pp. 48-62).

Kanacilo, Elisangela Mieko; Verbraeck, Alexander (2006). Simulation services to support the control design of rail infrastructures. In Proceedings of the 2006 Winter Simulation Conference (pp. 1372-1379).

Kelton, W. David; Sadowski, Randall P.; Swets, Nancy B. (5th edition 2010) Simulation with ARENA, International Edition 2010. McGraw Hill ISBN 978-007-126771-7

Kleijnen, Jack P.C. (2008) Design and Analysis of Simulation Experiments. Springer Science+Business Media ISBN 978-0-387-71812-5

Krueger, Harald (1999). Parametric Modeling in Rail Capacity Planning. In Proceedings of the 1999 Winter Simulation Conference (pp. 1194-1200).

Law, Averill M. (4th international edition 2007) Simulation Modeling and Analysis, McGraw Hill ISBN 978-007-125519-6

Li, Feng; Gao, Ziyou; Li, Keping; Yang, Lixing (2008). Efficient scheduling of railway traffic based on global information of train. Transportation Research Part B, Volume 42 (pp. 1008-1030).

Liu, Shi Quiang; Kozan, Erhan (2011). Scheduling Trains with Priorities: A No-Wait Blocking Parallel-Machine Job-Shop Scheduling Model. In Transportation Science Vol. 45, No. 2 (pp. 175-198).

Minin, Vladimir A.; Johnston, Donald A.; Lepsky, Natalie A. (1995). A General Model for Optimization of Track Circuit Parameters. Retrieved from IEEE 0-7803-2556-7/95 (pp. 113-117)

Montgomery, Douglas C. (7th international edition 2009) Design and Analysis of Experiments. John Wiley & Sons ISBN 978-0-470-39882-1

Montgomery, Douglas C.; Runger, George C. (5th edition SI version 2011). Applied Statistics and Probability for Engineers. John Wiley & Sons ISBN 978-0-470-50578-6

Muthmann, Thilo (2004). Rechnerische Bestimmung der optimalen Streckenauslastung mit Hilfe der Streckendurchsatzleistung. (Doctoral dissertation) Retrieved from Universität Darmstadt 2004, D17. ISSN 1614-9300 (in German).

Murali, Pavankumar (2010). Strategies for Effective Rail Track Capacity Use. (Doctoral Dissertation) Retrieved via ProQuest UMI Number 3434473.

Nießen, Nils (2008). Leistungskenngrößen für Gesamtfahrstraßenknoten. (Doctoral Dissertation) Received from Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen (in German).

Nunez, F.; Reyes, F.; Grube, P.; Cipriano, A. (2010). Simulating Railway and Metropolitan Rail Networks: From Planning to On-Line Control, in IEEE Intelligent Systems Magazine (v2, n4, pp.18-30)

O´Brien, Daniel K. (2005). Optimal Estimation and Rail Tracking Analysis. (Doctoral Dissertation) Retrieved via ProQuest UMI Number 3161157.

Paolucci, Massimo, Pesenti, Raffaele (1999). An Object-Oriented approach to Discrete-Event Simulation Applied to Underground Railway Systems. Simulation, Vol. 72, No. 6 (pp. 372-383).

Pawar, Sam Paul Singh (2011). An Analysis of Single Track High Speed Rail Operation. (Master of Science Dissertation). Retrieved from University of Birmingham, College of Engineering, School of Civil Engineering.

Pegden, Claude Dennis; Shannon, Robert E; Sadowski, Randall P. (1995). Introduction to Simulation Using SIMAN 2nd. Edition McGraw-Hill ISBN 0-07-049320-0

Rebreyand, Paul (2005). DisTrain: A simulation Tool for train dispatching. In Proceedings of the 8th IEEE Conference on Intelligent Transportation Systems, Vienna Austria, Sept. 2005 (pp. 801-806)

Saranen, Juha; Hilmola, Olli-Pekka (2007). Evaluating the competitiveness of railways in timber transports with discrete-event simulation. World Review of Intermodal Transportation Research, Vol. 1, No. 4 (pp. 405-458).

Sargent, Robert G. (2010). Verification and Validation of Simulation Models. In Proceedings of the 2010 Winter Simulation Conference (pp. 166-183).

Schachtebeck, Michael (2009). Delay Management in Public Transportation: Capacities, Robustness, and Integration. (Doctoral Dissertation) Received from Georg –August-Universität Göttingen.

Schmidt, Christine (2009). Beitrag zur experimentellen Bestimmung der Wartezeitfunktion bei Leistungsuntersuchungen im spurgeführten Verkehr. (Doctoral Dissertation) Retrieved via Universität Stuttgart (in German).

SMA und Partner AG (2011). Audit zur Betriebsqualitätsüberprüfung Stuttgart 21 Schlussbericht (Version 1-00), Downloaded on 26.7.2011 from www.bahn.de. (in German)

Stanley, Peter (editor) (2011). ETCS for Engineers. Eurailpress ISBN 978-3-7771-0416-4

Sturrock, David T. (2010). Tips for successful Practice of Simulation. In Proceedings of the 2010 Winter Simulation Conference (pp. 87-94).

Theeg, Gregor; Vlasenko, Sergej (editors). (2009) Railway Signalling & Interlocking – International compendium. Eurailpress ISBN 978-3-7771-0394-5

UIC, International Union of Railways (2004). UIC Code 406 Capacity. ISBN 2-7461-0802-X (English Version).

UIC, (2008). ETCS Implementation Handbook, ISBN (English Version) ISBN 2-7461-1499-2

UIC Line comparison study group (2011). Line Comparison Study. UIC International Union of Railways ISBN 978-2-7461-1970-3 (English Version).

UIC ERTMS Training Programme 2010 (2010). Signalling programme, contributions by different authors. Different presentations held on an UIC training seminar 23rd and 24th of June 2010 in Paris. Retrieved from Internet www.uic.org on 2011, August, 27th.

Vakhtel, Sergey (2002). Rechnergestützte analytische Ermittlung der Kapazität von Eisenbahnnetzen. (Doctoral Dissertation) Retrieved from Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen (in German).

Van der Meer, Ir. Dirk; Goverde, R.M.P;, Hansen, I.A. (2009). Prediction of train running times using historical track occupation data. Retrieved from Technische Universiteit Delft, Department of Transport & Planning, Faculty of Civil Engineering and Geosciences.

Van Dijk, Nico M. (2000). On Hybrid Combination of Queueing and Simulation. In Proceedings of the 2000 Winter Simulation Conference (pp. 147-150).

Van Oort, Niels (2011). Service Reliability and Urban Public Transport Design. (Doctoral Disseration) Retrieved from TRAIL Research School, Delft, The Netherlands, IISBN 978-90-5584-142-4.

Verkehrswissenschaftliches Institut der Rheinisch-Westfälischen Technischen Hochschule (RWTH) Aachen (2008). Influence of ETCS on line capacity Generic study. UIC International Union of Railways ISBN 2-7461-1455-0 (English Version).

Verkehrswissenschaftliches Institut der Rheinisch-Westfälischen Technischen Hochschule (RWTH) Aachen (2010). Einfluss des European Train Control System (ETCS) auf die Leistungsfähigkeit von Knoten. UIC International Union of Railways ISBN 978-2-7461-1844-7 (German Version).

Wende, Dietrich (2003) Fahrdynamik des Schienenverkehrs. Vieweg + Teubner ISBN 978-3-519-00419-6 (in German)

Winter, Peter (UIC, editor). (2009) Compendium on ERTMS European Rail Traffic Management System. Eurailpress ISBN 978-3-7771-0396-9

Xu, Xiaoming; Li, Keping; Yang, Lixing; Ye, Jingjing (2014). Balanced train tabeling on a single-line railway with optimized velocity. In Applied Mathematical Modelling 38 (2014) (pp. 894-909)

Yoo, Seung Pil; Lee, Doo Yong; Son, Hyoung Il (2001). Design and Verification of Supervisory Controller of High-Speed Train. ISIE 2001, Pusan, Korea (pp. 1290-1295).

Yuan, Jianxin (2010). Capturing stochastic variations of train event times and process times with goodness-of-fit-tests. (12 pages).

Zhu, Xiang (2001). Computer-based Simulation Analysis of Railway Carrying Capacity Utilization. Retrieved from IEEE 0-7803-7010-04/01 (pp. 107-112).

APPENDIX A

CALCULATING THE TIME NEEDED TO TRAVEL FOR A GIVEN DISTANCE

The travel time can be calculated by mathematical integration. Another possibility is a stepwise iterative procedure. As the relation between speed and corresponding acceleration is given as step-function (not smooth) the integration is not simple, so stepwise integration was chosen.

The corresponding calculations are based on (Wende 2008, pages 64 to 66) together with UIC 2008.

The accelerations are given as constant in certain speed ranges (see UIC 2008). This is a good approximation for the physical behavior of a Eurocity-train.

At first, there is the given acceleration, which has the highest value of the whole driving area. The acceleration is limited by the adhesion between train wheels and the track. As this is (normally) a steel-steel adhesion, there is a limit on the possible transmissible forces, due to the fixed weight of the train, respective the locomotive. The electronic control unit of the train limits this acceleration, otherwise the wheel could slip on the track and so no acceleration would be transmitted and the train and the track could be damaged.

Additionally, higher accelerations are not feasible in train service, as passengers could not feel comfortable if train acceleration would be much higher. For the Eurocity train, the acceleration is limited to 0.5 m/s² and is considered constant until a switching speed of 12.5 m/s.

In the next speed range the acceleration is limited to a lower value, in the case of the Eurocity train to an acceleration of 0.4 m/s². The reason for the decrease is the fact, that the power of the train is limited (either by the diesel engine(s) or by the traction motor(s)). The acceleration slightly in reality slightly decreases with higher speeds, but for this calculations is was approximated by a constant acceleration until a train speed 23.611 m/s.

At this point the acceleration further drops down to a value of 0.3 m/s², the physical reason for this is the decreasing due to the fixed power of the motor(s) or engine(s) of the train.

In the final step, the train reaches his upper limit of speed at 55.556 m/s. This is the maximal allowed driving speed of the train. The limit is set by external factors, like the comfort level of the train, the tolerable tear and wear and for safety reasons. Normally, the train would be able to go at a somewhat higher speed – this is very often requested in railway business by law for safety reasons – but this driving at higher speeds is not allowed. On poor tracks or for safety reasons, for example on switches (no derailment should occur) or in working areas the speed limit is further lowered. In this paper, such variance however is not considered.

As shown in the table, the time-distance relation, which is unknown, has been calculated in a delta s approach, as for the distance s (travel distance since start) equidistant intervals are considered. With only ten meter, they are quite small compared to the full travel distance of 100 km. In contrast to the delta-v approach as used by the authors of (UIC 2008) and (Vakhtel 2009), by this method it is possible to determine the travel time at certain travel distance in a single step calculation. With the delta-v-method in a second calculation step would have been necessary to calculate the travel distance for a given distance, because the support points of the curve are based on speeds and not on the distances.

For a calculation, the distance interval and the train´s speed at the entrance are the parameters. From a separate table, the acceleration is read, corresponding to the entrance speed. This is only valid, if the corresponding signal in the model states free travel. If it would signal halt, then the corresponding deceleration for the brake would be read from another table. By this method, also complicated curves like exponential dependencies could be calculated. This acceleration is the considered as constant across the distance interval. The resulting speed as f(v_{start}, s) at the exit is then calculated using equation 3.

$$v_{exit} = SQRT((2 * (delta\ s) * v_{start}) + {v_{start}}^2) \quad (3)$$

The corresponding time interval delta t as f(delta s, v_{start},v_{exit}) for traveling the distance is then calculated via equation 4.

$$delta\ t = \frac{delta\ s * (v_{start} + v_{exit})}{2} \quad (4)$$

The resulting time at the exit of the distance interval is then calculated by equation 5.

$$t_{exit} = t_{start} + delta\ t \quad (5)$$

This method is applied for the whole travel distance.

In the following tables, only some results of interest are displayed.

Simulation High-Speed Line

Eurocity-Train

Given accelerations:

EC	EC	
Va	Ve	A
0	12.5	0.5
12.5	23.611	0.4
23.611	55.55	0.3
55.55	55.55556	0
1000		0

Acceleration

AC	
Vstart	Vmax
0	200

					Dt	T
s1	s2	Va	acc	ve	0	0
0	10	0	0.5	3.162278	6.324555	6.324555
10	20	3.162278	0.5	4.472136	2.619717	8.944272
20	30	4.472136	0.5	5.477226	2.010179	10.95445
30	40	5.477226	0.5	6.324555	1.694659	12.64911
40	50	6.324555	0.5	7.071068	1.493025	14.14214
50	60	7.071068	0.5	7.745967	1.349798	15.49193
60	70	7.745967	0.5	8.3666	1.241267	16.7332
70	80	8.3666	0.5	8.944272	1.155343	17.88854
80	90	8.944272	0.5	9.486833	1.085122	18.97367
90	100	9.486833	0.5	10	1.026334	20
100	110	10	0.5	10.48809	0.976177	20.97618
150	160	12.24745	0.5	12.64911	0.803324	25.29822

					Dt	T
s1	s2	Va	acc	ve	0	0
160	170	12.64911	0.4	12.96148	0.780927	26.07915
650	660	23.49468	0.4	23.66432	0.424097	52.83624
660	670	23.66432	0.3	23.79075	0.421451	53.25769
670	680	23.79075	0.3	23.91652	0.419223	53.67692
680	690	23.91652	0.3	24.04163	0.41703	54.09395
690	700	24.04163	0.3	24.16609	0.414871	54.50882
700	710	24.16609	0.3	24.28992	0.412746	54.92156
710	720	24.28992	0.3	24.41311	0.410652	55.33222

720	730	24.41311	0.3	24.53569	0.40859	55.74081
730	740	24.53569	0.3	24.65766	0.406559	56.14737
740	750	24.65766	0.3	24.77902	0.404558	56.55192
750	760	24.77902	0.3	24.8998	0.402586	56.95451
1000	1010	27.64055	0.3	27.74887	0.36108	66.45142
1010	1020	27.74887	0.3	27.85678	0.359676	66.8111
4860	4870	55.49775	0.3	55.55178	0.1801	159.1278
4870	4880	55.55178	0	55.55178	0.180012	159.3078
4990	5000	55.55178	0	55.55178	0.180012	161.4679
5000	5010	55.55178	0	55.55178	0.180012	161.6479
9990	10000	55.55178	0	55.55178	0.180012	251.4741
10000	10010	55.55178	0	55.55178	0.180012	251.6541
10010	10020	55.55178	0	55.55178	0.180012	251.8341
10020	10030	55.55178	0	55.55178	0.180012	252.0141

….

APPENDIX B

CALCULATING THE RELATION BETWEEN ENTRANCE SPEED AND TRAVEL TIME

For all railway simulations and also for the first ARENA model developed in this paper (see Chapter 4.3) it is essential to simulate the travel time correctly. The travel time for a given distance (like a block distance in a railway network) is dependent on different factors. At first there is the signal aspect of the signaling. If the train is forced to break, the train has to break otherwise the train could travel freely. Second the speed depends on the train type and its special characteristics. A freight train is not able to show the same performance with respect to acceleration than a high speed train. Then the speed depends on track parameters, like steepness, curve radii and speed limits. In this paper these influences are considered as constant, as their influence could be predicted and is therefore quite easy to introduce additionally. Also in the chosen reference scenarios these influences were neglected. There are also some unexpected influences on travel time, like the weather, the driver´s experience and others. A list of these influences can be found in chapter 1.3. The primary delays can be introduced by adding one or more corresponding stochastic delay element(s) in the travel of the train (in ARENA the entity). Here only the deterministic aspect of the travel time is considered.

It is shown in Appendix A, that the travel time can be calculated in a stepwise procedure. Another possibility would be the (mathematical) integration. But as we saw the relation between speed and corresponding acceleration is given as step-function (not smooth) the calculation is not simply forward but must follow a stepwise approach.

APPENDIX C

SEARCHING FOR A TIMETABLE

USING OPTQUEST

As described earlier in this paper, it was found that OptQuest is suitable for finding a near optimal timetable for a given railway network.

To illustrate the process, the following extracts from a calculation with OptQuest are shown.

At first, the controls for OptQuest have to be set.

These controls are the fixed arrival times (no variance, variance is introduced inside the model using additional delay modules) of the trains according to the timetable. There must be as much variables as different trains arriving during the timetable. It is notable, that the length (in time) of the timetable need not match with the total length of simulation. In this work, the timetable has been repeated, until the desired simulation time was reached, since we were interested in the performance in the steady state. If transition phase are in the centre of interest, this assumption is no longer valid.

As responses, the recorded travel times of the different train types are taken (see Figure 49).

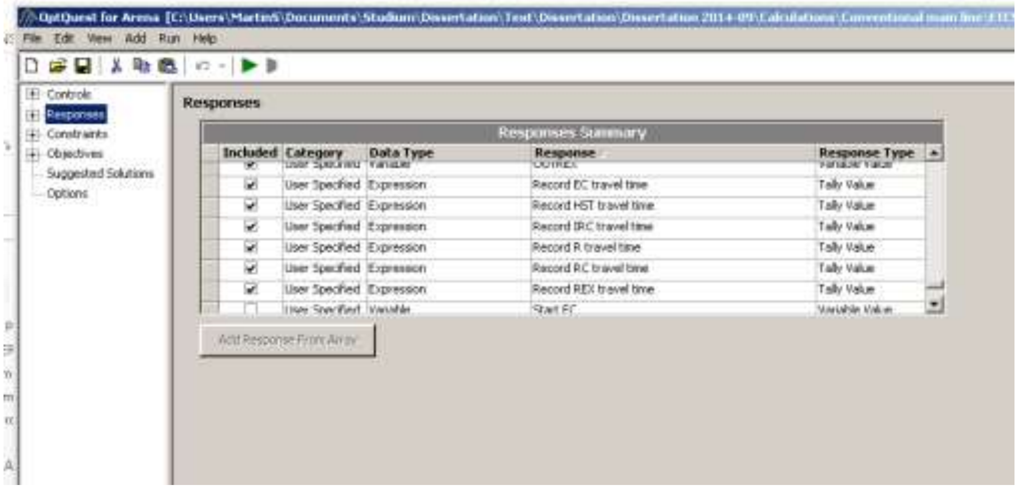


Figure 49: Responses of an OptQuest run

As constraints for the execution, the recorded travel times are used. This is a way to restrict the waiting time for the different train types separately, for example waiting time for freight trains is not so restricted than waiting time for high speed trains (see Figure 50). It is expected, that high speed train are running near top speed to minimize travel times, where possible.

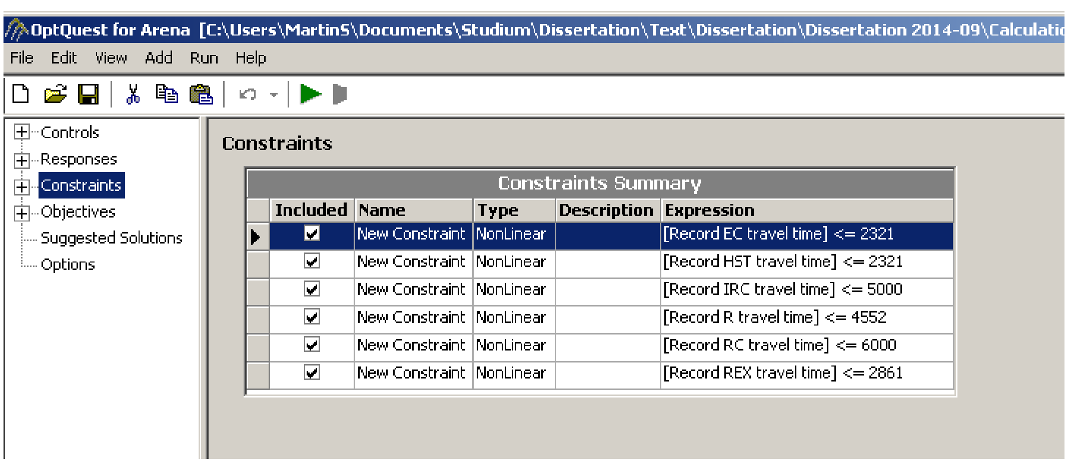


Figure 50: Constraints in OptQuest

For a successful run of OptQuest, one has to set an objective. Here it is quite simple: the objective is to maximize the number of trains traveling through the system in a given time, see also Figure 51.

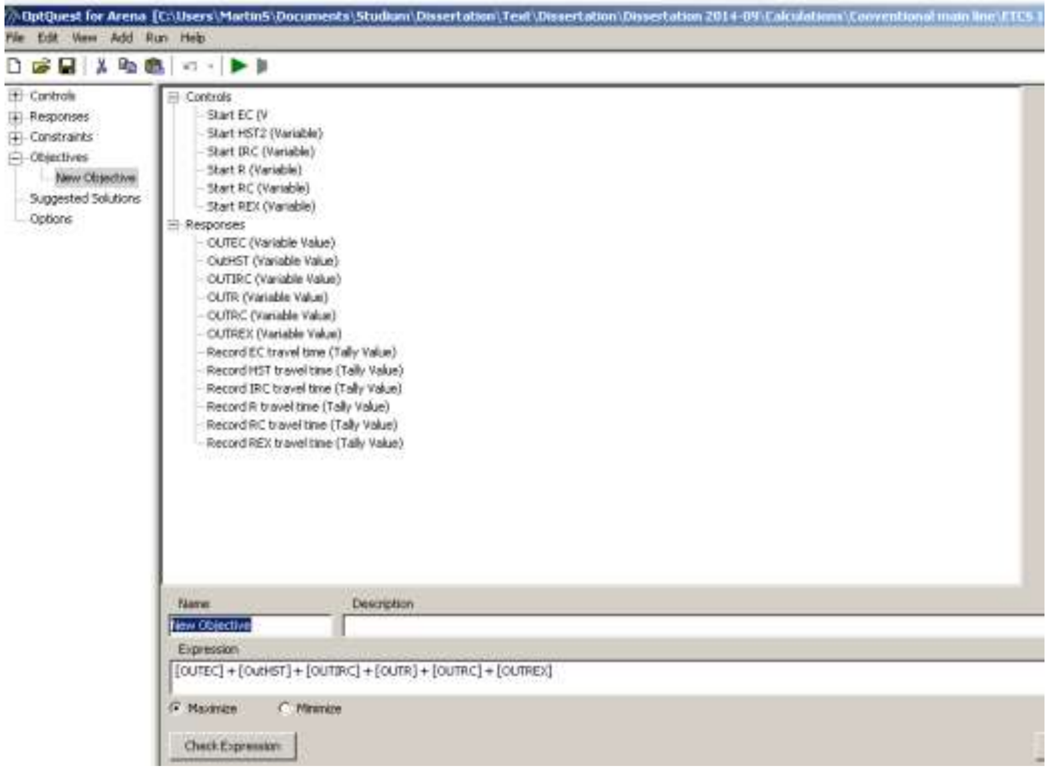


Figure 51: Objective for running OptQuest

Using a feasible starting point for the run (even a very “bad” timetable with low throughput), has been found beneficial. Otherwise it occurred quite frequently, that OptQuest could not be able to find feasible solutions. The controls of OptQuest have to be adjusted accordingly to accommodate this solution.

CURRICULUM VITAE

NAME:	Martin Johannes Reinhold Schmidt
DOB:	Guetersloh, Germany – November 4, 1966
EDUCATION & TRAINING:	Diplom-Ingenieur der Elektrotechnik Universität Hannover, Germany 1987 – 1993 Master of Business Administration Fachhochschule des Mittelstands, Bielefeld, Germany 2006 – 2008
PROFESSIONAL SOCIETIES:	Member of VDE (Verband der Elektrotechnik Elektronik Informationstechnik e.V.; Association for Electrical, Electronic & Information Technologies) Member of VDI (Verein deutscher Ingenieure; The Association of German Engineers)
PUBLICATIONS:	Seven Articles in the German Journal for Standardization “DIN-Mitteilungen” in 1/2008; 1/2010; 12/2010 and 1/2012 about meetings of IEC/TC48 and IEC/SC48B
EXPERIENCE:	WAGO Kontakttechnik GmbH & Co. KG, Minden Germany Compliance Manager for Standards 2011 – today HARTING KGaA, Espelkamp Germany Compliance Manager for Standards, Department director since 2006 2005 – 2011 HARTING Deutschland GmbH, Minden Germany Project engineer for Value Added Business 2000 – 2005

Felten & Guilleaume Anlagentechnik GmbH, Bielefeld Germany
Project engineer
1999 – 2000

DUEWAG AG, Krefeld Germany
Project engineer and project leader for diesel powered railway vehicles
1997 – 1999

ABB Trafo-BB GmbH, Brilon Germany
Development Engineer for dry-type transformers
1993 – 1997

REFERRED JOURNALS:	ELEKTRIE, Berlin 47 (1993) 11, B. R.Oswald, P. Riedel, M. J. R. Schmidt, Simulationsmodell eines supraleitenden Strombegrenzers, p. 413 – 417 (german)
BOOKS AND SYMPOSIA:	none
PRESENTATIONS AND LECTURES:	Presentation in the Seminar: Epoxidharze in der Elektrotechnik (Epoxy resins in electrotechnics) of the Technische Akadameie Esslingen,1995: New Aspects in the application of epoxy resins for dry-type-transformers