

Identifying Dependencies Among Delays

**Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen**

vorgelegt von

Carla Conte

aus Bassano del Grappa (VI)
Italien

Göttingen, 2007

D7
Referentin: Prof. Dr. Anita Schöbel
Korreferent: Prof. Dr. Axel Munk
Tag der mündlichen Prüfung: 17. Januar 2008

To Lia Zannoni

"...mortis nostræ animæ ..."

Abstract

*Weniger als 5 Minuten Verspätung:
Wir bitten um Ihre Verstandnis und Geduld.
Mehr als 5 Minuten Verspätung:
Wir bitten um Ihre Entschuldigung.*

DEUTSCHE BAHN

Mathematical optimization is a vital area in applied mathematics. Even if it came into existence about few decades ago, this field underwent not only theoretical advances. Many research efforts have been spent to study large-scale real-world applications. The enormous progress made in computer science together with the development of sophisticated numerical and algorithmic techniques are the reasons for the success in dealing with the large amount of data now available for the description of a practical problem.

In view of this background, *solving* a practical problem by means of applied mathematics is much more than applying mathematics. Implementation skills are of importance as well as the ability to communicate with practitioners and scientists from other disciplines. The proximity and interaction with computer science and statistic supply us with a spectrum of new ideas and techniques.

This application-oriented thesis deals with delay propagation in railway system, in particular with identifying dependencies among delays. The idea is to combine two different fields: stochastic and optimization. A stochastic approach points out relationship among variables representing (arrival/departure) delays of the trains at the stations of their journey. These dependencies are then transformed with a linear regression procedure into capacity constraints of a rescheduling problem. Finally the optimization problem is solved and the robustness of the new solution is checked through a comparison with the one of the uncapacitated model of the problem.

Nowadays, railway transportation needs to become more and more competitive, so new features are required to improve the planning process. In daily operations the goal is to compensate perturbations of the scheduled timetable, in particular to meet in a better way the passengers' needs concerning transfers and changes. This leads to the delay management problem. Delay propagation is often considered as one of the main reasons for the poor attractiveness of railway transport. In fact, a better re-scheduling of the timetable in order to minimize the disadvantages for the passengers will be possible if the critical points of the system are known. This includes knowledge about dependencies among delays, in order to be able to point out where the source delays are and how they spread out into the system. Delays and their behavior in railway systems have recently been investigated by [32], [40], [61], [72], [80] . . .

We distinguish, as usual in the literature (see e.g. [54]), between two types of delay: source delays, i.e. delays that are caused from the outside and not from other trains (see “Urverspätungen” in [73]), which usually spread out into the system inducing a second kind of delay, called “forced delay” (see “Folgeverspätungen” in [73]). We further distinguish between the following three types of delay propagation:

1. propagation along the same train. Delay is carried over along the path of each delayed train, i.e. if a train starts with a delay it is likely to reach its next station with a forced delay (propagation along a driving activity), and if it arrives at a station with a delay it will probably depart with a delay (propagation along a waiting activity);
2. propagation from one train to another due to connections. If a connecting train waits for a delayed feeder train, the delay of the feeder train may spread out to the connecting train (propagation along a transfer activity);
3. propagation from one train to another due to the limited capacity of infrastructure. If two trains share the same infrastructure (a part of a track or a platform) one of them has to wait until the other has left (propagation along a headway).

The first two types of delay propagation are easy to handle from an analytical point of view, since the minimal duration of every activity is known and is hence an explicit (i.e. given) parameter. However, the third kind of delay propagation is more complicated to deal with. This is due to the fact that it requires a detailed knowledge of the track system on a microscopic level.

In Chapter 1 an overview about approaches dealing with the third type of delay propagation is given.

In Chapter 2 the delay management problem is introduced. It deals with reactions in case of delays in public transportation. The timetable problem is formulated as a linear programming introducing the notion of Activity-on-arc Project Network (see [51]). The arrivals and departures of the trains in all the stations of their journey are regarded as the nodes of the network (we refer to them as the events of the system). The edges are defined by the scheduled activities of the system: waiting activities, driving activities and connections. Classical formulations of the timetable problem as Resource Constrained Project Scheduling problem (Ref: [64]) or as Job Shop Scheduling problem (Ref: [17] and [18]) are presented. The \mathcal{NP} -hardness of these formulations could be proved. Models suggested by other research groups are introduced to stress advantages and drawbacks with respect to the model considered in this thesis. In contrast to all these classical approaches, the goal of this work is to propose a procedure which enables us to detect dependencies among delays of the third type without explicit knowledge of all details of the infrastructure.

In Chapter 3 a stochastic approach to identify dependencies, called Tri-graph (see [75] and [76]), is presented. This is a simplified graphical modeling approach in which full conditional modeling is carried out in small subgraphs with only three vertices that will be then combined into the full model. The variables of the rescheduling problem are the events of an Activity-on-arc Project Network for the timetable problem in railway systems. The delay measurements of each event in this model correspond to the observations of the variables. Analyzing the dependencies of these variables with the Tri-graph method yields links among the events of the system, i.e. among the arrival and departure events of the trains in the stations. These links represent on one side the dependencies arising during the driving, waiting or transfer activities and on the other side the information about the third type of delay propagation, namely the propagation

due to capacity constraints that we are looking for. These “virtual activities” can be interpreted as inequalities that represent a lower bound for the re-scheduled time of an event ensuring that an event can not happen before another event has taken place: for example a train can enter in a station only if its assigned platform is free. Hence the “virtual activity” does not belong to the set of activities of the problem (waiting, driving and transfers), but it can be considered as a precedence constraint in the railway problem that has to be satisfied to avoid infrastructure conflicts (using the same track or the same platform) due to the limited capacity of the track system and to the operational rules of the security system.

The idea of our approach is to describe the delay propagation by using these virtual activities instead of the classical inventory constraints that are commonly used in re-scheduling problems (see Chapter 2). Indeed, the Tri-graph method has the advantage to be able to identify even complex dependencies without “a priori” knowledge of the track system. Moreover, compared to other statistical methods, Tri-graph is applicable also when the data sample is small compared to the number of random variables.

Chapter 4 presents the outcomes of the application of the Tri-graph method to real-world data of German railway, corresponding to some stations within the Harz area in Germany. The data has been provided by *Deutsche Bahn* within the context of a larger project named DisKon (see [11]) and it consists of delay data of all trains at all stations over a period of nine months. Apart from the delay, we also have information about the timetable and the infrastructure in this region, such that we are able to compare the stochastic results with the analytical constraints of the classical formulation of the problem. Moreover, a comparison with the outcomes of other standard techniques (in particular with Full Conditional Graph and with Covariance Graph) has been performed as validation of the theoretical results presented in Chapter 3.

In Chapter 5 an analysis, from a theoretical and practical point of view, of the modeling of the delay and the virtual activities has been carried out in order to transform the edges resulting from the Tri-graph procedure into time constraints for the delay management problem. Four possible linear regressions have been considered and the approximation errors have been compared considering two different strategies: the first strategy was to consider as the best-fit polynomial approximation the one that has the minimal sum of the deviations squared (least squares error), the second strategy was based on a robust estimator so that it will be relatively unaffected by outlying values (Huber’s error).

In Chapter 6 we introduce the concept of robustness and we test the “virtual” capacitated model of the problem according to three possible criteria: number of violated capacity constraints, cost (in second) of the violations and price (in second) to transform the solution into a feasible one for the Microscopical model. In order to have a larger set of observations, a new dataset is introduced as well as the corresponding numerical results associated to it. We applied the Tri-graph method with different values of the quantile and we transformed the outputs into sets of constraints for the timetable problem applying the linear regression described in Chapter 5. The “virtual” capacitated models are solved using an heuristic algorithm and the solutions are compared with the ones of the uncapacitated model according to the three defined criteria.

The results show that the Tri-graph method is able to locate the most important headways of the problem. It does not suffer so much of the transitivity property as the Covariance Graph (hence its output is much smaller) and it is much easier to apply than the Full Independence Covariance Graph, since it is defined on the covariance matrix. Moreover the sets of “virtual” constraints obtained from its output (especially in the case of absence of multistatistical correction) improve the robustness of the

(uncapacitated) solution in around 90% of the considered scenarios. Concerning the choice of the quantile, it was not possible to highlight any strong difference among the 5% and 10% values. Therefore we do suggest to consider the set of constraints corresponding to the Tri-graph method without multistatistical correction with a 5% quantile as “virtual” constraints.

Acknowledgments

When you arise in the morning, give thanks for the morning light, for your life and strength. Give thanks for your food and the joy of living. If you see no reason for giving thanks, the fault lies with yourself.

TECUMSEH SHAWNEE CHIEF (1768-1813)
American-Indian Warrior

During this period, that I definitely consider a journey, many people accompanied my work on this thesis in various respects. I am deeply indebted to all of them whose aid, stimulating suggestions and encouragement helped me in all the time of research and writing of this thesis. I would like to express my gratitude to all of them: for that journey, for its lessons and also for the help they offered me in my “personal” fight against the German language.

Danke! Ihr habt das alles erst möglich gemacht!

In particular I would like to express all my gratitude to my supervisor, Prof. Dr. ANITA SCHÖBEL, who introduced me to the field and community of railway network problems and who offered a fertile working environment, encouragement and support ever since.

I am obliged to Prof. Dr. AXEL MUNK and Prof. Dr. STEPHAN WAACK who both agree to become co-referees for this thesis.

A heartfelt thanks goes to HARALD BÖRNER at Deutsche Bahn for always giving his ears and sympathy.

I also want to show my gratitude to Dr. JULIANE SCHÄFER at ETH Zürich for the fruitful discussions about Tri-graph.

A special mention goes to the optimization group (ANDREAS GINKEL, MICHAEL SCHACHTEBECK, Dr. SILVIA SCHWARZE and Dr. SUSANNE WEIßMANN) for sharing information about public transportation systems and for giving me useful advice.

I really appreciate the countless discussions with Dr. THOMAS HOTZ: you had really an incredible patience!

I owe thanks to the *Deutsche Forschungsgemeinschaft* and the *Graduirtenkolleg “Identifikation in mathematischen Modellen: Synergie stochastischer und numerischer Methoden”* for the financial support.

I am indebted to Johannes Schuldt, who helped me to solve (some of) my starting technical problems with R and Python.

My thanks are extended to the whole Institute for Numerical and Applied Mathematics for providing me an excellent working surroundings. I very much enjoy the time spent with academic colleagues and I want to thank them for their help, support, interest and valuable hints. I won't name you all here, I hope to meet you again ... preferably not in Goslar.

Moreover I want to thank all the persons that share some ("un-mathematical") time of their life with me: thanks Anne for your "oldness" (and for the funny "Rødgrød med fløde" even I still cannot pronounce it), thanks Chiara for every single E-mail and for believing in me when I found it difficult to believe in myself, thanks Christian for your selfless support (like during our walk, running around the Kiessee), thanks Claudiu for your "nonchalance" (especially speaking about postcards), thanks Emmanuel for our "movie-time", thanks Gian Luca for your unconditional everything (this is more than I deserve), thanks Juri for your special vision of the everyday life, thanks Michel for my daily dose of irony (my dear old sparring partner), thanks Mihaela for your Salsa rhythm, thanks Mohammad for your midnight communicativeness, thanks Nouman for your easy-going energy ("ohne dich, Party nicht"), thanks Olena for saying what I have needed to hear instead of what I have wanted to hear, thanks Sonia for your beat and strength, thanks d. Stefano to remind me which are the important things in my life, thanks Ulrich for trusting me with your thoughts and disappointments, thanks Yuliya for your radiant splendor (and the great idea about how to use shopping carts), thanks Xiaoqin for your calm and patience (especially listening me letting off my steam) and for your incredible timing ("Mensch"), thanks God for keeping me sane ... and if I forgot to mention someone, thanks a lot to accept my apologize.

Certainly my family took on the hardest part - bearing my absence and when I was present, my moods. *Grazie*.

Last but not least: thanks G'Ma for just have been out of scale!

I have no other words to say Thanks, but I mean that with all my heart.

Contents

Abstract	I
1 Introduction to the Problem	1
1.1 Timetabling and Resource Allocation	1
1.1.1 Railroad traffic control system	2
1.2 Analysis of delay propagation	5
1.3 Research objectives	6
1.4 Panorama of possible approaches	7
2 Analytical Description	13
2.1 Formulation of the Problem	13
2.2 \mathcal{NP} -completeness	18
2.3 [CTM-3] is \mathcal{NP} -hard	20
2.4 Shop Scheduling Problem	22
2.5 [CTM-4] is \mathcal{NP} -hard	26
2.6 Capacity-Constrained Problem	27
2.7 Train traffic deviation handling using TS and SA	29
2.8 Resource Constrained Project Scheduling Approach	32
2.9 [RCTM] is \mathcal{NP} -hard	34
2.10 Pointing out the difference between [RCTM] and [CTM-4]	36
2.11 Identifying dependencies through a stochastic approach	40
3 Stochastic Approach	43
3.1 Probability and Independence	43
3.2 Statistical test	44
3.3 Chi-squared Test on a Contingency Table	45
3.4 Graphical Methods	48
3.4.1 Graph: notation and terminology	48
3.5 Continuous method	52
3.5.1 Full Conditional Independence Graph	53
3.5.2 Covariance Graph	53
3.5.3 Tri-graph	53
3.5.4 Some examples	54
3.5.5 Some theoretical results	60
3.6 Statistical Multicorrection	64
3.7 Tri-graph estimation algorithm	65
3.8 Complexity of the Algorithm	65

4 Numerical Results	69
4.1 The DisKon Project	69
4.1.1 The raw data	70
4.1.2 Working with the data	72
4.2 Outputs of the samples $W = 30$ and $W = 195$	74
4.3 Outputs of some subsamples of $W = 195$	79
4.4 Outputs of the Saturdays and Sundays samples	83
4.5 Conclusion	85
5 Modeling delays and virtual constraints	87
5.1 Distribution of the delays	87
5.2 How to represent delays	89
5.3 Modeling train delays in Urban Networks	90
5.4 Some considerations about the normal distribution.	93
5.5 About the covariance	95
5.6 Writing the “virtual” constraints	97
5.6.1 Four alternatives	99
5.6.2 Least squares error	102
5.6.3 Empirical comparison	104
5.6.4 Robust Estimators	109
5.6.5 Slope and intercept	110
6 About the Robustness	121
6.1 Robustness	121
6.2 The system	122
6.3 Final consideration	134
7 Conclusion	135
A Maximum Likelihood Estimation	137
A.1 A short introduction	137
A.2 Computational Problem	138
A.3 The Likelihood Ratio Test	138
A.4 Errors	140
Bibliography	141
Curriculum Vitae	146

List of Figures

1.1	Timetable Scheduling Project	2
1.2	Home signal limits and block sections	3
1.3	Home signal limits with intermedium interlocking signals	4
1.4	Train gate controller automat (from [40])	9
2.1	Example of Public Transportation Network	14
2.2	Example of Activity-on-arc Project Network	15
2.3	Blocks sections	17
2.4	Explanation of the parameter r_{ie}	33
2.5	Makespan vs. total tardiness	38
3.1	Example of undirected cycle in the railway system	49
3.2	Example of the separation concept on a graph	51
3.3	Example without cycles in the Full Conditional Independence Graph	55
3.4	Example with cycles in the Full Conditional Independence Graph	56
3.5	Example of Covariance Graph as subgraph of Full Conditional Independence Graph	57
3.6	single track line	58
3.7	Tri-graph coincides with FCIG	59
3.8	Triangular connection	59
3.9	Tri-graph coincides with CG	61
4.1	Harz area	70
4.2	Location of the considered stations	71
5.1	The variable Max(S,T)	95
5.2	Offsets of the linear regression	98
6.1	Concept of robustness	122
6.2	Robustness criteria: headways function	129
6.3	Robustness criteria: second function	130
6.4	Robustness criteria: price function	131
6.5	Robustness criteria: Comparison for NC 5%	132
7.1	... just an “ironic” conclusion	136
A.1	Likelihood function with local minima	138
A.2	MLE errors of type I and II	140

List of Tables

2.1	Reductions among objective functions	24
2.2	Particular addition to Table 2.1	25
2.3	Reductions among constraints of the Job Shop Scheduling problem	26
4.1	Characteristic of the samples $W - 30$ and $W - 195$	73
4.2	Outputs of the Contingency Table test for the samples $W - 30$ and $W - 195$	74
4.3	Output of the Covariance Graph for the samples $W - 30$ and $W - 195$	75
4.4	Output of the Tri-graph for the sample $W - 30$	76
4.5	Output of the Tri-graph for the sample $W - 195$	76
4.6	Comparison of different methods on the samples $W - 30$ and $W - 195$	77
4.7	Outputs of the Tri-graph for the samples $MWF - 117$, $TT - 78$ and $Mo - 39$	78
4.8	Comparison of Tri-graph outputs for the samples $W - 195$, $MWF - 117$, $TT - 78$ and $Mo - 39$	79
4.9	Outputs of the Tri-graph for the sample $Tu - 39$	80
4.10	Output of the Tri-graph for the sample $We - 39$	80
4.11	Output of the Tri-graph for the sample $Th - 39$	81
4.12	Output of the Tri-graph for the sample $Fr - 39$	81
4.13	Comparison of the outputs of the Tri-graph for samples $W - 195$, $Mo - 39$, $Tu - 39$, $We - 39$, $Th - 39$ and $Fr - 39$	82
4.14	Comparison of the virtual edges for the samples $W - 195$, $Mo - 39$, $Tu - 39$, $We - 39$, $Th - 39$ and $Fr - 39$	82
4.15	Comparison of virtual edges for the samples $W - 195$, $MWF - 117$, $TT - 78$ and $Mo - 39$	83
4.16	Output of the Tri-graph for the sample $Sa - 39$	84
4.17	Output of the Tri-graph for the sample $Su - 39$	84
4.18	Characteristic of the samples $Sa - 39$ and $Su - 39$	85
5.1	Comparison of LSE on edges TG (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$	105
5.2	Comparison of LSE mean on edges TG (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$	105
5.3	Comparison of LSE on the virtual connections (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$	106
5.4	Comparison of LSE mean on the virtual connections (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$	106

5.5	Comparison of LSE on edges TG (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78 and <i>Mo</i> –39	107
5.6	Comparison of LSE mean on edges TG (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78 and <i>Mo</i> –39	107
5.7	Comparison of LSE on the virtual connections (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78 and <i>Mo</i> –39	108
5.8	Comparison of LSE mean on the virtual connections (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78 and <i>Mo</i> –39	108
5.9	Mean of slope and intercept on all the edges of TG (no corr.) for the samples <i>W</i> –30, <i>W</i> –195, <i>Sa</i> –39 and <i>Su</i> –39	112
5.10	Mean of slope and intercept on all the edges of TG (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78, and <i>Mo</i> –30	113
5.11	Mean of slope and intercept on all the waiting activities (no corr.) for the samples <i>W</i> –30, <i>W</i> –195, <i>Sa</i> –39 and <i>Su</i> –39	114
5.12	Mean of slope and intercept on all the waiting activities (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78, and <i>Mo</i> –30	115
5.13	Mean of slope and intercept on all the driving activities (no corr.) for the samples <i>W</i> –30, <i>W</i> –195, <i>Sa</i> –39 and <i>Su</i> –39	116
5.14	Mean of slope and intercept on all the driving activities (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78, and <i>Mo</i> –30	117
5.15	Mean of slope and intercept on all the “virtual connection” (no corr.) for the samples <i>W</i> –30, <i>W</i> –195, <i>Sa</i> –39 and <i>Su</i> –39	118
5.16	Mean of slope and intercept on all the “virtual activities” (no corr.) for the samples <i>MoWeFr</i> –117, <i>TuTh</i> –78, and <i>Mo</i> –30	119
6.1	Outcome of the Tri-graph for the samples 2 <i>HH</i> , 4 <i>HH</i> and 6 <i>HH</i>	124
6.2	Number of violated headways	126
6.3	Percentage of violated headways	126
6.4	Cost in seconds of the violated headways	127
6.5	Percentage of the cost in seconds of violated headways	127
6.6	Price in seconds of correction	128
6.7	Percentage of price in seconds of correction	128
6.8	Nr. of scenarios with improvement w.r.t. the uncapacitated model	133

Introduction to the Problem

Delay is preferable to errors.

THOMAS JEFFERSON (1743-1826)
3rd President of the United States

1.1 Timetabling and Resource Allocation

A (re-)scheduling of the railway-timetable is a project involving a considerable amount of money, personnel and equipment. It is usually initiated either by some need of the railway system or by some particular customers' requests (e.g. soccer world championship). According to the definition of project management and resource allocation given in [64], its life cycle can be schematized into five consecutive phases.

Starting point is the *timetable concepting phase* in which some proposals and analyses are considered in order to decide whether or not the project will be performed. The *timetable definition phase* fixes the objectives of the project (e.g. minimizing the total traveling time of the trains), the different tasks to be performed (e.g. journeys of the trains) and the resources (trains, tracks, platforms, crews, ...) that will be assigned to the project. Then, in the *project planning phase*, every task is decomposed into a sequence of precedence-related activities. The time and the resource estimations provide the total duration and the resource requirements for every activity as well as temporal constraints between activities related by a precedence relationship. The result of this analysis can be represented in a network modeling the activities and the prescribed priority list among them. Moreover the timetable constraints define temporal windows bounded by the earliest and latest starting times as well as slack times, in respect of the resources availability. The last issue of the timetable is the most complicated and consists in allocating the limited resources to the execution of the activities. During this phase, called *timetable execution phase*, it must be checked if any suggested decision is against the schedule that has been established in the planning phase. In case of conflict the resource allocation has to be performed again to guarantee the feasibility of the project. The last step, *timetable termination phase*, consists in evaluating the new timetable to facilitate future management of the project.

The process can be schematically represented as in Figure 1.1.

The complexity of resource allocation comes from the interaction among *activities* by explicit and implicit dependencies, which may be subject to some degree of uncertainty. Explicit dependencies are given by the precedence relationships among activities arising from technical or organizational requirements. In this work, these dependencies will be transformed into temporal constraints. The limited resources give implicit dependencies among activities that can be formulated as resource constraints for all the activities competing for the same resource (it is also possible to add a penalty in the object function to avoid excesses).

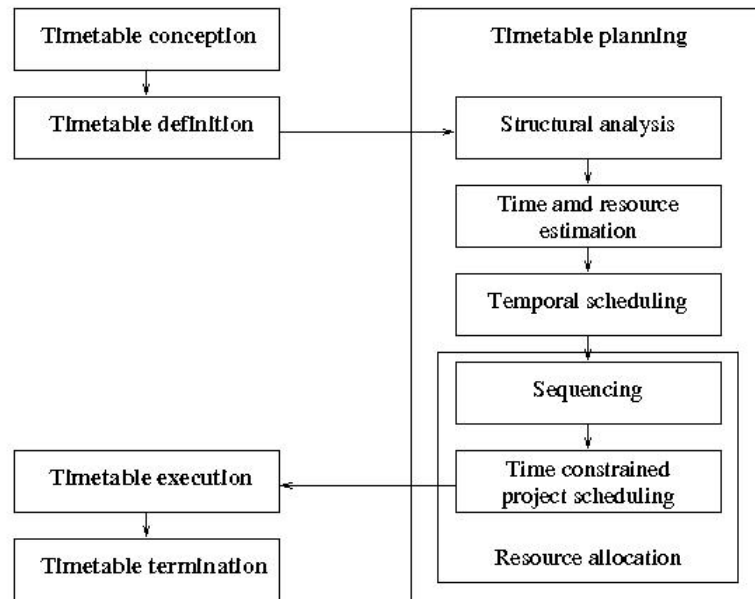


Figure 1.1: Timetable Scheduling Project

In the *timetable allocation* problem time windows are assigned to every activity corresponding to feasible execution time, taking into consideration all the temporal and resource constraints (the so called *resource-constrained problem*).

Within the *Resource allocation* sub-phases it is possible to distinguish two subproblems: *sequencing* and *time-constrained project scheduling*.

The sparsity of resources implies the definition of another precedence relationship among activities during the resource allocation task, that can be expressed by temporal constraints. In contrast to the previous ones, these new constraints depend on subjective decisions to build up the priority list of activities, that is the core of the timetable. The schedule of time-constrained project is based on computing a feasible solution, given all the temporal constraints, that also optimizes an object function representing the goal of the project.

1.1.1 Railroad traffic control system

Railroads publish documents that help railroaders understanding how trains are to be moved across their systems. Employees are required to study these documents, to have them available while working and to meticulously obey them. These documents define the procedures by which operations will be conducted: the "when and how" of railroading. Accordingly the employee timetable is more detailed than the passenger one and contains a treasure of detailed information about the railroad and how it operates: for example it indicates the geographical location of facilities (switches, tracks, ...) and distances in kilometers to help the workers controlling which type of authority is in effect for each piece of track. Every train must have the authority to occupy a track before it can be moved and even if there are different kinds of authorities (e.g. maintenance people must also have the authority to be on or to obstruct a track), usually only one type is in effect on any given piece of a track.

Examples of rules that have to be considered during the reschedule of a timetable are:

- *security distance*: the headway must be calculated in order to satisfy all security rules of the system, e.g. the breaking space required to stop a vehicle;
- *maximal speed*: in every block of tracks, for every kind of train, it is defined a speed limit in order to avoid conflicts. This boundary defines the minimal possible duration of the journey and consequently the available slack time;
- *kind of train*: due to particular characteristic of the machines, not all trains can travel on all blocks.

Even if the rules are quite straight, it is in any case necessary to direct train movements. Before giving a small introduction to the authorities that govern the decision process (see also [14] and [15]) we prefer to introduce the concept of blocks (see [54]).

In a railroad system we can distinguish between two different kinds of tracks: station tracks and tracks of the open line. The first type can be found only inside a kind of station that is called *Bahnhof* in German. A *Bahnhof* is a set of tracks limited by opposing home signals, with at least a turnout, where trains may depart, pass by, arrive, overtake and turn. The tracks outside the home signal limits are part of the open line set (Figure 1.2).

The signals that allow a train to leave the home signal limits are called exit signals (Figure 1.3). All the other signals inside a *Bahnhof* are called intermedium interlocking signals and they identify the limits of the track sections inside the home signal limits (that can not be referred to as block section). They govern either a junction station or a crossover station (that are junctions or crossovers inside a station).

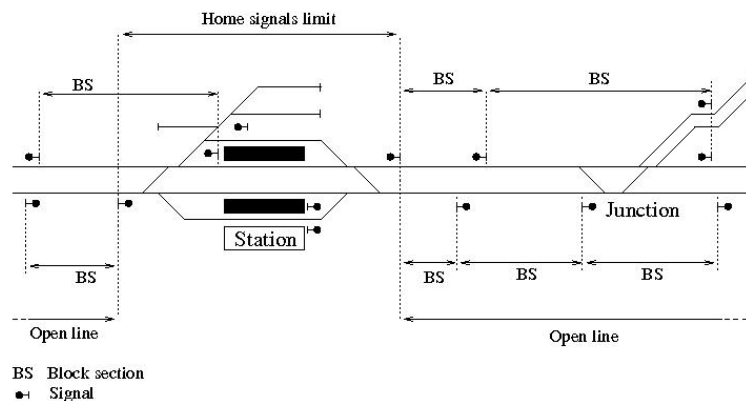


Figure 1.2: Home signal limits and block sections

According to these definitions we can introduce the most common kind of authorities: interlocking signal, centralized traffic control, track warrant control, direct traffic control and speed limits.

Interlocking signals

Interlocking signals allow movements within the limits of the interlocking plant bounded by those signals. As the name implies, they are appliances (once mechanical, now electronic) that are designed so that conflicting movements cannot be authorized.

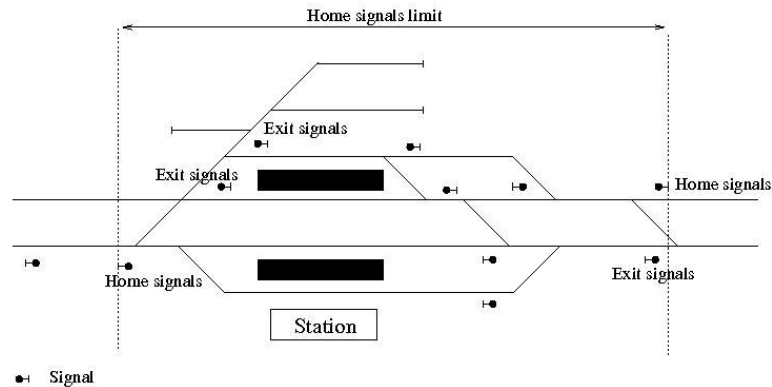


Figure 1.3: Home signal limits with intermedium interlocking signals

They can be manually or automatically controlled. The second type is the most common, for example in remote crossings of two railroads, and usually operates on a first-come-first-served base (but this rules can be changed to whatever the operator decides).

To unlock the signal so that it can be cleared again for another train, it is necessary to check the following conditions:

- the block section and the overlap must be clear;
- the train must have passed the signal at the end of the block section;
- the signal at the end of the block section must be in stop position.

These conditions can be verified with a sequence locking, i.e. a section can only be released after the next section has been occupied.

Centralized Traffic Control

The Centralized Traffic Control, also called Traffic Control System, is commonly found on high-or-medium-density lines, where the signal equipment cost can be justified by the reduction in train delays. Conceptually it is a series of interlocking signals controlled by a dispatcher. Trains are governed by signal indicators, some of which provide movement authorities (a "green/red light") in order to prevent conflicting authorities. From a console the dispatcher remotely controls signals and powered switches so that the trains need only to observe the controlled signals to obtain movement authority.

Track Warrant Control

The Track Warrant Control authorizes the dispatcher to verbally instruct the train to proceed, usually via radio. The dispatcher selects the stations and blocks of tracks among which the train may move. The authority is limited on the basis of "not in effect until after the arrival of ...". The train crew writes the instructions on a Track Warrant Form and repeats them to the dispatcher for verification. When only this authority is used nothing prevents the dispatcher from erroneously issuing overlapping or conflicting warrants. Therefore computers are usually used to check for conflicts to minimize the effects of (human) errors.

Direct Traffic Control

Direct Traffic Control is similar in execution and application to Track Warrant Control, but the railroad is subdivided into pre-defined "blocks." The dispatcher authorizes a train to proceed in one or more of the blocks but does not have the flexibility in the selection of authority boundaries available under Track Warrant Control.

Speed limits

Speed limits authorize any train to move at a speed that allows it to avoid conflicts. This is the railroad equivalent of aviations visual flight rules. A "speedmaster" may direct speed movements, but does not provide movement authorities. Trains and engines must still watch out for each other.

Each system of train movement authorities must also include provision for protecting maintenance vehicles and workers, as well as any work train that may be moving back and forward within a section of track.

1.2 Analysis of delay propagation

Even if a timetable that respects all the constraints and rules is given, it is always possible that somewhere in the system a problem occurs and so delays arise. Since delays are interactions among many random variables involved in the railway system (people, vehicles, infrastructure, weather, . . .), it is difficult to prevent them.

We can subdivide the delays into two main classes:

primary delays (also called source delay) are caused on a train from the outside and not from other trains (e.g. malfunctioning of infrastructure, bad weather conditions, accident at road-railroad crossing, . . .);

secondary delays (also called forced delay) are forced by earlier delays or other trains (e.g. sharing of the same infrastructure, fixed connections, transfer in crew schedules, . . .).

The main reasons for a delay to appear are:

- bad weather conditions;
- individual driving behavior of the engine drivers, that can affects the way train movements take place;
- strong fluctuations in boarding/transferring time of passengers;
- delayed setting up of the required route by the dispatchers or by the automatic train regulation;
- obstacles on the rail;
- sharing of the same infrastructure among a large number of rolling stocks (with different characteristics).

Except the last one that is a secondary delay, all the other reasons represent source delays that are not easily controllable in order to prevent delay propagation. In particular, it is not always possible to avoid that the delay of a single train affects several other trains. On the contrary, the propagation of delays is the main source for delays.

Delay propagation occurs because, on the one hand, there is a large number of trains using the same infrastructure, whereas on the other hand, the signaling systems impose strong restrictions on the way these trains are operated (e.g. security rules, . . .). The punctuality of the system can be improved by preventing or reducing delay propagation.

1.3 Research objectives

We focus our attention on the analysis of delay propagation: railway stations or single tracks are considered as the bottlenecks of the system (in a *Bahnhof*, as defined in the previous section, a train can wait for a green light either to reach or to leave a particular track, and a single track implies a priority list).

Such “fault” will reduce what is here called the infrastructure capacity. The concept of capacity is loosely defined, for a specific part of the railway line it usually expresses the maximum number of trains per time unit that could possibly operate on the line. This maximum number of trains is normally considered as a constant but precisely stated it is not, since it does depend not only on the hard technical facts that are captured in the functional and technical state, but also on how the trains are scheduled in the timetable and on the mixture of speed among the trains (e.g. if on a single track one train at a time is alternatively dispatched in each direction, fewer train can travel on it than if two trains at a time are alternatively dispatched in the same direction on it). Therefore the “value” of the capacity will be different if calculated for different periods of the day. In this work we consider the concept of capacity as one of the constant parameters of the system.

It results that the *Bahnhof* areas of the network have a limited (infrastructure) capacity compared to other block sections, due to the larger occupation time of infrastructure (e.g. the platform tracks) and the extra safety margins that are required to set up and release routes at crossings and switches. Consequently delay propagation is most likely to be expected in these areas since:

- trains arriving/traveling/departing in time receive a red signal if the block section ahead is still occupied by a delayed train (e.g. if two trains need to use the same junction to reach the assigned platform);
- trains may be forced to stop longer than planned at a platform to maintain connections.

Moreover since the safety level on the railway network has a very high importance, when faults occur rather than reduce the security distance, delays are allowed, slowing down or even stopping other trains in order to prevent catastrophic consequences. Therefore the railway system is very sensitive to disturbances on the timetable. The timetable in fact is designed so that it should be feasible (i.e. no delays or conflicts) if no disturbances occur. To make the timetable more “robust” in case of disturbance (i.e. the performance of the system will not be so sensitive to (small) deviations from the scheduled timetable) slack times are introduced. Obviously the easiest way to avoid the delay propagation would be to choose bigger slack times. It is always possible to design a timetable so that passengers train will stop several minutes at each station to be sure that there will be enough time for boarding and/or scheduled connections. These actions in fact reduce the variability of the travel time but that results in longer scheduled traveling times for the passengers. Another possibility will be to run fewer trains on a particular line, then automatically we will have larger time margins and hence a more robust timetable, but we will also reduce the utilization of the capacity

of the system, i.e. the money spent on the infrastructures will not be well “invested”. Since these approaches are not acceptable (either for the passengers or from the railway system), to prevent the propagation of delays it is necessary to identify the dependencies between primary and secondary delays. In such a way we may be able to find an optimal compromise among all the conditions that the system has to satisfy, so that all the passengers will be able to reach their final destination with the minimal delay. This can imply some missing connections to limit the propagation of the perturbation on the timetable.

Delay propagation analysis can be a useful tool to verify whether the capacity of the system is sufficient for a given timetable at a predefined level of punctuality (if not, the infrastructure management might decide to adjust or extend the existing infrastructure). In addition it provides new criteria for the timetable designers in order to obtain insights in how to optimize a timetable concept by adding buffer times to the “most sensitive” train successions.

In this thesis we will focus our attention on the analysis of delay propagation to find appropriate wait and depart decisions in delay management.

1.4 Panorama of possible approaches

The Train Timetable Problem is nowadays a well known problem and much progress has already been achieved in particular concerning the delay management. In a contest of an overall view, we will give in the next pages a short description of a few models that have been applied to delay propagation and/or prediction, drawing attention on some of their drawbacks. In particular we will introduce the following techniques: Process Algebra (see [5] or [9]), Markov Chain (see [80]), Linear Regression (see [13]), Wakob’s Approach (see [23]). Detailed explanations on the application of the procedure and on the results can be found in the quoted articles.

Instead of the method that will be tested in this work, a stochastic technique originally applied in the genetic field, will be separately presented in Chapter 3.

Process Algebra

Process Algebra is an active area of research in computer science and corresponds to an algebraic approach to the study of concurrent processes to ensure that they are correctly designed.

The word *Algebra* denotes that we take an algebraic or axiomatic approach in describing the behavior. So a Process Algebra is then any mathematical structure satisfying the axioms given for the basic operators. Within this structure calculations with processes that are the elements of the Process Algebra can be performed (by the axioms). The basic operators and the axioms must be defined according to the concurrent system under examinations (see [5]). Hence a deep knowledge of the process that has to be simulated is required.

Concurrent systems consist of a possibly huge number of components that not only work independently but also communicate with each other from time to time (e.g the railway signaling system). If the total amount of states is quite large, the number of possible actions can become too big to be considered.

In the deterministic case, the focus of Process Algebra is on verifying that the execution of specific actions is guaranteed by a fixed deadline after some event has happened, e.g that if a train is approaching a railroad crossing, then bars must be guaranteed to be lowered on due time.

In the stochastic case, instead, are considered systems, whose behavior cannot be deterministic predicted as it fluctuates according to some probability distribution.

Due to economical reason, these systems are referred to as shared resources systems because there is a varying number of demands competing the same resources. The consequences are mutual interference, delays due to contention, and varying service quality. In this case, the focus of process algebra is on evaluating the performance of the systems: e.g. in the railway system we may be interested in minimizing the average train delay or studying the characteristics of the flow of passengers.

Since these process can describe just certain aspects of behavior, disregarding others, they are always considered an abstraction of the “real” behavior of the system. Moreover a huge amount of details has to be taken into account (e.g. interconnection and synchronization structure, allocation and management of resources, real time constraints,...) so that it is necessary to have a close collaboration of many people with different skills in the project.

As applicative example we introduce the case of train-gate-controller presented in [40]. The problem is composed by three components: a train, a gate and a controller. A train approaches a gate from a great distance with a speed between 48 and 52 m/s. As soon as it passes the detector signal placed (1000 m) backward from the gate, an approaching signal is sent to the controller. The train may slow down (speed between 40 and 52 m/s) and pass the gate. As soon as it passes the detector placed (100 m) forward from the gate, an exit signal is sent to the controller. A new train may come after the current one has passed the second detector, but only at a security distance (1500 m). The gate is able to receive lower and raise signals from the controller at any time. As soon as the gate receives a lower signal, it lowers from 90° to 0° . As soon as it receives a raise signal, it raises from 0° to 90° . The controller is able to receive approaching and exit signals from the train detectors at any time. When the controller receives an approaching signal, it sends a lower signal to the gate. When it receives an exit signal, a raise signal is sent to the gate. Because of security procedure, an approaching signal should always cause the gate to go down, and exit signals should be ignored while the gate is going down. The train gate controller specifications have following environment variables:

- x for the distance of train from gate;
- r for the angle of gate with the ground (90° up, 0° down);
- d for possible delay controller;
- y for speed of the train.

It is assumed that initially there is no train at a distance smaller than 1400 m from the gate, the gate is open and the controller is idle. In transition labels, x, r, d denote values of variables before the transition. The primed variables x', r', d' represent values of variables in the new location after the transition. The problem can be represented with three graphs as in Figure 1.4. The complete system of commands to describe the system requires more than three pages and it can be found in [40]. In any case it should be evident that the need of different graphs (one for the trains, one for the signals and one for the infrastructure) make the procedure neither easy to define nor easy to understand.

In the last years many Process Algebras have been formulated, extended with data, time, mobility, probability and stochastic (e.g. [9]). Unluckily they have not always been satisfactory, because the presence of concurrency, communication, synchronization and nondeterminism makes the study of the correctness of concurrent systems particularly difficult, especially when the structure is not regular.

Markov Chain

Markov Chains are a special case of random processes which can be used to model

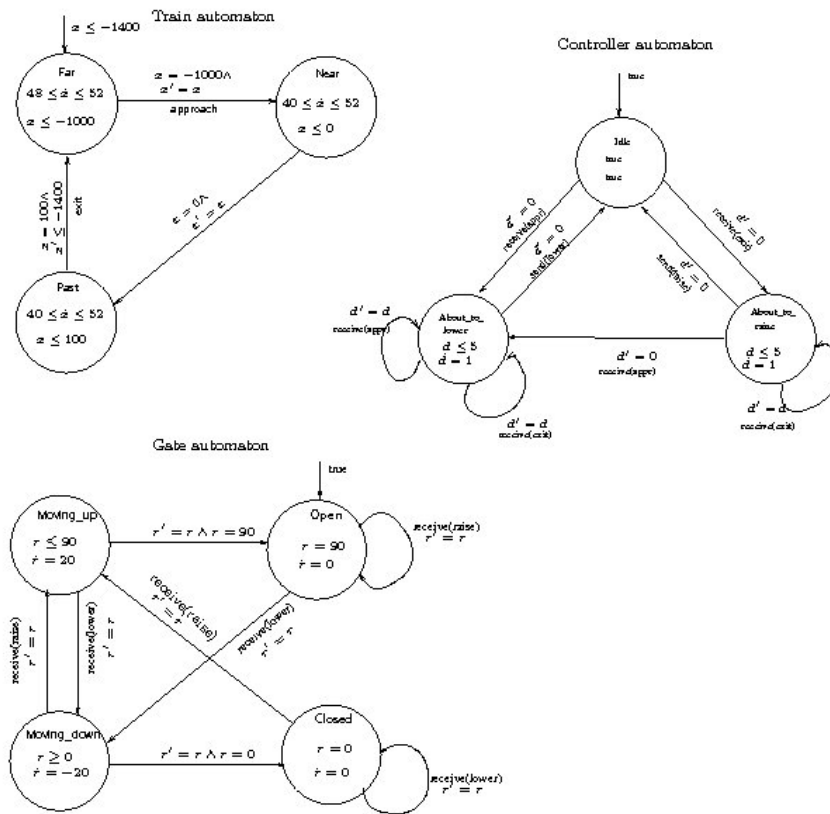


Figure 1.4: Train gate controller automaton (from [40])

various processes in queuing theory and statistics.

A random process is a collection of random variables X indexed by some set T taking values in some sets I .

- T is the index set, usually time (in the delay management $T = \mathbb{Z}$ but it can also be chosen as \mathbb{R} or \mathbb{R}^+);
- I is the state spaces (in the delay management $I = \mathbb{Z}^+$ but it can also be \mathbb{R} or $\{1 \dots n\}$ or $\{a, b, c\}$).

We classify random processes according to both the index set (discrete or continuous) and the state space (finite, countable or uncountable, continuous) A random process is called Markov Chain if conditioned on the current state of the process, its future is independent of its past. Mathematically we can write this property as

$$\mathbb{P}(X(t+1) = i_{t+1} | X(t) = i_t, \dots, X(1) = i_1, X(0) = i_0) = \mathbb{P}(X(t+1) = i_{t+1} | X(t) = i_t)$$

where $\mathbb{P}(A|B)$ is the conditional probability defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

The definition states that only the present state gives information on the future behavior of the process. Knowledge of the history of the process does not add any new

information.

The controlling factor in a Markov Chain is the *transition probability*, i.e. a conditional probability for the system to go to a particular new state given the current state of the system. This means that we get fairly efficient estimates if we can determine the proper transition probabilities.

A Markov Chain can be applied to predict the next state of a system, given the information of the system in the previous states. Under the assumption that the **Markov property** is valid, it is possible to neglect all the information coming from the history of the process except the most recent one. So the transition probability is defined as

$$p(i, j) = \mathbb{P}(X(t+1) = j | X(t) = i)$$

and it does not depend on the time t . Intuitively this value gives the “rules of the game” since it is the basic information needed to describe a Markov Chain and, due to the definition of probability, we must have

$$p(i, j) \leq 1 \quad \forall i, j \quad \text{and} \quad \sum_j p(i, j) = 1 \quad \forall i$$

A state j is said to be accessible from state i (written $i \rightarrow j$) if, given that we are in state i , there is a non-zero probability that at some time in the future, we will be in state j .

In the railway system, the variables $X_1, \dots, X_n \in \mathbb{Z}$ can represent the delays of the considered trains in the system. Then the Markov property can be explained as follow: the value of the delays of the trains at time t , $(x_{1,t}, \dots, x_{n,t})$, depends only on the previous measurement of their value, i.e. $(x_{1,t-1}, \dots, x_{n,t-1})$.

A recent application of Markov chains for the railway-timetable problem is [80], in which different distributions (and hence probability functions) are considered to evaluate the level of punctuality (e.g number of punctual trains) of the systems in case of delay.

Regression Model

Another approach often used in the railway delay management is the linear regression model (see [13] or [49]). Let’s consider a response variable, also called dependent variable, $Y \in \mathbb{R}$ (in our case the delay of an event), and some explanatory variables (independent variables) $X_1, \dots, X_{p-1} \in \mathbb{R}$ (delays of “previous” events).

The regression model tries to explain Y through a systematic component based on the X_i and an error ϵ to cover possible discrepancy:

$$Y = r(X_1, \dots, X_{p-1}) + \epsilon$$

As first assumption we suppose that the function $r(\cdot)$ is linear, so that we can rewrite the model as

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + \epsilon$$

If we suppose that we have more than one response variable Y_1, \dots, Y_n we get the system of equalities:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_{p-1} X_{i(p-1)} + \epsilon_i \quad \forall i = 1, \dots, n$$

where for each Y_i we can consider a different set of explanatory variables $X_{i1}, \dots, X_{ip-1} \in \mathbb{R}$. If we define the vector $Y = (Y_1, \dots, Y_n)^T$, the error vector $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$, the regression parameter $\beta = (\beta_0, \dots, \beta_{p-1})$ and the regression matrix

$$X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1n} \\ 1 & X_{21} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{(p-1)1} & X_{(p-1)2} & \dots & X_{(p-1)n} \end{pmatrix}$$

containing the values of the p explanatory variables, we can rewrite the model as $Y = \beta X + \epsilon$.

As second order hypothesis we consider (homoscedasticity):

$$\begin{aligned}\mathbb{E}(\epsilon) &= 0 \\ \text{Var}(\epsilon) &= \sigma^2 I_n \quad \exists \sigma^2 > 0 \text{ unknown} \\ \text{rank}(X) &= p\end{aligned}$$

so that

$$\mathbb{E}(Y) = \mu = \beta X \text{ and } \text{Var}(Y) = \sigma^2 I_n$$

If all that hold, we have defined a linear model that is particularly suited for the mathematical description of problems arising from controlled experiments, where experimenters can control the values taken by relevant factors to examine the corresponding values of the response variables. In this setting, X contains the value of the experimental factors, which are non-stochastic since they are chosen by the experimenters. The error term is due to measuring errors (which explains its name) and, if the instruments are not biased, it follows that $\mathbb{E}(\epsilon) = 0$. Finally, if the various experiments are conducted in such a way not to influence each other, then the stochastic independence assumption is satisfied, implying uncorrelated errors. To be able to define the probability distribution of Y , we introduce an alternative criterion to the Likelihood principle that would require additional hypothesis on the distribution of ϵ . We choose β such that it minimizes $\|Y - \mu\| = \sqrt{(Y - \beta X)^T (Y - \beta X)} = Q(\beta)$ (Least Squares Error). This method will be applied in Chapter 5 to evaluate the coefficient of the "virtual activities", i.e. the capacity constraints related to the outcome of the Tri-graph method (see Sections 2.11, 3.5.3 and 5.6).

In case the matrix X is regarded as non-stochastic, we can use the model to predict the values of Y . Consideration of non-stochastic X is supported by the following argument. In most common cases, the distribution of the explanatory variables does not contain any information on the relationship with the response variables, since we are interested in making interferences on this relationship not on the distribution of the explanatory variables. Therefore, we examine the variables conditionally on the values taken by X . In other words, we can operate within the conditional principle, which stipulates that interferences should be based not on the distribution itself but on the conditional distribution.

The main problem we found out in the application of this procedure is the choice of the variable X that should be considered in the definition of the model to evaluate the propagation of the delays. In fact this choice is strictly dependent on the knowledge of the set of dependencies of the system, that is the aim of this thesis. Therefore there is a really good interaction between Tri-graph method and linear regression: the first procedure identifies the dependencies, whereas the second one gives the specific "degree" of dependencies, i.e. the slope (see Chapter 5).

Wakob's Approach

Wakob has proposed an analytical framework for capacity assessment of railway stations which is based on queuing theory. More precisely, he applies queuing theory to predict the waiting time incurred by the simultaneous arrival and random processing of two trains as isolated part of the infrastructure.

Wakob's approach does not provide a queuing model for an entire railway station, but it proposes an analytical framework for capacity planning. It partitions the station into specific parts of the infrastructure (blocks, switches, platforms, ...) and it describes the performance of them as single element (single server) instead of the station as a whole.

Let a set C consists of basic infra-element s_1, s_2, \dots, s_p and assume that C carries the single server identity. If an arbitrary infra-element $s_i, i \in \{1, \dots, p\}$ of C is occupied

by one train, then the server identity implies that all other infra-elements inside C are blocked. Consequently C can not be used by any other train during the same time slot. So C can be seen as a "*Teilfahrstraßenknoten*" (TFK), i.e. as common parts of several routes. Wakob's Approach assumes that all the TFKs have infinite queuing space, to prevent a train from getting locked/blocked if a queue has reached its capacity, and that allows the approach to consider the TFKs separately. Since the size of a TFK is small compared to the length of a train, several TFK will be occupied simultaneously. Other assumptions of the method are Erlang distribution for the interarrivals processed at a TFK and for the service times, and random arrival order of the trains. The method evaluates the average time of a queue by the Pollackzek-Khintchine formula. In order to calculate the total amount of waiting time it evaluates the mean queue length by Little's formula, so that the total waiting time is computed by multiplying the expected queue length by the predefined observation period.

The approximation is very accurate but the waiting times are generally larger than those obtained via simulations. Moreover, it is a "timetable"-free approach, hence it cannot be compared with daily observations since a specific timetable can not be used to verify or falsify it. Therefore this method should only be adopted as a first approximation for the capacity assessment of railway stations. The studied cases indicate that the approach is indeed able to locate the bottlenecks section in a station. However, it seems to be rather uncomfortable for the practical use by railway staff due to the substantial efforts that are required to implement and to maintain the algorithms.

Analytical Description

It is not worthwhile to travel in the first class. You will arrive with the same delay of the second one.

Scrooge McDuck
fictional Scottish character created by Carl Barks

2.1 Formulation of the Problem

The scheduling of a timetable can be considered as a project in which a set of interacting tasks (journeys of the trains) require time (e.g. driving time, waiting time ...) and resources (e.g. tracks, platforms ...) to be completed.

Given two sets, \mathcal{T} for the trains and \mathcal{V} for the stations, that have to be studied, we represent the railway system by a network, the so called **Public Transportation Network** $PTN = (\mathcal{V}, \mathcal{B})$ in which every node represents a station and every edge is a set of (blocks of) tracks connecting two different stations (see an example in Figure 2.1).

The PTN is intuitive but the information it contains is not enough to study the problem from an analytical point of view. Accordingly we will instead consider the so called **Activity-on-arc Project Network** $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ (Ref: [51] and [54]).

We define a set of events \mathcal{E} corresponding to the arrivals and departures of all trains in all stations of their journeys, and a set of activities \mathcal{A} (driving along an edge, waiting in a station or connection between two trains) so that

$$(2.1) \quad \mathcal{E} = \mathcal{E}^{\text{dep}} \cup \mathcal{E}^{\text{arr}}$$

where

$$\begin{aligned} \mathcal{E}^{\text{dep}} &= \{ (t, v, \text{dep}) : t \in \mathcal{T} \ v \in \mathcal{V} : t \text{ departs from } v \} \\ \mathcal{E}^{\text{arr}} &= \{ (t, v, \text{arr}) : t \in \mathcal{T} \ v \in \mathcal{V} : t \text{ arrives in } v \} \end{aligned}$$

and

$$(2.2) \quad \mathcal{A} = \mathcal{A}^{\text{drive}} \cup \mathcal{A}^{\text{wait}} \cup \mathcal{A}^{\text{change}}$$

where

$$\begin{aligned} \mathcal{A}^{\text{drive}} &= \{ ((t, v, \text{dep}), (t, u, \text{arr})) \in \mathcal{E}^{\text{dep}} \times \mathcal{E}^{\text{arr}} : (v, u) \in \mathcal{B} \} \\ \mathcal{A}^{\text{wait}} &= \{ ((t, v, \text{arr}), (t, v, \text{dep})) \in \mathcal{E}^{\text{arr}} \times \mathcal{E}^{\text{dep}} \} \\ \mathcal{A}^{\text{change}} &\subseteq \{ ((t, v, \text{arr}), (t', v, \text{dep})) \in \mathcal{E}^{\text{arr}} \times \mathcal{E}^{\text{dep}} : t, t' \in \mathcal{T} \} \end{aligned}$$

and the set of guaranteed connections contained in $\mathcal{A}^{\text{change}}$ should be defined according to the passengers' needs.



Figure 2.1: Example of Public Transportation Network

The graph in Figure 2.2 is a small example of how an Activity-on-arc Project Network looks like when two connections between two trains are possible.

We define now the parameters

- w_i weight of event $i \in \mathcal{E}$
- π_i scheduled time of event $i \in \mathcal{E}$
- d_i source delay associated to event $i \in \mathcal{E}$, $d_i \geq 0$
- L_a minimal duration of activity $a \in \mathcal{A}$

and the variables

- y_i secondary delay of event $i \in \mathcal{E}$
- x_i re-scheduled time of event $i \in \mathcal{E}$

so that

$$(2.3) \quad x_i = \pi_i + y_i \quad \text{and} \quad y_i \geq d_i$$

If we consider just delays of the first type (propagation along the same train) and of the second type (propagation from one train to another due to connections), we can

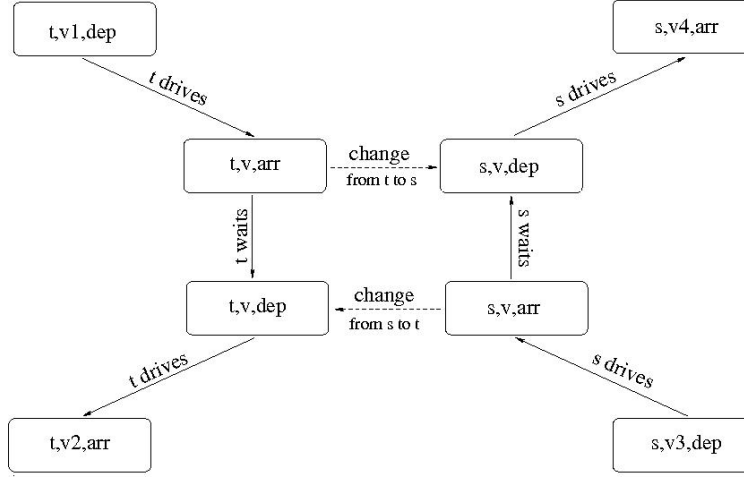


Figure 2.2: Example of Activity-on-arc Project Network

write our Timetable Model [TM-1] as:

$$(2.4a) \quad \min \sum_{i \in \mathcal{E}} w_i x_i$$

$$(2.4b) \quad x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A}$$

$$(2.4c) \quad x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E}$$

$$(2.4d) \quad x_i - \pi_i \leq T \quad \forall i \in \mathcal{E}$$

$$(2.4e) \quad x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E}$$

The objective Function (2.4a) is equivalent to the delay function $\sum_{i \in \mathcal{E}} w_i y_i$ since we defined $x_i = \pi_i + y_i$, where the scheduled timetable π_i is a constant parameter of our model.

The constraints represent the time limits of our problem:

- (2.4b) the real duration of an activity must respect the (technically) minimal one, i.e. the real duration must be larger than the given lower bound;
- (2.4c) the real timetable must respect the scheduled one and the delays;
- (2.4d) the delay of an event must be smaller than the period T of the model;
- (2.4e) the variables x_i are in \mathbb{Z}^+ since minutes (or seconds) are the minimal time units of the system.

Constraint (2.4d) is an (“implicit”) condition in a periodic timetable. If the delay of a train at one station y_i is greater than the period T ($y_i > T$), it is preferable (in order to avoid delay propagation) to cancel the train and ask the passengers to get on the next scheduled train. This condition gives an upper bound for delays, that can be interpreted as a deadline for every activity of the system. The Timetable Model can be read as a problem in which every activity has to be executed inside a time window (i.e. a time interval) defined by the scheduled timetable and the (“pre-defined”) deadline:

$$x_i \geq \pi_i + d_i \quad \text{and} \quad x_i - \pi_i \leq T$$

that is

$$x_i \in [\pi_i + d_i, \pi_i + T]$$

This is a broad interpretation of these “implicit” constraints since, inside a periodic timetable, every set of trains traveling on the same route has a specific period, which is usually smaller than the general period T . For example Hannover and Göttingen are connected by a train every hour, but the trains between the main station in Hannover and the Hannover airport have a higher frequency. Therefore it makes sense to consider, instead of the constant period T , a specific period T_i that depends on the route of train t corresponding to event i . We can rewrite the Timetable Model as [TM-2]

$$\begin{aligned}
(2.5a) \quad & \min \sum_{i \in \mathcal{E}} w_i x_i \\
(2.5b) \quad & x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A} \\
(2.5c) \quad & x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E} \\
(2.5d) \quad & x_i - \pi_i \leq T_i \quad \forall i \in \mathcal{E} \\
(2.5e) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E}
\end{aligned}$$

In case of its feasibility, [TM-2] can be solved by the critical path method (CPM) which looks for a longest path in the event-activity-network. In a connected network there always exists a longest path between two nodes if and only if it does not contain any direct cycle with positive length. We can assume the absence of direct cycles with positive length since the event-activity-network is a time-expanded network such that a cycle would represent a sequence of meaningless precedences.

Now we introduce in the [TM-2] the third type of delay (propagation from one train to another due to limited capacity of infrastructure), that is the one we mainly want to investigate. A possible way to proceed is to avoid any overlapping between two consecutive events. This can be interpreted as a capacity constraint since formally we forbid that two trains can use simultaneously the same track/platform.

To mathematically define these capacity constraints we could use the following Capacitated Timetable Model [CTM-1]

$$\begin{aligned}
(2.6a) \quad & \min \sum_{i \in \mathcal{E}} w_i x_i \\
(2.6b) \quad & x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A} \\
(2.6c) \quad & x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E} \\
(2.6d) \quad & x_i - \pi_i \leq T_i \quad \forall i \in \mathcal{E} \\
(2.6e) \quad & g_{ije}(x_j - x_i - L_a) \geq 0 \quad \forall i, j \in S_e \\
& \quad \text{where } a = (i, k) \in \mathcal{A}^{drive} \\
(2.6f) \quad & (1 - g_{ije})(x_i - x_j - L_{a'}) \geq 0 \quad \forall i, j \in S_e \\
& \quad \text{where } a' = (j, k') \in \mathcal{A}^{drive} \\
(2.6g) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E} \\
(2.6h) \quad & g_{ije} \in \{0, 1\} \quad \forall i, j \in \mathcal{E} \forall e
\end{aligned}$$

Here S_e refers to the set of (departure) events that use the same edge e of the underlying physical network in their next (driving) activity and g_{ije} is a binary variable equal to 1 if event $i \in S_e$ happens before event $j \in S_e$ on the edge e , 0 otherwise. The Constraints (2.6e) and (2.6f) are the capacity constraints on every edge: before a new (driving) activity starts on the edge, the previous scheduled one must have been completed (and to define the considered driving activities we introduced the corresponding arrival events k and k' of the two trains). These constraints are more restrictive than what we need. In practice, two trains can simultaneously use the same edge. To guarantee the safety while doing so, each edge is separated into blocks as introduced in

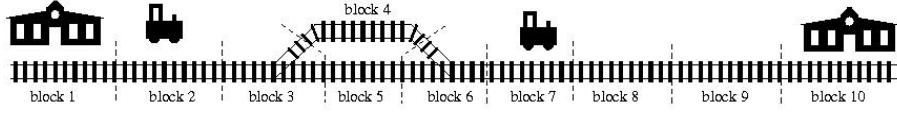


Figure 2.3: Blocks sections

Section 1.1.1, an example of blocks on a single track is shown in Figure 2.3. We hence rewrite [CTM-1] in a more detailed formulation. Firstly we define the set \mathcal{M} as the blocks (part of tracks or platforms) of the system. For every block $m \in \mathcal{M}$ we also define a set S_m of events in \mathcal{E}^{dep} that have to take place on it. Besides we introduce a new binary variable g_{ijm} equal to 1 if event $i \in S_m$ happens before event $j \in S_m$, zero otherwise, and a new parameter h_{ijm} , called headway, corresponding to the security distance that events i and j have to respect on block m . The Capacitated Timetable Model can be written as [CTM-2]

$$\begin{aligned}
 (2.7a) \quad & \min \sum_{i \in \mathcal{E}} w_i x_i \\
 (2.7b) \quad & x_j - x_i \geq L_a & \forall a = (i, j) \in \mathcal{A} \\
 (2.7c) \quad & x_i \geq \pi_i + d_i & \forall i \in \mathcal{E} \\
 (2.7d) \quad & x_i - \pi_i \leq T_i & \forall i \in \mathcal{E} \\
 (2.7e) \quad & g_{ijm}(x_j - x_i - h_{ijm}) \geq 0 & \forall m \in \mathcal{M} \forall i, j \in S_m \\
 (2.7f) \quad & (1 - g_{ijm})(x_i - x_j - h_{jim}) \geq 0 & \forall m \in \mathcal{M} \forall i, j \in S_m \\
 (2.7g) \quad & x_i \in \mathbb{Z}^+ & \forall i \in \mathcal{E} \\
 (2.7h) \quad & g_{ijm} \in \{0, 1\} & \forall m \in \mathcal{M} \forall i, j \in S_m
 \end{aligned}$$

Note that (2.7e) and (2.7f) can be replaced by

$$\begin{aligned}
 (2.7e') \quad & g_{ijm}(x_j - x_i - \bar{h}_{ijm}) \geq 0 & \forall m \in \mathcal{M} \forall i, j \in S_m \\
 (2.7f') \quad & (1 - g_{ijm})(x_i - x_j - \bar{h}_{jim}) \geq 0 & \forall m \in \mathcal{M} \forall i, j \in S_m
 \end{aligned}$$

where $\bar{h}_{ij} = \max_{\substack{m \in \mathcal{M} \\ i, j \in S_m}} h_{ijm}$. This choice reduces significantly the number of parameters of

the system since we do not need anymore to introduce the concept of block into the model. Hence we will get a formulation really similar to [CTM-1] but with sharper capacity constraints (i.e. (2.7e') and (2.7f')) instead of (2.6e) and (2.6f).

The model can be linearized or treated using disjunctive constraints (see the investigation in [60]).

To simplify a little bit the notation we introduce two new parameters

$$\tilde{\pi}_i = \pi_i + d_i \text{ and } \tilde{T}_i = T_i - d_i$$

so that [CTM-2] can be rewritten as [CTM-3]

$$\begin{aligned}
(2.9a) \quad & \min \sum_{i \in \mathcal{E}} w_i x_i \\
(2.9b) \quad & x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A} \\
(2.9c) \quad & x_i \geq \tilde{\pi}_i \quad \forall i \in \mathcal{E} \\
(2.9d) \quad & x_i - \tilde{\pi}_i \leq \tilde{T}_i \quad \forall i \in \mathcal{E} \\
(2.9e) \quad & g_{ijm}(x_j - x_i - h_{ijm}) \geq 0 \quad \forall m \in \mathcal{M} \forall i, j \in S_m \\
(2.9f) \quad & (1 - g_{ijm})(x_i - x_j - h_{jim}) \geq 0 \quad \forall m \in \mathcal{M} \forall i, j \in S_m \\
(2.9g) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E} \\
(2.9h) \quad & g_{ijm} \in \{0, 1\} \quad \forall m \in \mathcal{M} \forall i, j \in S_m
\end{aligned}$$

If there exists a source delay that is “big enough”, the solution $x_i = \tilde{\pi}_i \forall i \in \mathcal{E}$ is not any longer a feasible one.

Next we show that [CTM-3] is an \mathcal{NP} -hard problem.

2.2 \mathcal{NP} -completeness

Many mathematical problems are known to be \mathcal{NP} -complete and discovering that property on a model is usually the first step of a long work on the problems, even if that correspond also to a low priority in looking for an efficient, exact algorithm. The notation, the theorems and lemmas reported in this section, wove as introduction into \mathcal{NP} -completeness theory, are drawn from the book of Garey and Johnson [30].

The *time complexity function* for an algorithm expresses the largest amount of time it requires to solve a problem instance of a given size. Of course, this function is not well-defined until we do not fix the way to determine the input length of a problem and which computers are going to be applied. Nevertheless these choices have a small effect on the broad distinction made in the \mathcal{NP} -completeness theory.

To distinguish which algorithms are “efficient enough” and which are “too inefficient” is difficult, since it depends on the situation we want to consider. However computer scientists recognized a simple distinction between polynomial and exponential time algorithms that offers considerable insight into these matters.

A *polynomial time algorithm* is defined to be one whose time complexity function is $O(p(n))$ for some polynomial $p(n)$, i.e. there exists a constant c so that the complexity function $f(n)$ is bounded by $|f(n)| \leq c|p(n)|$ for all values of $n \geq 0$, where n is used to define the input length.

Any algorithm whose complexity function cannot be bounded in this way is called *exponential time algorithm*, although it should be noted that this class of algorithms includes also those whose complexity function is not normally regarded as an exponential function, e.g. $n^{\log(n)}$.

Most exponential time algorithms are merely variations on exhaustive search, whereas polynomial time algorithms are generally made possible through the gain of some deeper insight into the structure of the problem. There is a wide agreement that a problem has not been “well-solved” until a polynomial time algorithm is known for it. Hence we refer as *intractable* a problem so hard that no polynomial time algorithm is known. This is a rough approximation of the formal use of “intractable” due to the fact that, although exponential time algorithms are known for many problems, only few of them are regarded as being very useful in practice.

This definition of “intractable” turns out to be essentially independent from the particular encoding scheme and the computer model used for determining time complexity since it can be verified that standard schemes used in practice for any particular problem always seem to differ at most polynomially from each other. Similar comments

can be made for the choice of computer models.

The class of problems that can be solved with a polynomial time algorithm is defined as \mathcal{P} , while the problems whose solutions can be verified in polynomial time are referred to as \mathcal{NP} . An alternative informal definition of the \mathcal{NP} -class refers to “*all the problems that can be solved in polynomial time by a nondeterministic Turing machine*” (for a complete reference see [30]). A nondeterministic Turing machine is formed by parallel Turing machines that can take many computational paths simultaneously, without being able to communicate with each other. A Turing machine is a theoretical computing machine invented by Alex Turing (1937) to serve as an idealized model for mathematical calculations.

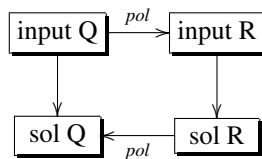
A problem belonging to the \mathcal{P} -class is always also an \mathcal{NP} -problem, i.e. $\mathcal{P} \subseteq \mathcal{NP}$.

It is an important open question to understand if $\mathcal{P} = \mathcal{NP}$. There is a theory developed by Cook ([22]) which provides strong evidence that $\mathcal{P} \neq \mathcal{NP}$.

Before we introduce the notion of \mathcal{NP} -completeness, it is necessary to point out two important concepts.

Firstly that, technically, the only problems which can be defined as \mathcal{NP} -complete are problems for which the answer is either yes or no. Thus an optimization problem must be rephrased as a feasibility problem (i.e. decision problem) in order to put it into the desired form. We may associate with each problem a decision problem by defining a threshold k for the corresponding objective function f . This decision problem is: does there exist a feasible solution S such that $f(S) \leq k$? When a problem is formulated as a decision problem there is an important asymmetry between those input whose output is “yes” and those whose output is “no”. A “yes” answer can be certified by a small amount of information, the feasible schedule S with $f(s) \leq k$. Given this certificate, the “yes” answer can be verified in polynomial time. This is not the case of the “no” answer.

Secondly the notion of reduction. Given two decision problems Q and R , we say that Q reduces to R (denoted by $Q \leq R$) if there exists a polynomial time function g that transforms inputs of Q into inputs of R such that x is a “yes”-input of Q if and only if $g(x)$ is a “yes”-input of R .



The importance of polynomial transformation comes from the following lemmas:

Lemma 2.1 [30] *If $Q \leq R$, then $R \in \mathcal{P}$ implies $Q \in \mathcal{P}$ (and equivalently $Q \notin \mathcal{P}$ implies $R \notin \mathcal{P}$).*

Lemma 2.2 [30] *If $Q \leq R$ and $R \leq S$, then $Q \leq S$.*

A problem Q is called \mathcal{NP} -complete if $Q \in \mathcal{NP}$ and for all $R \in \mathcal{NP}$ we have $R \leq Q$. Moreover the \mathcal{NP} -completeness means that any algorithm used to solve such problems in its full generality is likely to require exponential time in the worst case and thus is impractical except for relatively small input, where it is possible to implement it in such a way to find “rapidly” optimal solutions.

Therefore if a single \mathcal{NP} -complete problem could be solved in polynomial time, all

problems in \mathcal{NP} could be solved in polynomial time and we would have $\mathcal{P} = \mathcal{NP}$. The proof of \mathcal{NP} -completeness of a problem is divided into two parts. Firstly showing that the problem can be solved in polynomial time by a nondeterministic Turing machine. This is rather technical and therefore we are going to omit this part of the proof. The main part of the proof of \mathcal{NP} -completeness is the “reduction”. We must show that a known \mathcal{NP} -complete problem Q can be reduced, or transformed, into our problem R , i.e. given a specific input q for the problem Q we must show how to find a corresponding input r to problem R , such that the answer for r is “yes” if and only if the answer for q is “yes”. Moreover the input length of r and the time needed to generate it must be bounded by polynomial functions of the input length measured for problem Q . Summarizing to prove that new problems are \mathcal{NP} -complete we can refer to the following theorem:

Theorem 2.3 [30] *If Q and R belong to \mathcal{NP} , Q is \mathcal{NP} -complete and $Q \propto R$, then R is \mathcal{NP} -complete.*

We introduced the notion of \mathcal{NP} -completeness for decision problems, although usually we are not dealing with this kind of problems but rather with optimization problems, i.e. minimization/maximization problems with a specific objective function. An optimization problem is called \mathcal{NP} -hard if the corresponding decision problem is \mathcal{NP} -complete. Therefore as \mathcal{NP} -hard are defined those problems which can be solved with an algorithm that can be transformed into one for solving any other \mathcal{NP} -problem. Consequently \mathcal{NP} -hard problems are “at least as hard as any \mathcal{NP} -problem”, but they might be harder.

2.3 [CTM-3] is \mathcal{NP} -hard

To prove the \mathcal{NP} -hardness of [CTM-3] we introduce a well-known \mathcal{NP} -complete problem: *Sequencing within Intervals Problem* [SIP] (Ref: [30]). The [SIP] determines whether it is possible to find a feasible schedule for the tasks c of a finite set C , with a given duration l_c , in order to obey temporal restrictions (the execution has to happen inside a pre-defined time interval), with at most one task ever being executed at a time. A formal description of [SIP] is the following: given the durations of the activities l_c and temporal restrictions as minimal starting times p_c and deadlines t_c , we are looking for a function $\sigma : C \rightarrow \mathbb{Z}^+$ such that for each $c \in C$, $\sigma_c \geq p_c$, $\sigma_c + l_c \leq t_c$ and, if $c' \in C \setminus \{c\}$, then either $\sigma_{c'} + l_{c'} \leq \sigma_c$ or $\sigma_{c'} \geq \sigma_c + l_c$. This decision problem is \mathcal{NP} -complete, so every optimization problem associated with it results \mathcal{NP} -hard, no matter which objective function is chosen.

As objective function we consider the minimization of the weighted total flow time of the problem, i.e. the sum of the weighted completion times of the tasks in the set C . Accordingly to the definition of [SIP], we can write the problem as IP in such a way

to evidence the likeness with our formulation of [CTM-3].

$$\begin{aligned}
(2.10a) \quad & \min \sum_{c \in C} w_c \sigma_c \\
(2.10b) \quad & \sigma_c \geq p_c && \forall c, c' \in C \\
(2.10c) \quad & \sigma_c + l_c \leq t_c && \forall c, c' \in C \\
(2.10d) \quad & g_{cc'}(\sigma_c - \sigma_{c'} + l_c) \leq 0 && \forall c, c' \in C \\
(2.10e) \quad & (1 - g_{cc'}) (\sigma_{c'} - \sigma_c + l_{c'}) \leq 0 && \forall c, c' \in C \\
(2.10f) \quad & \sigma_c \in \mathbb{Z}^+ && \forall c \in C \\
(2.10g) \quad & g_{cc'} \in \{0, 1\} && \forall c, c' \in C
\end{aligned}$$

where the binary variable $g_{cc'} = 1$ if task c is executed before c' , 0 else. Consequently a task c is started at time σ_c , it is completed at time $\sigma_c + l_c$, it cannot be started before time p_c , it must be completed by time t_c , and its execution cannot overlap the execution of any other task c' .

We are now going to show that [SIP] can be rewritten as a special case of [CTM-3] and hence we will have that

Proposition 2.4 *Capacitated Timetable Model [CTM-3] is an \mathcal{NP} -hard problem.*

Proof: We show that [SIP] reduces to [CTM-3], i.e. [SIP] \propto [CTM-3].

Given an instance of [SIP] we interpret the tasks $c \in C$ as events $i \in \mathcal{E}$ of the problem [CTM-3] and obtain an instance of [CTM-3] by the following correspondences. We define $\mathcal{E} = \mathcal{E}^{dep} = C$ and also $S_m = C$ since we consider just one block, $\mathcal{M} = \{m\}$. Then we are able to define the parameters of [CTM-3] as follows.

$$\begin{aligned}
\tilde{\pi}_i &= p_i && \forall i \in \mathcal{E} \\
\tilde{T}_i &= (t_i - p_i) - l_i && \forall i \in \mathcal{E} \\
h_{ijm} &= l_i && \forall i, j \in \mathcal{E}
\end{aligned}$$

leaving the w_i as they are given in [SIP], where $T_i \geq 0$ otherwise the problem would be infeasible, since it would be required to complete a task in less than its minimal execution time. We also define $L_a = -\infty$ so that Constraint (2.9b) can be neglected from the model.

Now we have to show that x is a feasible timetable for [CTM-3] if and only if σ is feasible for [SIP].

x is a feasible solution of [CTM-3] if and only if x satisfies (2.9b), (2.9c), (2.9d), (2.9e) and (2.9f). According to the identification written above, we have that x is a feasible solution of [CTM-3] if and only

$$\begin{aligned}
x_i &\geq p_i && \forall i \in \mathcal{E} \\
x_i &\leq t_i - l_i && \forall i \in \mathcal{E} \\
g_{ijm}(x_j - x_i - l_i) &\geq 0 && \forall i, j \in \mathcal{E} \forall m \in M \\
(1 - g_{ijm})(x_i - x_j - l_i) &\geq 0 && \forall i, j \in \mathcal{E} \forall m \in M \\
x_i &\in \mathbb{Z}^+ && \forall i \in \mathcal{E} \\
g_{ijm} &\in \{0, 1\}
\end{aligned}$$

Considering the identity $x_i = \sigma_i$, these constraints coincide with (2.10b), (2.10c), (2.10d), (2.10e) and (2.10f). Hence it results that x is a feasible solution of [CTM-3] if and only if σ is a feasible solution of [SIP]. Furthermore, since for both problems we consider the same objective values (weighted total flow time) and the schedule x coincides with σ , we have that σ is an optimal solution of [SIP] if and only if x is an optimal solution of [CTM-3].

Hence [CTM-3] is \mathcal{NP} -hard. □

Although we proved that our problem is \mathcal{NP} -hard, we are not satisfied with this result since for the proof we referred to a timetable with specific periods T_i introduced in the Constraint (2.5d).

We remove now these parameters and consider the problem [CTM-4]

$$\begin{aligned}
(2.11a) \quad & \min \sum_{i \in \mathcal{E}} w_i x_i \\
(2.11b) \quad & x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A} \\
(2.11c) \quad & x_i \geq \tilde{\pi}_i \quad \forall i \in \mathcal{E} \\
(2.11d) \quad & g_{ijm}(x_j - x_i - h_{ijm}) \geq 0 \quad \forall m \in \mathcal{M} \forall i, j \in S_m \\
(2.11e) \quad & (1 - g_{ijm})(x_i - x_j - h_{jim}) \geq 0 \quad \forall m \in \mathcal{M} \forall i, j \in S_m \\
(2.11f) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E} \\
(2.11g) \quad & g_{ijm} \in \{0, 1\} \quad \forall m \in \mathcal{M} \forall i, j \in S_m
\end{aligned}$$

We are going to prove that [CTM-4] is \mathcal{NP} -hard by showing that it is a generalization of a Job Shop Scheduling problem.

2.4 Shop Scheduling Problem

According to the notation of [30] and [16], the **general Shop Scheduling Problem** may be defined as follows. We have m machines, $\mathcal{M} = \{M_1, \dots, M_m\}$, and n jobs, $j \in \{1, \dots, n\}$, such that each job j consists of a collection of operations $k \in \{1, \dots, n_j\}$. The set of all the operations corresponding to all the possible jobs is denoted as K and its cardinality is $N = \sum_{j=1}^n n_j$. Every operation is associated with a processing time

p_k and a specific machine $m_k \in \{M_1, \dots, M_m\}$ where it has to be processed. Each operation can be processed only by one machine and each machine can process only one operation at a time. Furthermore there may be defined some other restrictions like: a release time for the first operation of every job or a deadline for every job, pre-emption (or job splitting, i.e. job processing may be (several times) interrupted and resumed later on, either on the same machine or on another one), unitary processing time, precedence relations between the operations or batches (i.e. sets of jobs that must be processed jointly on one machine). We call a solution feasible if it satisfies all the restrictions used to define the scheduling problem.

The problem is to find a feasible schedule $\sigma : K \rightarrow \mathbb{Z}^+$ that minimizes some (regular) objective functions of the completion times of the jobs, $C_j = \sigma_{n_j} + p_{n_j}$, $j \in \{1, \dots, n\}$ (where n_j is the index of the last operation of job j).

The **Job Shop Scheduling problem** is a special case of the general Shop Scheduling in which every job j consists of an ordered sequence of operations (j_1, \dots, j_{n_j}) which must be executed in this order since there exist precedence constraints of the form $j_k \rightarrow j_{k+1}$.

It is also assumed that two consecutive operations of the same job j cannot be executed on the same machine, i.e. $m_{j_k} \neq m_{j_{k+1}}$ for $k \in \{1, \dots, n_j - 1\}$. Hence a schedule, that is an allocation of the operations to time intervals to machines, is a feasible solution for the job shop scheduling problem, if it respects all the precedence constraints between the operations, the condition on the changing of machines between two consecutive operations and the capacity constraints.

Again the problem is to find a feasible solution that minimizes some (regular) objective functions of the completion times of the job.

The majority of known results for the Job Shop Scheduling problems has been obtained for makespan minimization, i.e. $C_{max} = \max_{j=1 \dots n} C_j$ and for the minimization

of the total flow time, i.e. the minimization of the sum of the job completion times, $\sum_{j=1}^n C_j$. Nevertheless other objective function can be considered. For example if we introduce a due date d_j associated to every job $j \in J$, we can define for every job j its:

$$\begin{array}{ll}
 \text{lateness} & L_j = (C_j - d_j) \\
 \text{earliness} & E_j = \max(0, d_j - C_j) \\
 \text{tardiness} & T_j = \max(0, C_j - d_j) \\
 \text{absolute deviation} & D_j = |C_j - d_j| \\
 \text{squared deviation} & S_j = (C_j - d_j)^2 \\
 \text{unit penalty} & U_j = \begin{cases} 1 & \text{if } C_j - d_j \geq 0 \\ 0 & \text{else.} \end{cases}
 \end{array}$$

With each of these functions f_j we can get four possible objective functions: $\max_{j \in J} f_j$, $\max_{j \in J} w_j f_j$, $\sum_{j \in J} f_j$ and $\sum_{j \in J} w_j f_j$, where the parameters w_j are weight linked to every job j . According to [16], the most commonly used objective functions are:

- makespan $C_{max} = \max_{j=1..n} C_j$;
- total flow time $\sum_{j=1}^n C_j$;
- weighted total flow time $\sum_{j=1}^n w_j C_j$;
- maximum lateness $L_{max} = \max_{j=1..n} L_j$;
- total tardiness $\sum_{j=1}^n T_j$;
- weighted total tardiness $\sum_{j=1}^n w_j T_j$;
- total unit penalty $\sum_{j=1}^n U_j$;
- weighted total unit penalty $\sum_{j=1}^n w_j U_j$

From here forth we extend the concept of reduction to objective functions meaning that an objective function reduces to another if the Job Shop Scheduling problem with one of these objective functions reduces to a Job Shop Scheduling problem with the other objective function. Hence $\sum_{j=1}^n f_j$ reduces to $\sum_{i=1}^n w_j f_j$ by setting $w_j = 1 \forall j$ and $C_{max}, \sum_{j=1}^n C_j, \sum_{j=1}^n w_j C_j$ reduce to $L_{max}, \sum_{j=1}^n T_j, \sum_{j=1}^n w_j T_j$ respectively by setting $d_j = 0 \forall j$.

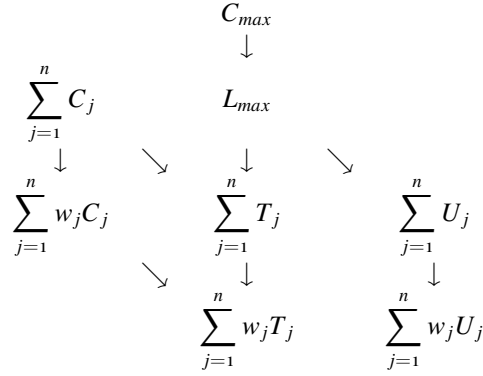


Table 2.1: Reductions among objective functions

Furthermore we have that L_{max} reduces to $\sum_{j=1}^n T_j$ and $\sum_{j=1}^n U_j$:

$$\begin{aligned}
L_{max} \leq k &\Leftrightarrow \max_{j=1..n} (C_j - d_j) \leq k \\
&\Leftrightarrow C_j - d_j \leq k \quad \forall j \\
&\Leftrightarrow C_j - (d_j + k) \leq 0 \quad \forall j \\
&\Leftrightarrow \max\{0, C_j - (d_j + k)\} \leq 0 \quad \forall j
\end{aligned}$$

Since the maximum between zero and a quantity x is equal to zero ($\max\{0, x\} = 0$) if and only if the quantity x is not positive ($x \leq 0$). We can write

$$L_{max} \leq k \Leftrightarrow \sum_{j=1}^n \max\{0, C_j - (d_j + k)\} \leq 0$$

Considering the new deadlines $d'_j = d_j + k \forall j$, we get a new definition of the tardiness as $T'_j = \max\{0, C_j - (d'_j)\} \forall j$, hence

$$\begin{aligned}
L_{max} \leq k &\Leftrightarrow \sum_{j=1}^n T'_j \leq 0 \\
&\Leftrightarrow \sum_{j=1}^n U'_j \leq 0.
\end{aligned}$$

As drawn in [16] these reductions can be represented graphically as in Figure 2.1. where an arrow $A \rightarrow B$ indicates a reduction of A to B . From now on we will use the symbol \rightarrow to indicate the reduction of a single function (or constraint) to another one. To indicate the reduction of a problem to another one we will consider the symbol \Leftarrow as done in Section 2.2.

Therefore if it has been proved that a problem is \mathcal{NP} hard given one of these objective functions, then the problem is also \mathcal{NP} hard given all the other objective functions that can be reached in the graph moving from the given objective function following the direction of the arrows.

For our specific timetable problem we introduce further objective functions:

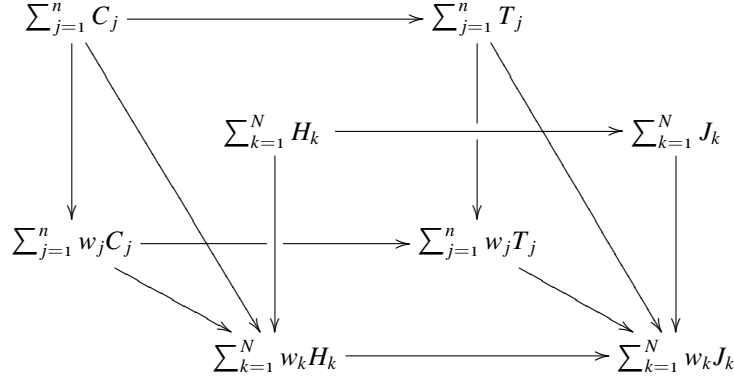


Table 2.2: Particular addition to Table 2.1

- total task-flow time $\sum_{k=1}^N H_k$ where H_k is the completion time of task $k \in K$;
- weighted total task-flow time $\sum_{k=1}^N w_k H_k$;
- total task-tardiness $\sum_{k=1}^N J_k$ where $J_k = \max\{0, H_k - d_k\}$ is the tardiness of task $k \in K$ given a deadline d_k ;
- weighted total task-tardiness $\sum_{k=1}^N w_k J_k$.

Likewise we have that $\sum_{k=1}^N f_k$ reduces to $\sum_{k=1}^N w_k f_k$ by setting $w_k = 1 \forall k \forall j$ and $\sum_{k=1}^N H_k$, $\sum_{k=1}^N w_k H_k$ reduce respectively to $\sum_{k=1}^N J_k$, $\sum_{k=1}^N w_k J_k$ by setting $d_k = 0 \forall k \forall j$. Moreover we have that $\sum_{j=1}^n C_j$ and $\sum_{j=1}^n T_j$ reduce respectively to $\sum_{k=1}^N w_k H_k$ and $\sum_{k=1}^N w_k J_k$ by setting $w_k = 1$ if $k = n_j$ for some $j \in J$, and $w_k = 0$ otherwise. Similarly $\sum_{j=1}^n w_j C_j$ and $\sum_{j=1}^n w_j T_j$ reduce respectively to $\sum_{k=1}^N w_k H_k$ and $\sum_{k=1}^N w_k J_k$ by setting $w_k = 0$ for all $k \neq n_j$ and letting w_{n_j} as given.

These reductions are depicted in Figure 2.2.

Similarly there are reductions between the constraints of the problem as shown in [16]. In Table 2.3, we represent some of these reductions, where \circ indicates the absence of the restriction. In the first column of Table 2.3 it is written that if the JSS problem is \mathcal{NP} -complete in case of 2 machines, then it will be \mathcal{NP} -complete also in case of an arbitrary number m of machines and hence in the case where the number of machines of the problem is not specified. While the second column means that if the problem is \mathcal{NP} -complete in absence of precedence constraints, then it will also be \mathcal{NP} -complete

machine	precedence	release time	deadline	processing time	nr. of jobs
2	○	○	○	$p_k = 1$	2
↓	↓	↓	↓	↓	↓
3	prec	r_j	d_j	$p_k = p$	3
↓				↓	↓
m				○	n
↓					↓
○					○

Table 2.3: Reductions among constraints of the Job Shop Scheduling problem

if we define some precedence among the jobs. The other columns can be analogously interpreted.

2.5 [CTM-4] is \mathcal{NP} -hard

We want to prove that [CTM-4] is \mathcal{NP} -hard, therefore we introduce the planning version of [CTM-4], denoted by [CTM-4p], which is obtained by replacing Constraints (2.11c) by

$$(2.11c') \quad x_i \geq 0$$

Since (2.11c') is a special case of (2.11c), where the feasibility of π does not matter since d_i can be chosen arbitrarily, we obtain that [CTM-4p] is a special case of [CTM], hence the \mathcal{NP} -hardness of [CTM-4p] yields the \mathcal{NP} -hardness of [CTM-4]. Accordingly, our goal is to show that [CTM-4p] is \mathcal{NP} -hard.

To this end we consider the Job Shop Scheduling problem defined by m machines, n jobs with an arbitrary number of operations ($k \in K$) where each of them associated to a proceeding time (p_k) and a machine (m_k). As in the previous section we define N as the total number of operations of the problem and n_j as the index of the last operation of job j .

Furthermore, let a set of precedence constraints $\mathcal{A}^{prec} \subseteq K \times K$ be given, where a pair $(i, j) \in \mathcal{A}^{prec}$ (with operations i and j belonging to different jobs) corresponds to the requirement that i has to be processed before j is allowed to start. The goal is to find a schedule $\sigma : K \rightarrow \mathbb{Z}^+$ so that all precedence constraints are satisfied and the *total flow time*, i.e. the sum $\sum_{j \in J} C_j = \sum_{j \in J} \sigma_{n_j} + p_{n_j}$ of all completion times of all jobs is minimized. We denote this problem as $Jm|prec|\sum_{j=1}^n C_j$ as usually done in literature.

By setting $w_k = 1$ if $k = n_j$ and $w_k = 0$ else, we can reduce $Jm|prec|\sum_{j=1}^n C_j$ to a Job Shop Scheduling problem with a new objective function $\sum_{k \in K} w_k H_k = \sum_{k \in K} w_k (\sigma_k + p_k)$, i.e. we minimize the sum of completion times over all operations (*total task flow time*).

Proposition 2.5 *The $Jm|prec|\sum_{k=1}^N w_k H_k$ reduces to the planning version of Capacitated Timetable Model [CTM-4p].*

Proof: We reduce [CTM-4p] to the job shop scheduling problem with total task flow time. We choose the blocks $\mathcal{M} = \{M_1, \dots, M_m\}$ as the machines of the Job Shop Scheduling problem, define the set of trains as $\mathcal{T} := J$, and add a departure and an

arrival event for each operation, i.e.

$$\begin{aligned}\mathcal{E}^{dep} &:= \{t_k^{dep} : k \in K\} \\ \mathcal{E}^{arr} &:= \{t_k^{arr} : k \in K\}\end{aligned}$$

For each job $j \in J$ consisting of operations $(j_1, j_2, \dots, j_{n_j})$ we define $\mathcal{A}^{drive} := (j_k^{dep}, j_k^{arr})$ for $k = 1, \dots, n_j$ and $\mathcal{A}^{wait} := (j_k^{arr}, j_{k+1}^{dep})$ for $k = 1, \dots, n_j - 1$. Finally, $\mathcal{A}^{change} := \mathcal{A}^{prec}$. Then we can define the parameters of [CTM-p] as

$$\begin{aligned}L_a &= p_i \text{ for all } a = (i, j) \in \mathcal{A}^{drive} \cup \mathcal{A}^{change} \\ L_a &= 0 \text{ for all } a = (i, j) \in \mathcal{A}^{wait}, \text{ and} \\ h_{ijm} &= p_i \text{ for all } i, j \in \mathcal{E}^{dep} \text{ with } m_i = m_j = m,\end{aligned}$$

where m_i and m_j are the machines on which operations i and j are scheduled. We leave w_i as they are given in $Jm|prec| \sum_{k \in K} w_k H_k$. The result of [CTM-4p] then is a timetable x , from which we obtain $\sigma : N \rightarrow \mathbb{Z}^+$ by the identity $\sigma(i) = x_i$.

For the correctness, we show that x is an optimal timetable for [CTM-4p] if and only if σ is optimum for $Jm|prec| \sum_{k=1}^N w_k H_k$.

x is a feasible solution of [CTM-4p] if and only if x satisfies (2.11b), (2.11c'), (2.11d) and (2.11e). According to the definition above it results that x is a feasible solution of [CTM-4p] if and only if

$$\begin{aligned}x_j - x_i &\geq p_i \quad \forall a = (i, j) \in \mathcal{A}^{drive} \cup \mathcal{A}^{change} \\ g_{ijm}(x_j - x_i - p_i) &\geq 0 \quad \forall m \in \mathcal{M} \forall i, j \in S_m \\ (1 - g_{ijm})(x_i - x_j - p_j) &\geq 0 \quad \forall m \in \mathcal{M} \forall i, j \in S_m \\ x_i &\in \mathbb{Z}^+ \quad \forall i \in \mathcal{E}\end{aligned}$$

Due to the identity $\sigma(i) = x_i$, we have that x is a feasible solution of [CTM-4p] if and only if σ is a feasible solution of $Jm|prec| \sum_{k=1}^N w_k H_k$.

Furthermore, since the objective function of both the problems is the weighted total flow time, and since the schedule x coincides with σ , we can conclude that σ is an optimal solution of $Jm|prec| \sum_{k=1}^N w_k H_k$ if and only if x is an optimal solution of [CTM-4p].

□

2.6 Capacity-Constrained Problem

In the previous section we showed that our model [CTM-4] is a generalization of a Job Shop Scheduling problem with precedence constraints and release times. In such a problem the capacity constraints are represented by inequalities in the form

$$(2.12) \quad g_{ijm}(x_j - x_i - p_i) \geq 0$$

that can be read as: before a new task j can be processed on one machine m the previous task i on it must be completed. The precedence is given by the binary variables g_{ijm} as explained in the previous sections.

This is one of the many approaches that can be applied to represent the capacity constraints. Here we want to introduce another way to deal with these constraints: the Resource Constrained Project Scheduling Problem (RCPS).

In particular we will show that the RCPS, in the case of the railway system, can be seen as a generalization of the Job Shop Scheduling problem.

Resource Constrained Project Scheduling Problem (RCPS) and its application to train timetables

RCPS involves assigning tasks of a project to a resource or a set of resources with limited capacity, in order to meet some predefined objectives. We suppose that every task has just one execution mode and both its duration and its resources requirements are assumed to be fixed. We also require to complete a task once it is started on a machine. In the case of the railway system, the tasks are the trains traveling into the system which have to be assigned to blocks of tracks and platforms in the railway network in order to guaranteed the desired trips and connections. We suppose that the sequence of blocks that a train will use is determined and all the traveling times are known. We also suppose that a train will receive a red light on a track outside a station only to prevent dangerous situation (i.e. no preemption is allowed).

Hence the project can be represented as a graph where the nodes correspond to the events of the trains \mathcal{E} and the arcs to the precedence relationships among them. If an arc $a = (i, j)$ appears in the graph, then event i must be completed prior to performing event j .

Resources are characterized in terms of their availability (capacity) to perform tasks in each time periods of the problem and the maximal resource availability is constant. Thus we can define for every block of the system $m \in \mathcal{M}$ a maximal capacity k_m , so that the total consumption of this resource in every period of time τ satisfies the condition

$$\sum_{i \in \mathcal{E}} r_{im\tau} \leq k_m \quad \forall m \in \mathcal{M} \text{ and } \forall \tau$$

where

$$r_{im\tau} = \begin{cases} 1 & \text{if event } i \text{ uses block } m \text{ at time } \tau \\ 0 & \text{else} \end{cases}$$

represents the consumption of the resource m at time τ due to event i .

Proposition 2.6 *The Job Shop Scheduling problem is a special case of the Resource Constrained Project Scheduling Problem.*

Proof: We consider a special case of RCPS in which every block has a capacity of one, i.e. $k_m = 1 \quad \forall m \in \mathcal{M}$. In other words, we are claiming that just one task at a time can use a machine. Furthermore we define c_m as the necessary time for a task to be completed on machine m (plus the time to set up the machine for the next task). The next task that can take place on the machine m cannot start before c_{ijm} units of time from the beginning of the previous task on the machine, i.e.

$$x_j - x_i \geq c_{ijm}$$

If we identify c_{ijm} with p_i and we fix the priority order so that j has to be executed before i (i.e. $g_{ijm} = 1$), we get exactly the Job Shop Scheduling Constraint (2.12). \square

Remark 2.7 In the case of the railway system this corresponds to consider as block the single tracks/platforms of the railway system, so that, due to security rules, only one train at a time can be on at it. c_m is defined as the necessary time for a train to pass through the block m (plus the “security distance” time). The next event that can take place on the block m cannot start before c_{ijm} units of time from the beginning of the previous event on the block.

In Section 2.8 we are going to present a possible formulation of the renewable resource constraints for the Capacitated Timetable Model under the RCPS method. Meanwhile we present another possible formulation of the problem that has been presented in [71] in order to compare our notation with other possible ones.

2.7 Train traffic deviation handling using TS and SA

The timetable problem has been studied in detail by a research group of the Blekinge Institute of Technology, Karlshamm, Sweden (see [72]). Their formulation includes many aspects that have not been considered in the described [CTM-4] model. For example they define a set of constraints to decide if a connection should be kept or not. Therefore we prefer here to present the “Swedish” model in its original form and to compare afterwards this formulation with the one introduced in the previous sections to highlight the similarities.

We report here the legend of the “Swedish” formulation as explained in [72].

Sets:

- T = set of trains;
- B = set of blocks;
- K_i = set of n_i events for train i , where $i \in T$ so $K_i = \{1 \dots n_i\}$;
- L_j = set of m_j events for block j , where $j \in B$ so $L_j = \{1 \dots m_j\}$.

Parameters:

the index i is associated with a train ($i \in T$), while j is the index for the blocks ($j \in B$). k denotes the index of the event of a train ($k \in K_i$) and l the events of a block ($l \in L_j$). Thus the same event can belong both to a set K_i and a set L_j .

- g_j^{block} = minimum time separation between occupation of train associated with block j ;
- c_{ik}^{delay} = penalty per time unit for delays for event k and train i ;
- d_{ik} = minimum running time for event k of train i ;
- $b_{ik}^{initial}$ = initial start of event k of train i according to the timetable;
- $e_{ik}^{initial}$ = initial end of event k of train i according to the timetable;
- $b_{ik}^{earliest}$ = earliest start (due to imposed delay) of event k of train i ;
- $e_{ik}^{earliest}$ = earliest end of event k of train i ;
- a_{jl}^{train} = train occupying block j during event l on the block;
- a_{jl}^{event} = corresponding index of event l of block j within the list of events of train a_{jl}^{train} ;
- h_{ik} = binary value equal to 1 if the events k for train i starts from a station where the start time is fixed, else 0;
- f_{ik}^{con} = binary value equal to 1 if the events k for train i has a connecting train that should wait, else 0;
- f_{ik}^{train} = specifies the connecting train of event k for train i if $f_{ik}^{con} = 1$ else 0;
- f_{ik}^{event} = index of the connecting event of train i and event k if $f_{ik}^{con} = 1$ else 0;
- f_{ik}^{gap} = specifies min time difference between end of event k of train i and its connecting event if $f_{ik}^{con} = 1$ else 0;
- c_{ik}^{con} = fixed penalty cost for missing the connection of event k for train i if $f_{ik}^{con} = 1$ else 0;
- c_{ik}^{con2} = penalty cost per minute for missing the connection of event k for train i if $f_{ik}^{con} = 1$ else 0;
- K = constant with a large positive value that covers the time period of the optimization problem.

Variables:

- x_{ik}^{begin} = start time of event k for train i ;
- x_{ik}^{end} = end time of event k for train i ;
- z_{ik} = magnitude of delay for event k for train i ;
- y_{ik}^{con} = time difference between end of event k for train i plus f_{ik}^{gap} and start of its connecting event;
- u_{ik}^{con} = binary variable, specifies if event k for train i has missed its connecting event, equal to 1 if $y_{ik}^{con} > 0$, else 0;
- $s_{jl\nu}$ = binary variable, specifies if event l on block j occurs before or after event $l + \nu$, where $\nu \in \{1 \dots (m_j - j)\}$, i.e. equal to 1 if event l happens before event $l + \nu$.

The “Swedish” model considers two possible different objective functions:

$$(2.13) \quad \min \sum_{i \in T} z_{i,n_i}$$

to minimize the total delay of the traffic system (i.e. the sum of the final delays of the trains when they arrive at their final destinations)

and

$$(2.13') \quad \min \sum_{i \in T} \sum_{k \in K_i} [c_{ik}^{delay} z_{ik} + c_{ik}^{con} u_{ik}^{con}]$$

to minimize the total cost calculated assigning a cost to every minute of delay of every single train and a penalty for every missing connection.
Just one of the two objective functions is considered at a time.

Subject to the constraints:

$$(2.14a) \quad x_{ik}^{end} \leq x_{ik+1}^{begin} \quad i \in T, k \in \{1 \dots (n_i - 1)\}$$

$$(2.14b) \quad x_{ik}^{end} = x_{ik}^{begin} + d_{ik} \quad i \in T, k \in K_i$$

$$(2.14c) \quad x_{a_{j,l+v}^{train} a_{j,l+v}^{event}}^{begin} - x_{a_{jl}^{train} a_{jl}^{event}}^{end} \geq g_j^{block} s_{jlv} - K(1 - s_{jlv}) \quad j \in B, l \in \{1 \dots (m_j - 1)\} \\ v \in \{1 \dots (m_j - l)\}$$

$$(2.14d) \quad x_{a_{j,l+v}^{train} a_{j,l+v}^{event}}^{begin} - x_{a_{jl}^{train} a_{jl}^{event}}^{end} \leq K s_{jlv} - (1 - s_{jlv}) \cdot \\ \cdot (g_j^{block} + d_{a_{jl}^{train} a_{jl}^{event}} + d_{a_{j,l+v}^{train} a_{j,l+v}^{event}}) \quad j \in B, l \in \{1 \dots (m_j - 1)\} \\ v \in \{1 \dots (m_j - l)\} \quad i \in T, \\ k \in K_i : h_{ik} = 1$$

$$(2.14e) \quad x_{ik}^{begin} \geq b_{ik}^{initial} \quad i \in T, k \in K_i$$

$$(2.14f) \quad x_{ik}^{begin} \geq b_{ik}^{earliest} \quad i \in T, k \in K_i$$

$$(2.14g) \quad x_{ik}^{end} \geq e_{ik}^{earliest} \quad i \in T, k \in K_i$$

$$(2.14h) \quad x_{ik}^{end} - e_{ik}^{initial} \leq z_{ik} \quad i \in T, k \in K_i : f_{ik}^{con} = 1$$

$$(2.14i) \quad x_{f_{jk}^{train} f_{jk}^{event}}^{begin} - x_{ik}^{end} + y_{ik}^{con} \geq f_{jk}^{gap} \quad i \in T, k \in K_i : f_{ik}^{con} = 1$$

$$(2.14j) \quad y_{ik}^{con} \leq K u_{ik}^{con} \quad i \in T, k \in K_i$$

$$(2.14k) \quad x_{ik}^{begin}, x_{ik}^{end}, z_{ik}, y_{ik}^{con} \geq 0 \quad i \in T, k \in K_i$$

$$(2.14l) \quad u_{ik}^{con} \text{ binary} \quad i \in T, k \in K_i$$

$$(2.14m) \quad s_{ikv} \text{ binary} \quad j \in B, l \in \{1 \dots (m_j - 1)\} \\ v \in \{1 \dots (m_j - l)\}$$

A short explanation of the constraints can be given as:

(2.14a) every event must be completed before the next event in the train list starts;

(2.14b) every event ends after its run time;

(2.14c) minimum time separation between event l and the successive event $l + v$ to use the same block in case l happens before $l + v$;

(2.14d) the reverse of (2.14c) can be applied in the opposite case: $l + v$ happens before l ;

(2.14e) the scheduled timetable for start activities must be respected;

(2.14f) if there is a delay the new departure time can not be scheduled before the earliest possible start;

- (2.14g) the scheduled timetable for end activities must be respected;
- (2.14h) the magnitude of the delay for every event is recordable by variable z_{ik} ;
- (2.14i) specify how large is the gap between two trains to penalize missed connections;
- (2.14j) activation of the binary variables.

We can compare this model with the one presented in Section 2.1 if we introduce a new constant M (that is a number “big enough”), a new parameter b_M for the capacity constraints, and a new binary variable z_a for the missing connections ($z_a = 1$ if the connection is kept, else 0), then

- (2.13) is equivalent to (2.4a);
- (2.14a) and (2.14b) correspond to (2.4b);
- (2.14c) and (2.14d) can be rewritten as $|x_j - x_i| \geq b_M$;
- (2.14e), (2.14f) and (2.14g) correspond to (2.4c);
- (2.14h) is equivalent to $x_i - \pi_i \leq y_i$;
- (2.14i) and (2.14j) can be rewritten as $x_j - x_i - Mz_a \geq L_a$;

Constraint $|x_j - x_i| \geq b_M$ can be linearized by introducing a boolean variable z_{ij} defines as

$$z_{ij} = \begin{cases} 0 & \text{if } x_j - x_i \geq b_M \\ 1 & \text{else} \end{cases}$$

and a new constant M_1 so that

$$\begin{aligned} z_{ij}(x_j - x_i - b_M) &= 0 \\ z_{ij} + z_{ji} &= 1 \end{aligned}$$

We conclude that the two models can describe the same aspects of the problem and that the main difference between them is the choice of the indices of the variables. The Swedish model considers different sets of indices hence it can describe more in details the constraints of the problem. But this choice results in a heavier formulation which is not easily readable. This is one of the reasons we prefer to adopt the model presented in Section 2.1 even if in an simplified version. Another one is that we do not want to consider the optimization problem itself, but the interaction it can have with the stochastic methods that will be introduced in Chapter 3.

For the record we write hereafter the linearization of the Swedish Mix Integer Linear Programming (MILP) model that is applied during the experiments. It is based on the replacement of the fixed cost for a missing connection by a dynamic one c_{ik}^{con2} so that (2.13') can be rewritten as

$$(2.13'') \quad \min \sum_{i \in T} \sum_{k \in K_i} [c_{ik}^{delay} z_{ik} + c_{ik}^{con2} y_{ik}]$$

and on the assignment of fixed values to the binary variables s_{ikv} so that Constraints (2.14c) and (2.14d) can be rewritten together in a single inequality

$$(2.14c') \quad x_{a_{jl}^{rain} a_{jl}^{event}}^{end} + g_j^{block} \leq x_{a_{j,l+1}^{rain} a_{j,l+1}^{event}}^{begin} \quad j \in B$$

$$l \in \{1 \dots (m_j - 1)\}$$

The problem to be solved is:

Minimize: (2.13)

or

(2.13'')

Subject to: (2.14a),(2.14b),(2.14c'), (2.14e)-(2.14i) and (2.14k)

The Linear Programming is dealt with an iterative two-level process. The lower bound is to allocate the start and end times for each train and the blocks it will occupy according to a fixed order of trains for each block (Linear Programming LP). The upper bound is to determine the order of the trains on the blocks, which is carried out by using two heuristics, either Tabu Search (TS) or Simulated Annealing (SA). Hence, the IP model calculates the effects of the modifications carried out by the heuristic(s). The general procedure, independent of which heuristic is used, is as follow: given a disturbance, a program calculates if the disturbance has any impact, and if so, it calls a method that clears all events that ended before the disturbed event occurred. All events that started but not ended before the disturbance appeared are assigned the value of their initial starting time to their earliest starting time, except the one that was disturbed, which gets a value according to the delay. Then one of the two heuristics is used to create and evaluate different solutions by making swaps of neighboring events for the block (one at a time).

Not all the swaps need to be considered since some of them may be infeasible. The feasibility of a swap is checked by a specific algorithm. Swapping two trains that will meet is only possible if the order of trains in the other blocks allows it. The neighborhood used in TS and SA is then further reduced by considering only the swap that satisfy also a constraint on the dual variable of (2.14c'). This dual variable indicates how much we could gain by letting two trains occupy the same block simultaneously for one additional unit of time. Thus, the only moves allowed are feasible moves which correspond to a potential reduction of the objective function value.

The swaps are carried out by switching places between two events in the event list vector of a block. After each swap, the LP is run in order to optimize the start and end times of each event. The values of the objective function used are compared to the best value found so far, and the process is repeated until a stopping condition is satisfied (depending on which heuristic is applied).

2.8 Resource Constrained Project Scheduling Approach

The train scheduling problem with capacity constraints can be formulated according to the project planning with resource conflicts model presented in the article "Project Scheduling with inventory constraints" of Neumann-Schwindt ([52]). A train in fact can be seen as a project that to be completed, i.e. to reach the final destination of its journey, has to be processed in a sequel of machines (stations connected by routes in the network), which have limited capacity (limited number of platforms/tracks) as described in Section 2.5.

We define for every train the path (corresponding to tracks and platforms) that it is going to use. We introduce a set E of resources which elements are the edges and nodes of the PTN, so $E = V \cup B$. For every element in E we assign a capacity \bar{R} corresponding to the number of trains that can use the same resource simultaneously.

$$\bar{R} : E \rightarrow \mathbb{Z}$$

$$\bar{R}^e = \begin{cases} 1 & \text{single track} \\ 2 & \text{double tracks} \\ \dots & \end{cases}$$

Due to mathematical reasons, we also define a lower bound for the capacity

$$\underline{R}_e = 0 \quad \forall e \in E$$

For every activity in $\mathcal{A}' = \mathcal{A}^{\text{drive}} \cup \mathcal{A}^{\text{wait}}$ we define a function

$$(2.16) \quad e : \mathcal{A}' \rightarrow E$$

that indicates which resource $e \in E$ is used to perform the activity $a \in \mathcal{A}'$.

It is also necessary to define a function that expresses the level of "consumption" of a resource and this is done through another function that gives the status of a train with respect to a particular resource.

$$(2.17) \quad r : \mathcal{E} \times E \rightarrow \{-1, 0, 1\}$$

$$r_{ie} = \begin{cases} 1 & \text{if } \exists j \in \mathcal{E} : a = (i, j) \in \mathcal{A}' \text{ satisfies } e(a) = e \\ -1 & \text{if } \exists j \in \mathcal{E} : a = (j, i) \in \mathcal{A}' \text{ satisfies } e(a) = e \\ 0 & \text{otherwise} \end{cases}$$

We try to explain this parametric function with a visual example.

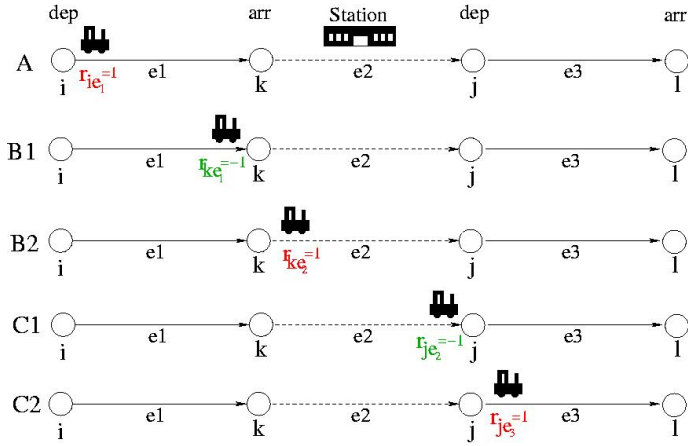


Figure 2.4: Explanation of the parameter r_{ie}

In Figure 2.4 case (A) the train has just departed from a station (event i), so it is using the block of track e_1 and this resource is occupied, hence $r_{ie_1} = 1$. In the second case (B₁) the train has just arrived in a station (event k) so it is not using the track e_1 anymore (then $r_{ke_1} = -1$) but it has to wait (e.g. for fixed connections) on a platform e_2 of the station (case B₂) so the parameter $r_{ke_2} = 1$. When it departs from the station (event j , case C₁) it lets the platform e_2 free (so $r_{je_2} = -1$) and it occupies the block of tracks e_3 (case C₂), then $r_{je_3} = 1$. And so on...

So the capacity of an element $e \in E$ with respect to a given time t and a timetable Π is sufficient if

$$(2.18) \quad \underline{R}_e \leq r_e(\Pi, t) \leq \bar{R}_e \quad \forall t \geq 0$$

where

$$(2.19) \quad r_e(\Pi, t) = \sum_{i \in \mathcal{A}(\Pi, t)} r_{ie}$$

$$\mathcal{A}(\Pi, t) = \{i \in \mathcal{E} : \pi_i \leq t\}$$

and the parameter π_i represents the scheduled timetable of event i .

If a delay occurs in the system, we will not consider the scheduled timetable π_i but the new timetable x_i , i.e. $\mathcal{A}(X, t) = \{i \in \mathcal{E} : x_i \leq t\}$ and the corresponding $r_e(X, t)$.

We can describe this alternative formulation of the Capacitated Timetable Model [CTM-4] as [RCTM]

$$\begin{aligned} (2.20a) \quad & \min \sum_{i \in \mathcal{E}} \omega_i x_i \\ (2.20b) \quad & x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A} \\ (2.20c) \quad & x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E} \\ (2.20d) \quad & \underline{R}_e \leq r_e(X, t) \leq \bar{R}_e \quad \forall e \in E \text{ and } t \geq 0 \\ (2.20e) \quad & r_{ie} \in \{-1, 0, 1\} \quad \forall e \in E \forall i \in \mathcal{E} \\ (2.20f) \quad & x_i \in \mathbb{Z} \quad \forall i \in \mathcal{E} \end{aligned}$$

The parameter $\omega_i, L_a, \pi_i, d_i, r_{ie}$ are given and (2.20d) is the capacity constraints described previously in this section.

This model is an adaptation of the model presented by Neumann and Schwindt and we are going to point out the four major differences between the two models in Section 2.10.

2.9 [RCTM] is \mathcal{NP} -hard

To show that [RCTM] is \mathcal{NP} -hard we could have reduced it to the [CTM-4] model, which has been proved to be \mathcal{NP} -hard in Section 2.5. Instead we prefer to introduce another proof based on the Minimum Parallel Processor Scheduling problem, MPPS. We consider a special case of [RCTM], if we can prove that this special case is \mathcal{NP} -hard then also the generalized problem must be \mathcal{NP} -hard.

We consider a special network with only one line e connecting two stations (for example Hannover Messe and Hannover Hauptbahnhof). This line has a capacity $\bar{R}_e = m$ so no more than m trains can travel simultaneously on this line (the number of tracks between two stations in a big city can reasonably considered bigger than 2, hence we have $m \geq 2$) and we do not consider the capacity of the station itself since it is usually bigger than the one of the single line passing through it. Moreover we neglect all the waiting and changing activities, i.e. $\mathcal{A}^{wait} = \mathcal{A}^{change} = \emptyset$. In other words we are interested in rescheduling the events in \mathcal{E} considering just driving activities \mathcal{A}^{drive} . Moving from these considerations we can rewrite the capacity Constraint (2.19) as

$$r(\Pi, t) = \sum_{i \in \mathcal{A}_1(\Pi, t)} r_i + \sum_{j \in \mathcal{A}_2(\Pi, t)} r_j$$

where we neglect the index e since we are considering just one line and where

$$\begin{aligned} \mathcal{A}_1(X, t) &= \{i \in \mathcal{E}^{dep} : x_i \leq t\} \\ \mathcal{A}_2(X, t) &= \{j \in \mathcal{E}^{arr} : x_j \leq t\}. \end{aligned}$$

Due to (2.20b) we have that $x_j \geq x_i + L_a \forall a = (i, j) \in \mathcal{A}^{drive}$, hence we can rewrite $\mathcal{A}_2(X, t)$ as

$$\bar{\mathcal{A}}_2(X, t) = \{i \in \mathcal{E}^{dep} : x_i + L_a \leq t \quad a = (i, j) \in \mathcal{A}^{drive}\}.$$

since we neglect preemption on the traveling time. Now we notice that r_i can be equal to one if and only if $i \in \mathcal{E}^{dep}$ and it can be equal to -1 if and only if $i \in \mathcal{E}^{arr}$. The capacity constraint can be now written as

$$r(X, t) = \sum_{i \in \mathcal{A}_1(X, t)} r_i - \sum_{j \in \overline{\mathcal{A}}_2(X, t)} \bar{r}_j$$

where $\bar{r}_j = -r_j \forall j \in \overline{\mathcal{A}}_2(X, t)$, and this is equivalent to

$$r(X, t) = \sum_{i \in \tilde{\mathcal{A}}_1(X, t)} r_i$$

where $\tilde{\mathcal{A}}_1(X, t) = \{i \in \mathcal{E}^{dep} : x_i \leq t \leq x_i + L_a \mid a = (i, j) \in \mathcal{A}^{drive}\}$. Therefore we can rewrite [RCTM] as [RCTM-2]

$$\begin{aligned} (2.21a) \quad & \min \sum_{i \in \mathcal{E}^{dep}} \omega_i x_i \\ (2.21b) \quad & x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E}^{dep} \\ (2.21c) \quad & 0 \leq r(X, t) \leq m \quad \forall t \geq 0 \\ (2.21d) \quad & r_i \in \{0, 1\} \quad \forall i \in \mathcal{E}^{dep} \\ (2.21e) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E}^{dep} \end{aligned}$$

where we neglect the precedence constraints since we consider just departure events and driving activities.

If we consider the case in which one of the source delay d_i is considerably large compared to the scheduled timetable, then the solution $x_i = \tilde{\pi} = \pi_i + d_i$ is not anymore feasible and the rescheduling problem can be compared to a scheduling problem and hence we can prove that it is \mathcal{NP} -hard.

To this end we introduce now the Minimum Parallel Processor Total Flow Time Scheduling Problem (MPPS) as given in [47]: given a number of processors $m \in \mathbb{Z}^+$ and a set of operations O , every $o \in O$ is associated with a release time r_o and a length l_o , we look for a m -processor schedule $\sigma : T \rightarrow \mathbb{Z}_0^+$ such that the number of tasks that are executed at time t cannot exceed m (i.e. no overlapping is allowed) and such that the time required to schedule all operations is minimized.

To model this problem it is a standard procedure to introduce a binary variable u such that $u_o = 1$ if the operation o is in execution at time t , 0 otherwise,

$$u_o = \begin{cases} 1 & \text{if } \sigma_o \leq t \leq \sigma_o + l_o \\ 0 & \text{otherwise} \end{cases}$$

so that $u(\Sigma, t) = \sum_{o \in (O, t)} u_o \leq m \forall t$ where $(O, t) = \{o \in O : \sigma_o \leq t \leq \sigma_o + l_o\}$.

The problem can be written as [MPPS]:

$$\begin{aligned} (2.22a) \quad & \min \sum_{o \in O} \omega_o (\sigma_o + l_o) \\ (2.22b) \quad & \sigma_o \geq r_o \quad \forall o \in O \\ (2.22c) \quad & 0 \leq u(\Sigma, t) \leq m \quad \forall t \geq 0 \\ (2.22d) \quad & u_o \in \{0, 1\} \quad \forall o \in O \\ (2.22e) \quad & \sigma_o \in \mathbb{Z}^+ \quad \forall o \in O \end{aligned}$$

[MS] is \mathcal{NP} -hard (as proved in [47]).

Proposition 2.8 $[MS] \infty [RCTM-2]$

Proof: We interpret the events $i \in \mathcal{E}^{dep}$ as operation $o \in O$, so we can identify the release time $\tilde{\pi}_i = r_i$ and the length $L_a = l_o$. Moreover we can identify the scheduling functions $x : \mathcal{E}^{dep} \rightarrow \mathbb{Z}^+$ and $\sigma : O \rightarrow \mathbb{Z}^+$ so that it results $\tilde{\mathcal{A}}_1(X, t) = (O, t)$. x is an optimal solution of [RCTM-2] if and only if σ is an optimal solution of [MS] since the two objective functions differ just for a constant (i.e. $\sum_{o \in O} l_o$). Consequently we state that [RCTM-2] is \mathcal{NP} -hard. \square

2.10 Pointing out the difference between [RCTM] and [CTM-4]

The first difference between our [RCTM] model and the model presented by Neumann and Schwindt in [52], is the fact that we are not going to study the *makespan* of the problem, that is minimizing the completion time of the last activity of the project $\min \max_{i \in V} x_i$, but the *total flow time* of it, that is minimizing the sum of the completion time of all the activities in the project $\min \sum_{i \in V} x_i$.

Neumann and Schwindt focused their attention on the minimization of the starting point of the last activity of the system, we want instead to minimize the sum of the delays.

Lemma 2.9 *In the uncapacitated case, the two objective functions $\min \max_{i \in \mathcal{E}} x_i$ and $\min \sum_{i \in \mathcal{E}} x_i$ are equivalent.*

Proof: We are working on a network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$. If \mathcal{N} is not connected, then there exists at least one node $i \in \mathcal{E}$ that can not be reached through any path from any other node in \mathcal{E} , i.e. whatever will happen this node will not be influenced from what happens in any other node of \mathcal{N} . Therefore we are allowed to neglect this case and to consider \mathcal{N} connected.

Both the makespan and the total tardiness problems have to satisfy the Constraints (2.20b) and (2.20c) (i.e. precedence constraints and earliest starting times). These constraints are defined through some positive integer parameters L_a , π_i and d_i (expressed in minutes or seconds). We suppose that:

- no activity can be interrupted after it has been started;
- no activity can start before the beginning of the project;
- the project ends when every activity has been completed.

Under these hypothesis, both problems can be solved by finding the longest path, i.e. the longest sequence of tasks related by precedence constraints, needed to fulfill the specified constraints.

In the uncapacitated model, precedence constraints are the only bonds on the execution of the jobs. If there would be no precedence constraints at all the optimal solution would be given by the sum of the durations of the single jobs (since JSS does assume precedence constraints among the tasks of every job). In case of precedence constraints among jobs, minimizing the longest path, we minimize not only the earliest starting time of all the tasks that belong to it, but also the earliest starting time of all

the tasks related to them by precedence constraints. Hence we do not only minimize the completion time of the project (makespan) but also its total flow since this solution gives the earliest starting time of every activity.

To find this solution we are going to use a variation of the algorithm of the Distance Matrix.

The distance matrix is defined as $W = (w_{ij})_{i,j \in \mathcal{E}}$ such that

- $w_{ii} = 0$, where 0 indicates the first "dummy" activity;
- $w_{0i} = \pi_i + d_i$ earliest starting time of activity i given precedence constraints;
- $w_{ij} = \max_{k \in \mathcal{E}: (k,j) \in \mathcal{A}} (w_{ik} + L_a)$ where $i, j \in \mathcal{E}$, $i \neq j$, $a = (k, j) \in \mathcal{A}$.
 w_{ij} longest direct walk from i to j .

Algorithm of the Distance Matrix

Given a diagraph $\mathcal{N} = (\mathcal{E}, \mathcal{A})$

- **Step 1** $i, j \in \mathcal{E}$
 - if $i = j$ $w_{ij} = 0$
 - else $i \neq j$
 - if $(i, j) \in \mathcal{A}$
 - if $i = 0$ $w_{0j} = \pi_j + d_j$
 - else $w_{ij} = L_a$ where $a = (i, j)$
 - else $w_{ij} = -\infty$
 - **Step 2** $i, j, k \in \mathcal{E}$
 - if $w_{ik} > -\infty$ and $w_{kj} > -\infty$
 - if $w_{ij} \leq w_{ik} + w_{kj}$ $w_{ij} = w_{ik} + w_{kj}$
-

The sequence of steps of the algorithm does not depend on the chosen objective function (makespan and/or total tardiness), hence except for the output, the solution procedure is exactly the same in both cases. \square

It is still necessary to prove that the solution of the algorithm is the optimal one.

Lemma 2.10 *The earliest starting times w_{0i} processed by the Distance Matrix algorithm is a feasible solution for [CTM-4].*

Proof: We want to show that the constraints (2.4b) and (2.4c) are satisfied.

The definition of the longest direct walk in the case $i = 0$ is:

$$w_{0j} = \max_{k \in \mathcal{E}: (k,j) \in \mathcal{A}} (w_{0k} + L_a) \text{ where } j \in \mathcal{E}, j \neq 0, a = (k, j) \in \mathcal{A}$$

This can be rewritten as

$$w_{0j} = \max\{\pi_j + d_j, \max_{k \in \mathcal{E} \setminus \{0,j\}: (k,j) \in \mathcal{A}} (w_{0k} + L_a)\}$$

where the case $k = j \neq 0$ has been highlighted with respect to the case $k \in \mathcal{E} \setminus \{0, j\}$. Since w_{0i} is the earliest starting time of activity i given the precedence constraints we have that x_j , the re-scheduled timetable of event j , must be bigger than it, i.e. $x_j \geq w_{0i}$. Hence $x_j \geq \pi_j + d_j$, i.e. (2.4c) is fulfilled. Furthermore

$$w_{0j} \geq \max_{k \in \mathcal{E} \setminus \{0,j\}: (k,j) \in \mathcal{A}} (w_{0k} + L_a)$$

Thus $w_{oj} \geq (w_{ok} + L_a) \forall k \in \mathcal{E} \setminus \{o, j\} : (k, j) \in \mathcal{A}$.

The re-scheduling problem is a minimization problem, therefore the new timetable of an activity in the optimal solution will be scheduled as early as possible, i.e. $x_j = w_{oi}$. Hence we have $x_j \geq x_k + L_a \forall k \in \mathcal{E} \setminus \{o, j\} : (k, j) \in \mathcal{A}$, i.e. Constraint (2.4b) is also fulfilled.

In other words w_{oi} is a feasible solution. \square

Remark 2.11 Since the network is connected, there exists a longest path from i to j if and only if there does not exist any direct cycle with positive length.

We can assume the absence of direct cycles with positive length since they would represent an impossible sequence of precedences.

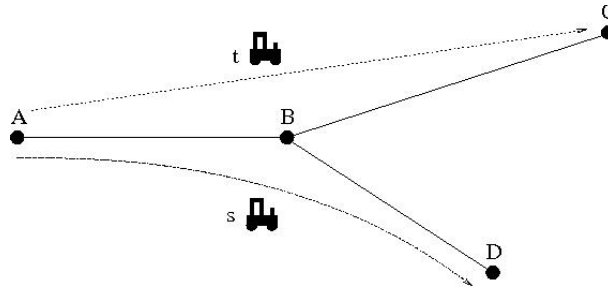


Figure 2.5: Makespan vs. total tardiness

To verify that the two objective functions are not equivalent in the capacitated case, we can consider the network in Fig. 2.5.

Example 2.12

Two trains t and s are leaving from station A and they have to follow the same single track line until station B where their paths separate: the first train is going to station C and the second is going to D . Suppose that the durations of the activities along the arcs are as follows:

$$\begin{aligned} L_{AB}^t &= 50 \text{ and } L_{BC}^t = 40 \text{ for train } t \\ L_{AB}^s &= 45 \text{ and } L_{BD}^s = 35 \text{ for train } s \end{aligned}$$

We have two possible cases:

A: t moves before s . The train s can not overtake the train t until they reach the station B , so it has to go slower and follow the other train, i.e. its traveling time between A and B will not be $L_{AB}^s = 45$ but $\bar{L}_{AB}^s = L_{AB}^s + (L_{AB}^t - L_{AB}^s) = 45 + (50 - 45) = 50$. The arrival times at the final destinations will be $x_C^t = 50 + 40 = 90$ minutes and $x_D^s = 50 + 35 = 85$ minutes.

B: s moves before t . The arrival times will be $x_C^t = 50 + 40 = 90$ minutes and $x_D^s = 45 + 35 = 80$ minutes

If we now consider the *makespan problem* both possibilities give the same solution (the last activity ends after 90 minutes) but, if we consider the *total tardiness problem*, it is clear that the case **B** is better than the case **A** (265 minutes of total traveling time against 275).

In this example we omitted the fact that the trains have to respect a security distance but if we fix it, e.g. $d_{AB} = 3$ minutes, the difference between the two objective functions become even more clear.

A: t moves before s . Again is $\bar{L}_{AB}^s = 50 + 3$ as explained before. The arrival times at the final destinations will be $x_C^t = 50 + 40 = 90$ minutes and $x_D^s = 53 + 35 = 88$ minutes since the second train has to respect also the security distance.

B: s moves before t . The arrival times will be $x_C^t = 3 + 50 + 40 = 93$ minutes and $x_D^s = 45 + 35 = 80$ minutes

In that situation the minimal makespan is reached in the case **A** (90 minutes against 93) instead the minimal total tardiness comes from the case **B** (274 minutes against 284). \triangleleft

The second difference is that we limited the value of "resources consumption / replenishment" to the set $\{-1, 0, 1\}$.

The third difference is that we fixed the lower bound for the capacity to zero.

The fourth difference is the way we are going to consider the equation (2.20d) of the model, the resource constraint. This constraint is not linear so, to linearize it, we introduce a new parameter $M = \max_{i \in \mathcal{E}} (T - x_i) + 1$ where T is an upper bound for the timetable (e.g. the period of the timetable), and new variables $v_{it} \in \{0, 1\}$ defined through the following inequalities:

$$(2.23) \quad v_{it} \leq 1 + \frac{(t - x_i)}{M}$$

$$(2.24) \quad v_{it} \geq \frac{(t - x_i) + \epsilon}{M}$$

where ϵ is a positive constant smaller than one, i.e. $\epsilon \in (0, 1)$. We have then the following proposition.

Proposition 2.13 For all x_i, v_{it} satisfying $x_i \geq 0$, $v_{it} \in \{0, 1\}$, (2.23) and (2.24), we have:

$$v_{it} = 1 \Leftrightarrow t \geq x_i$$

.

Proof: First we notice that $M > 0$ since the definition of $T > 0$, therefore there always exists an event scheduled before time T (at least the starting event).

$v_{it} = 1$

Replacing the value of v_{it} into (2.23) we have

$$0 \leq \frac{(t - x_i)}{M} \Rightarrow (t - x_i) \geq 0$$

that is $t \geq x_i$. Moreover (2.24) yields $M \geq (t - x_i) + \epsilon$ which is satisfied for $M = \max_{i \in \mathcal{E}} (T - x_i) + 1$ since $T \geq t \forall t$ and $\epsilon < 1$ by definition.

$t \geq x_i$

Then $\frac{(t - x_i) + 0.5}{M} > 0$. From (2.24) we obtain the $v_{it} > 0$, i.e. $v_{it} = 1$ since $v_{it} \in \{0, 1\}$. Moreover (2.23) is satisfied for both $v_{it} = 0$ and $v_{it} = 1$.

□

We can rewrite the function $r_e(X, t)$ defined in (2.19) as

$$(2.25) \quad r_e(X, t) = \sum_{i \in \mathcal{E}} r_{ie} v_{it}$$

and the Linearized Capacitated Timetable Model [LCapTM] is

$$(2.26a) \quad \min \sum_{i \in \mathcal{E}} x_i$$

$$(2.26b) \quad x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A}$$

$$(2.26c) \quad x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E}$$

$$(2.26d) \quad \underline{R}_e \leq \sum_{i \in \mathcal{E}} r_{ie} v_{it} \leq \bar{R}_e \quad \forall e \in E \text{ and } t \in [0, T]$$

$$(2.26e) \quad v_{it} \leq 1 + \frac{(t - x_i)}{M} \quad \forall i \in \mathcal{E} \text{ and } t \in [0, T]$$

$$(2.26f) \quad v_{it} \geq \frac{(t - x_i) + 0.5}{M} \quad \forall i \in \mathcal{E} \text{ and } t \in [0, T]$$

$$(2.26g) \quad v_{it} \in \{0, 1\} \quad \forall i \in \mathcal{E}$$

$$(2.26h) \quad x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E}$$

where $M = \max_{i \in \mathcal{E}} (T - x_i)$.

The new problem arising from this formulation is the increasing number of binary variables v_{it} necessary to define the model.

Proposition 2.14 *The cardinality of the set of parameters v_{it} grows to infinity as the period T of the problem increases.*

Moreover [LCapTM] is \mathcal{NP} -hard since it can be reduced to the MS problem which has been proved to be \mathcal{NP} -hard ([52] or [30] for the makespan objective function, or [48] for weighted total flow time objective function).

Since it is not easy to get around the \mathcal{NP} -hardness of the analytical constrained models, we looked for other possible approaches and we suggest now a new approach based on the concept of “virtual activities”.

2.11 Identifying dependencies through a stochastic approach

Our aim is to simplify the formulation of [CTM-4] without a loss of quality of the solution so that, even if the obtained Macromodel will not contain all the information of the corresponding Micromodel, its optimal solution will have such a quality level then it will not be necessary to re-run the Micro- and Macromodel many times to find a conflict-free solution for the rescheduling problem.

Instead of using all capacity constraints for each block as in [CTM-4] we want to apply a stochastic procedure which points out the critical points of the system, in particular the location of the source delays and how they spread out into the system. We can reduce the [CTM-4] formulation by restricting the set of capacity constraints to a smaller set of abstract constraints which contains just these critical points.

Using the delays of each event in this model as random variables we use a stochastic approach to analyze the dependencies among these variables. This method, that will be explained in details in Chapter 3, reveals the dependencies among the events of

the system, i.e. among the arrival and departure events of the trains in the stations. These dependencies represent information about all three types of delay propagation. While delay propagation of type 1 and 2 belongs to driving, waiting and changing activities, delay propagation of type 3 does not correspond to any $a \in \mathcal{A}$. Hence we introduce a set of “virtual activities” $\mathcal{A}^{virtual}$ describing the dependencies of type 3 which have to be identified by the stochastic approach. The “virtual” activities ensure that an event can not happen before another event has taken place: that means, for example, that a train can enter in a station only if its assigned platform is free. Hence a “virtual activity” does not belong to the set of activities defined in \mathcal{A} but it can be considered as a precedence constraint in the railway problem that has to be satisfied to avoid infrastructure conflicts (using the same track or the same platform) due to the limited capacity of the track system and to operational rules of the security system. The resulting model is similar to [CTM-3]:

$$\begin{aligned}
 (2.27a) \quad & \min \sum_{i \in \mathcal{E}} x_i \\
 (2.27b) \quad & x_j - x_i \geq L_a \quad \forall a = (i, j) \in \mathcal{A} \\
 (2.27c) \quad & x_i \geq \pi_i + d_i \quad \forall i \in \mathcal{E} \\
 (2.27d) \quad & x_j \geq v_{ij}^1 x_i + v_{ij}^2 \quad \forall i, j \in \mathcal{E} \text{ such that } a = (i, j) \in \mathcal{A}^{virtual} \\
 (2.27e) \quad & x_i \in \mathbb{Z}^+ \quad \forall i \in \mathcal{E}
 \end{aligned}$$

where we replace Constraints (2.6e) and (2.6f) with linear dependency constraints between pairs of events connected with “virtual” edges in the set $\mathcal{A}^{virtual}$ (2.27d) and the parameters v_{ij}^1 and v_{ij}^2 will be estimated by the stochastic approach (see Section 5.6). We call these constraints *virtual activities*.

Stochastic Approach

I would rather discover a single casual relationship, than be king of Persia.

DEMOCRITUS (450BC-370BC)
pre-Socratic Greek philosopher

In Chapter 2 we defined an analytical (Macro-)model for the railway system based on the Activity-on-arc Project Network. Moving from those definitions, we introduce in this chapter some stochastic approaches, that help us to identify the critical points of the system, i.e. the points where a detailed observation of the process (Micromodel) would be suitable to solve the possible conflicts among trains.

Our aim is to find an alternative formulation for the resource constraints of the Capacitated Timetable Model, in order to increase the quality of the solution of the Macro-model, i.e. to reduce the amount of conflicts in the Macro-solution so that the use of Micromodel could be limited.

3.1 Probability and Independence

The term *experiment* is used to refer to any process whose outcome is not known in advance. The *sample space* associated with an experiment is the set of all possible outcomes and it is usually denoted by Ω .

The goal of probability theory is to compute the probability of various events of interest. Intuitively, an event is a statement about the outcome of an experiment. Formally, an *event* A is a subset of the sample space, i.e. $A \subseteq \Omega$. Since events are just sets, we can perform the usual operations of the set theory on them: union, intersection, complement ...

In particular we say that two events A and B are *disjoint* if their intersection is the empty set, $A \cap B = \emptyset$.

A *probability* is a way of assigning numbers to events that satisfies the following axioms:

- for any event $A \subseteq \Omega$, $0 \leq \mathbb{P}(A) \leq 1$;
- $\mathbb{P}(\Omega) = 1$;
- for any countable sequence of disjoint events $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

In words, the probability is a positive function, normalized to one, that when applied to the union of disjoint events is equal to the sum of the probabilities of each event.

A probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is a measure space with a measure \mathbb{P} that satisfies the probability axioms. \mathcal{A} is the σ -algebra over the set Ω . In our further argumentation, we will always assume that $\Omega = \mathbb{R}^p$ where p is chosen according to the dimension of the multivariate variables of the model, \mathcal{A} is the corresponding Borel σ -algebra and \mathbb{P} is a suitable probability function. It is a much debated question, which probability \mathbb{P} should be chosen (see Section 5.1). For the graphical approach we will assume a normal distribution of the variables.

A random variable X is a measurable function from a sample space to the measurable space of possible values of the variable, $X : \Omega \rightarrow \Omega'$. In this thesis just real-valued random variables will be considered, i.e. $X : \Omega \rightarrow \mathbb{R}$.

A fundamental concept in probability and statistic theory is the one of *independence*. Two events A and B in Ω are said to be (marginally) *independent* if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$$

In particular two random variables X and Y are independent if

$$\mathbb{P}(\{X = x, Y = y\}) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(\{X = x\}) \mathbb{P}(\{Y = y\})$$

for every possible choice of the values x and y . In general, "misusing" the notation, we will write that two random variables X and Y are independent if

$$\mathbb{P}(X \cap Y) = \mathbb{P}(X) \mathbb{P}(Y)$$

If the probability of an event B is strictly positive, i.e. $\mathbb{P}(B) \neq 0$, we can define the *conditional probability* of two events A and B through the ratio

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

thus an equivalent definition of independence would be $\mathbb{P}(A|B) = \mathbb{P}(A)$. That can be read as: knowing that B occurs, does not change the probability that A occurs.

Analogously we can define this property for two random variables X and Y :

$$\mathbb{P}(\{X = x\}|\{Y = y\}) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(\{Y = y\})}$$

The central concept in graphical methods is the *conditional independence*.

We consider three random variables X, Y and Z . If X and Y are independent in the conditional distribution given each possible value Z , i.e

$$\mathbb{P}(X \cap Y|Z) = \mathbb{P}(X|Z)\mathbb{P}(Y|Z)$$

then we say that X and Y are conditionally independent given Z and we write $X \perp\!\!\!\perp Y | Z$. This notation is due to Dawid (1979), who discussed alternative characterizations of this property.

It can be also useful to consider the notion of *irrelevance*, in the sense that we can interpret the statement $X \perp\!\!\!\perp Y | Z$ as: "If we know Z , any information about Y is irrelevant in the knowledge of X ".

3.2 Statistical test

A statistical test provides a mechanism for making quantitative decisions about a process or processes [33]. The intent is to determine whether there is enough evidence

to “reject” a conjecture or hypothesis about the process. The conjecture is called the null hypothesis. Not rejecting may be a good result if we want to continue to act as if we “believe” the null hypothesis is true. Or it may be a disappointing result, possibly indicating we may not yet have enough data to “prove” something by rejecting the null hypothesis. A classic use of a statistical test is to check wherever two random variables, X and Y , are independent or not. The hypotheses of the test, are hence so defined: “null hypothesis” the variables are independent; “alternative hypothesis” the variables are dependent. The null hypothesis is so called, because it proposes something initially presumed true. It is rejected only when it becomes evidently false. That is, when there exists a certain degree of confidence that the data do not support the null hypothesis. This confidence is reached through the application of some statistical test, in our case the Chi-squared Test on a Contingency Table.

In other words, the null hypothesis is a statement about a belief. We may doubt that the null hypothesis is true, which might be why we are “testing” it. The alternative hypothesis might, in fact, be what we believe to be true. The test procedure is constructed so that the risk of rejecting the null hypothesis, when it is in fact true, is small. This risk α is often referred to as the significance level of the test. By having a test with a small value of α , we feel that we have actually “proved” something when we reject the null hypothesis. The choice of α is somehow arbitrary, although in practice values of 1%, 5% and 10% are common. A value $\alpha = 5\%$ implies that the null hypothesis is rejected 5% of the time when it is in fact true. Hence, the significance level is the probability that the null hypothesis will be rejected in error when it is true (a decision known as a Type I error, or “false positive”).

The risk of failing to reject the null hypothesis when it is in fact false is not chosen by the user but is determined, as one might expect, by the magnitude of the real discrepancy. This risk β is usually referred to as the Type II error. Large discrepancies between reality and the null hypothesis are easier to detect and lead to small errors of the second kind; while small discrepancies are more difficult to detect and lead to large errors of the second kind. Also the risk β increases as the risk α decreases.

Different α -levels have different advantages and disadvantages. A very small α -level (say 1%) is less likely to be more extreme than the critical value and so is more significant than high α -level values (say 10%). However, smaller α -levels run greater risks of failing to reject a false null hypothesis (a Type II error or “false negative”), and so have less statistical power.

In particular a Chi-squared Test is any statistical hypothesis test in which the test statistic has a Chi-squared distribution when the null hypothesis is true, or anyone in which the probability distribution of the test statistic (assuming the null hypothesis is true) can be approximated by a Chi-squared distribution as closely as desired by making the sample size large enough. A quantity has a Chi-squared distribution with r degrees of freedom if it can be rewritten as the sum of the squares of r standard normal independent variables (mean zero and variance equal to one).

Given a significance level α , the outcome of the test is compared with the critical value of a Chi-squared variable with the corresponding number of degrees of freedom. The “null hypothesis” is rejected if the value of the test is bigger than the critical value.

3.3 Chi-squared Test on a Contingency Table

An easy test to check the independence of variables is the Chi-squared Test on a Contingency Table (see [77]). This is based on the concept of conditional probability. From a data set we consider two random variables X and Y . Willing to test if they are independent, we define two hypotheses:

“null hypothesis”: the variables are independent;

“alternative hypothesis” : the variables are dependent.

We can think of the following situation: the values of the random variables X and Y can be subdivided respectively into r and s classes. As notation we have:

- n total number of observations;
- h and v class-indexes for X and Y ;
- n_{hv} number of elements belonging to the intersection of the classes X_h and Y_v .

We can write the Contingency Table as

n_{11}	n_{12}	n_{13}	\dots	n_{1v}	\dots	n_{1s}	n_1
n_{21}	n_{22}	n_{23}	\dots	n_{2v}	\dots	n_{2s}	n_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{h1}	n_{h2}	n_{h3}	\dots	n_{hv}	\dots	n_{hs}	n_h
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{r1}	n_{r2}	n_{r3}	\dots	n_{rv}	\dots	n_{rs}	n_r
n_1	n_2	n_3	\dots	n_v	\dots	n_s	n

where n_h and n_v are the partial sums on a line and on a column respectively, i.e.

$$n_h = \sum_v n_{hv} \text{ and } n_v = \sum_h n_{hv} .$$

The joint distribution of X and Y is given by the probabilities $\pi_{hv} = \mathbb{P}(X = h, Y = v)$, i.e. the probability that a single observation belongs to both the classes X_h and Y_v . We suppose that the observations are independent, i.e. every observation “chooses” the classes h and k without being influenced from any previous observations. Then the quantities N_{hv} , i.e. the random variables that correspond to the results n_{hv} , can be represented with a multinomial distribution of size n and parameters π_{hv} :

$$[N_{11}, N_{12}, \dots, N_{rs}] \sim M(n, \pi_{11}, \pi_{12}, \dots, \pi_{rs})$$

We are interested in the null hypothesis that the variables X and Y are independent i.e.

$$(3.1) \quad \mathbb{P}(X = h, Y = v) = \mathbb{P}(X = h)\mathbb{P}(Y = v)$$

If we define

$$\pi_h = \mathbb{P}(X = h) = \sum_v \pi_{hv} \text{ and } \pi_v = \mathbb{P}(Y = v) = \sum_h \pi_{hv}$$

we can then rewrite the “null hypothesis” (i.e. the variables are independent) as

$$\pi_{hv} = \pi_h \pi_v \quad \text{for every } h \text{ and } v$$

The Chi-squared Test is

$$T = \sum_{h,v} \frac{(N_{hv} - n\pi_h\pi_v)^2}{n\pi_h\pi_v}$$

The parameters π_h and π_v are nuisance parameters, since we are not interested in them. We approximate them through the data

$$\widehat{\pi}_h = \frac{N_h}{n} \text{ and } \widehat{\pi}_v = \frac{N_v}{n}$$

Thus the test can be rewritten as

$$(3.2) \quad T = \sum_{h,v} \frac{(N_{hv} - \frac{N_h N_v}{n})^2}{\frac{N_h N_v}{n}}$$

The quantity T has a Chi-squared distribution with a number of degrees of freedom equal to

$$\nu = (\text{number of classes}) - 1 - (\text{number of nuisance parameters})$$

The number of classes is given by the product rs . Concerning the nuisance parameters, we observe that $\widehat{\pi}_r$ is fixed once we define $\widehat{\pi}_1, \widehat{\pi}_2 \dots \widehat{\pi}_{r-1}$ since the probability is normalized to one, i.e. $\sum_h \widehat{\pi}_h = 1$. Therefore we need to approximate just $r - 1$ respectively $s - 1$ parameters. Then we have

$$\nu = rs - 1 - (r - 1) - (s - 1) = (r - 1)(s - 1)$$

Given a desired significance level, we can compare the value of (3.2) with the critical values of a Chi-squared variable characterized by $(r - 1)(s - 1)$ degrees of freedom: i.e. the independence (null) hypothesis is refused if the value of T is bigger than the critical value.

In the delay problem the variables X and Y can be seen as two random variables corresponding to the delays of two events in the graph. In our case the (two) classes for every variable correspond to the states:

- punctuality (e.g. earlier arrival or delays smaller than 3 minutes);
- delay (e.g. delays bigger than 3 minutes).

Thus in our case $r = s = 2$ and $\nu = 1$.

Applying the Contingency Table Test directly to every possible pair of variables we can check if there exists a dependency between the variables. What we would like to do is to look for more complex dependencies, in fact we want to see not only if two variables are independent, but also if they are independent given all the other variables and in particular given a third variable, i.e. if there exist a third variable that can explain the dependence (e.g. given at one station two guaranteed connections, between the arrival event of train t_1 and both the departure events of trains t_2 and t_3 . We do not expect any direct dependence between the two departure events, since their similar behavior can be explained through the arrival event).

The idea of considering triples of variables arises from the Tri-graph, a stochastic graphical model that will be explained later in this chapter (Subsection 3.5.3). Therefore given two variables X and Y corresponding to two different events in the system, we will consider a third variable Z (chosen among all other variables of the system except X and Y) and we will define two contingency tables \mathcal{P}_Z and \mathcal{D}_Z . In any of these tables we will insert the observed values of the pair (X, Y) , corresponding to the days in which the variable Z is punctual (\mathcal{P}_Z) or delayed (\mathcal{D}_Z) as defined above.

Then the test given by Formula (3.2) will be applied to both \mathcal{P}_Z and \mathcal{D}_Z to check the "null hypothesis" under the constraint of punctuality/delay of the third variable. We will repeat this step for every pair of events (X, Y) and for every possible choice of the third variable Z except the variables forming the considered pair, i.e. $Z \neq X, Y$.

The alternative hypothesis (X and Y are dependent) will be "accepted" if the direct contingency table test rejected the null hypothesis and given any third variable Z , the null hypothesis will be accepted in at least one table \mathcal{P}_Z or \mathcal{D}_Z (i.e. there does not exist any third variable Z that can explain the dependence between X and Y). This procedure should allow a direct comparison between the results of the method and the ones of the stochastic methods that will be introduced later in this chapter.

3.4 Graphical Methods

Graphical methods have their origin in several scientific areas and they can be considered as a marriage between probability theory and graph theory. They provide a natural tool for dealing with uncertainty and complexity. The basic idea is the notion of modularity, so that a complex system can be built by combining simpler parts. Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent and providing ways to interface models and data. Graph theory provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to design general-purpose algorithms. Therefore we can define them as a sort of multivariate analysis that uses graphs to represent models.

Probabilistic graphical models are graphs in which the nodes represent random variables and the (lack of) arcs represent conditional independence assumption. Hence they are a compact representation of joint probability distribution.

There are two main kinds of graphical models: undirected and directed (but it is also possible to have a model with both directed and undirected arcs, called chain graph). Undirected graphical models, also known as Markov networks, are more popular but they are not the most suitable to be applied in the railway system since a timetable is based on a sequence of actions in time and hence of ordered variables. Directed graphical models, also known as Bayesian networks, are then the one that will be considered in this thesis. Note that despite the name, Bayesian networks do not necessarily imply a commitment to Bayesian methods; rather they are so called because they use Bayes' rules for inference.

In a directed graphical model an arc from A to B can be informally interpreted as indicating that A "causes" B. Directed cycles are disallowed in the railway systems since they would represent an impossible sequence of precedences, but undirected cycles are still possible.

Example 3.1 (Undirected cycle in the railway system)

We consider two trains t and s . We suppose that s is an Euro City (EC) traveling from Basel to Berlin, and that t is an Inter City Express (ICE) traveling from München to Hamburg. Their journeys will lead them to Kassel, where passengers of the train t have the possibility to change to train s . Here, even if train s arrives after t , it will leave the station first due to a fixed priority list (e.g. internal decision to give precedence to ICE trains). Between Kassel and Göttingen, both trains have to use the same line, so in Göttingen the passengers of train s have the possibility to change to train t to reach their final destination. The problem is represented in Figure 3.4 both in the form of Public Transport Network and of Activity-on-arc Project Network, which is more intuitive to understand.

◀

3.4.1 Graph: notation and terminology

In graphical modeling the dependence pattern between variables is associated with a graph in which vertices encode the random variables and edges encode conditional dependence between the variables. Hence before going deeply into this topic, it is necessary to recall some definitions of the Graph Theory (the following notation refers to [45]).

A graph $G = (V, E)$ is defined through a set of vertices V and a set of edges E that is a subset of $V \times V$.

Edges $(v, w) \in E$ with both (v, w) and (w, v) in E are called *undirected*, whereas an

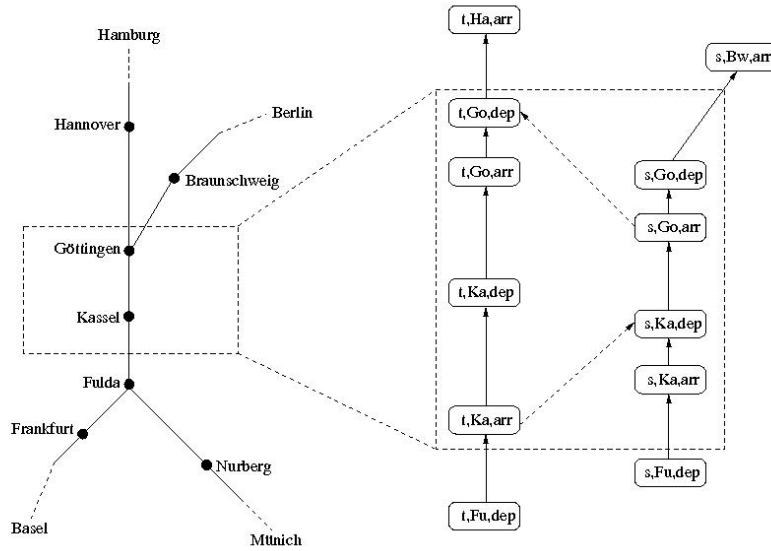


Figure 3.1: Example of undirected cycle in the railway system

edge (v, w) with its opposite (w, v) not in E is called directed.

The graphs we consider are directed, that is, they contain only directed edges drawn as *arrows*, $v \rightarrow w$, thus we can identify edges with *ordered* pairs of vertices (v, w) .

If $S \subset V$ is a subset of the vertex set, it induces a subgraph $G_S = (S, E_S)$, where the edge set $E_S = E \cap (S \times S)$ is obtained from G by keeping all edges with both endpoints in S .

Given two nodes $v, w \in V$, we say that they are *adjacent*, $v \sim w$, if there exists an edge between v and w , without distinction between $v \rightarrow w$ and $w \rightarrow v$. The set of vertices that are adjacent to a vertex is denoted as $ne(v)$. We say that v is a *parent* of w if there exists an edge $v \rightarrow w$, then w is called a *child* of v . The set of parents of w is denoted as $pa(w)$ and the children of v is denoted as $ch(v)$. If there is a directed path from v to w , then v is called an *ancestor* of w and w is called a *descendent* of v . The set of ancestors of w is denoted as $an(w)$ and the set of descendents of v is denoted as $de(v)$.

These definitions can be extended to set of nodes. Given a set $S \subset V$ we define:

- $ne(S) = (\cup_{v \in S} ne(v)) \setminus S$;
- $pa(S) = (\cup_{v \in S} pa(v)) \setminus S$;
- $ch(S) = (\cup_{v \in S} ch(v)) \setminus S$;
- $an(S) = (\cup_{v \in S} an(v)) \setminus S$;
- $de(S) = (\cup_{v \in S} de(v)) \setminus S$;

The *boundary* $bd(S)$ of a subset S of vertices is the set of vertices in $V \setminus S$ that are adjacent to vertices in S , i.e. $bd(S) = ne(S) \setminus S$.

A *path* (of length m) is a sequence of vertices $v_0, v_1 \dots v_m$ such that $v_i \sim v_{i+1} \forall i \in \{0 \dots m-1\}$.

A path is *directed* if $v_i \rightarrow v_{i+1} \forall i \in \{0 \dots m-1\}$.

If $v_0 = v_m$ the (directed) path is called a (*directed*) *cycle*. An m -cycle is a path of length m such that $v_0 = v_m$.

If there is a directed path from v to w we say that v leads to w and write $v \mapsto w$. If both $v \mapsto w$ and $w \mapsto v$ we say that v and w are connected and write $v \rightleftharpoons w$. The corresponding equivalence classes $[v]$, where

$$w \in [v] \Leftrightarrow v \rightleftharpoons w$$

are the (strong) *connectivity components* of G . If $v \in S \subset V$, the symbol $[v]_S$ denotes the (strong) connectivity component of v in G_S .

A subset $S \subset V$ is said to be a (v, w) -*separator* if all paths from v to w intersect S . Thus in an undirected graph S is a (v, w) -*separator* if and only if $[v]_{V \setminus S} \neq [w]_{V \setminus S}$.

The subset S_3 of V is said to separate the subsets S_1 and S_2 of V if it is a (v, w) -*separator* for every $v \in S_1$ and $w \in S_2$.

If $bd(v) \subset S$ for all $v \in S$, we say that S is an *ancestral set*. In a directed graph a set S is ancestral if and only if $an(v) \subset S$ for all $v \in S$. In an undirected graph, the ancestral sets are union of connectivity components. The intersection of a collection of ancestral sets is again ancestral. Hence for any subset S of vertices there is a smallest ancestral set containing S which is denoted by $An(S)$.

A *tree* is a connected, undirected graph without cycles. It has a unique path between two vertices. A *forest* is an undirected graph where all connectivity components are trees. The name forest arises from the concept that a forest consists only of (possibly disconnected) trees.

Chain graphs are graphs where the vertex set V can be partitioned into numbered subsets, forming a so-called *dependence chain* $V = V(1) \cup \dots \cup V(m)$ such that all edges between vertices in the same subset are undirected and all edges between vertices in different subsets are directed, pointing from the set with lower number to the one with higher number. Such graphs are characterized by having no directed cycles and the connectivity components form a partitioning of the graph into chain components. A graph is a chain graph if and only if its connectivity components induce undirected subgraphs. An undirected graph is a special case of chain graph. A directed, acyclic graph is a chain graph with all chain components consisting of one vertex.

For a chain graph G we define its *moral graph* G^m as the undirected graph with the same vertex set but with v and w adjacent in G^m if and only either $v \rightarrow w$ or $w \rightarrow v$ or if there are s_1, s_2 in the same chain component such that $v \rightarrow s_1$ and $w \rightarrow s_2$. If no edges have to be added to form the moral graph, the chain graph is said to be *perfect*. The name moral graph arises from the connection (marriage) of nodes that have a common child.

In graphical modeling, the focus is on models under which some conditional independence relations of the form $X \perp\!\!\!\perp Y \mid$ (some other variables) holds (see [25]). In particular, we focus on models for which these relations take the form $X \perp\!\!\!\perp Y \mid$ (the rest), whereas “the rest” means all other variables in the model. For such a model, we can construct an undirected graph $G = (V, E)$ where V is the set of variables in the model and E consists of edges between variable pairs that are not conditionally independent given the rest. In other words, for all pairs $\{X, Y\}$ such that $X \perp\!\!\!\perp Y \mid$ (the rest), the edge between X and Y is omitted; for all other pairs, an edge is drawn. Thus if two variables are not adjacent, then they are conditional independent given the rest. This is known as the pairwise Markov property for undirected graphs.

Graph separation is a very important example of a model for the conditional independence on undirected graphs. Let S_1, S_2 and S_3 be subsets of the vertex set V of a finite undirected graph $G = (V, E)$, the global Markov property states

$$\text{if } S_3 \text{ separates } S_1 \text{ from } S_2 \text{ in } G \text{ then } S_1 \perp\!\!\!\perp S_2 \mid S_3 .$$

As example we consider four variables W , X , Y and Z for which we know that $W \perp\!\!\!\perp Z \mid (X, Y)$ and $Y \perp\!\!\!\perp Z \mid (W, X)$. The edges $\{W, Z\}$ and $\{Y, Z\}$ must be absent from the graph on $V = \{W, X, Y, Z\}$. We obtain the graph shown in Figure 3.2, from which we can infer that $W \perp\!\!\!\perp Z \mid X$ and $Y \perp\!\!\!\perp Z \mid X$.

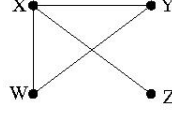


Figure 3.2: Example of the separation concept on a graph

The global Markov property allows a simple translation from a graph-theoretic property, separation, to a statistical property, conditional independence.

That is S_1 and S_2 are conditional independent given S_3 if every path from (a node in) S_1 to (a node in) S_2 passes through (at least one node in) S_3 . An analogous model can be defined also for directed graphs introducing the concept of moral graphs. Given S_1, S_2 and S_3 subsets of the vertex set V of a finite directed graph $G = (V, E)$, then the direct Markov property is

$$\text{if } S_3 \text{ separate } S_1 \text{ from } S_2 \text{ in } G_{An(S_1 \cup S_2 \cup S_3)}^m \text{ then } S_1 \perp\!\!\!\perp S_2 \mid S_3$$

where $G_{An(S_1 \cup S_2 \cup S_3)}^m$ is the moral graph of the smallest ancestral set containing $S_1 \cup S_2 \cup S_3$.

As a particular case, we can consider $S_1 = \{v_i\}$, $S_2 = \{v_j\}$ and $S_3 = \{v_k\}$, subsets of V . So the vertices v_i and v_j are independent given v_k , if every path from v_i to v_j in the moral graph $G_{An(\{v_i, v_j, v_k\})}^m$ includes the vertex v_k . We will write $v_i \perp\!\!\!\perp v_j \mid v_k$.

In general the procedure to check whether two nodes v_i and v_j are independent given a set S , i.e. $v_i \perp\!\!\!\perp v_j \mid S$, can be schematized as follows:

1. consider the ancestral set of $\{v_i, v_j\} \cup S$, that is $An(\{v_i, v_j\} \cup S) = A$;
2. draw the subgraph corresponding to the ancestral set $G_A \subset G$;
3. construct the moral graph of G_A , G_A^m :
 - connecting with undirected edges all pairs of vertices in $pa(v) \quad \forall v \in A$;
 - replacing all the edges in G_A with undirected edges;
4. check if there does not exist any path connecting v_i and v_j in G_A^m that does not involve any node of S (i.e. if S is separating v_i and v_j on G_A^m).

In this thesis, we will consider directed graphs with no directed cycles since in the railway network these would represent a set of impossible precedence constraints. In fact the railway timetable is a list of ordered time events. These can be enumerated in such a way that $i < j$ if and only if event i happens before event j .

3.5 Continuous method

In this section we describe three graphical methods based on the hypothesis of multivariate normal distribution of the variables. Two of them (*Full Conditional Independence Graph* and *Covariance Graph*) are classical methods, the third one (*Tri-graph*) has been suggested in 2004 by Wille and Bühlmann (see [75] and [76]) and is the method that will be mainly applied to this work.

Gaussian Assumption:

We suppose we have a p -dimensional random variable $X = (X_1 \dots X_p)$ with a multivariate normal distribution:

- mean $\mu = (\mu_1 \dots \mu_p)$
- covariance matrix $\Sigma = (\sigma_{ij})_{i,j \in \{1, \dots, p\}}$ and $\sigma_{ij} = \text{cov}(X_i, X_j)$
- precision matrix $\Omega = \Sigma^{-1} = (\omega_{ij})_{i,j \in \{1, \dots, p\}}$
- density $f_X = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu))$.

Under the Gaussian Assumption, we can rewrite the Markov property as follows:

Definition 3.2 [45] *A multivariate normal distribution $X = (X_1 \dots X_p)$ follows the (global) Markov property with respect to a graph G if for all vertices X_i and X_j and set of vertices K so that $X_i, X_j \notin K$ and K separates X_i and X_j in G , it holds that $X_i \perp\!\!\!\perp X_j \mid \{X_k : k \in K\}$*

Besides we can introduce the notion of faithfulness

Definition 3.3 [76] *A multivariate normal distribution $X = (X_1 \dots X_p)$ is faithful to a graph G if for all vertices X_i and X_j and set of vertices K so that $X_i, X_j \notin K$ and $X_i \perp\!\!\!\perp X_j \mid \{X_k : k \in K\}$, it holds K separates X_i and X_j in G .*

Assuming that the covariance matrix Σ is regular such that the precision matrix $\Omega = \Sigma^{-1}$ is well defined, the following theorem can be proved.

Theorem 3.4 [45] *Assume that $X = (X_1 \dots X_p)$ has a multivariate normal distribution with mean μ and (regular) covariance matrix Σ , then it holds for every X_i, X_j with $i \neq j$ that*

$$X_i \perp\!\!\!\perp X_j \mid X_{\{1, \dots, p\} \setminus \{i, j\}} \iff \omega_{ij} = 0$$

where $\Omega = (\omega_{ij})_{i,j \in \{1, \dots, p\}} = \Sigma^{-1}$ is the precision matrix of the distribution

The proof can be found in [45]. Moreover for the multivariate normally distributed variables the notion of correlation (i.e. covariance equals to zero) and independence are equivalent (see [46], [42], [43]) although in general this is not true. Hence the following theorem holds.

Theorem 3.5 [46] *Assume that $X = (X_1 \dots X_p)$ has a multivariate normal distribution with mean μ and covariance matrix Σ . Then for every X_i, X_j in X with $i \neq j$ we have X_i and X_j are uncorrelated if and only if they are independent.*

3.5.1 Full Conditional Independence Graph

In the Full Conditional Independence Graph (FCIG) an edge between two variables X_i and X_j is drawn if and only if the two variables are conditionally dependent given all other variables of the system $X_k \forall k \in \{1 \dots p\} \setminus \{i, j\}$.

Due to the Theorem 3.4, we can rewrite this condition as:

Definition 3.6 FCIG draws an edge $X_i \rightarrow X_j \iff \bar{\omega}_{ij} = \frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}} \neq 0$

where ω_{ij} is the element in the i^{th} row and j^{th} column of the precision matrix Ω .

The advantage of this method is the simultaneous inclusion of all the variables in the evaluation of the dependencies, since computing the precision matrix requires to consider the determinant of the covariance matrix. Its shortcoming is the numerical implementation: to be able to calculate the inverse of the covariance matrix a large sample of data is required to have an accurate estimation. Moreover to determine which elements of the precision matrix are equal to zero, it is necessary to perform a super exponential number of tests (Likelihood tests), and for a large number of variables this is hardly feasible.

3.5.2 Covariance Graph

A natural choice to avoid the problem arising from the use of the precision matrix in the FCIG would be to work directly with the covariance matrix. Then it would be possible to work even with a small number of data and a large set of variables.

The Covariance Graph (CG) is based on this idea. It that draws an edge between two variables X_i and X_j if and only if the correlation coefficient of the two variables is different than zero (see Theorem 3.5).

Definition 3.7 CG draws an edge $X_i \rightarrow X_j \iff \rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} \neq 0$

where σ_{ij} is the element in the i^{th} row and j^{th} column of the covariance matrix Σ .

The obvious disadvantage of this method is the neglectation of all the possible interactions of other variables in the explanation of the dependency of a particular pair, i.e. this method is not able to capture complex patterns.

3.5.3 Tri-graph

The Tri-graph (TG) (also called 0 – 1 conditional independence graph) is developed from the combination of the previous two methods, and it can be seen “somehow” between them (see [75] and [76]).

To evaluate the dependency of a couple of variables X_i and X_j instead of considering jointly all other variables, just a third one X_k is considered at a time, but the index k is varied in the whole set of variables, i.e. $k \in \{1 \dots p\} \setminus \{i, j\}$, so that all the possible choices for the third variable will be iteratively considered. The name Tri-graph refers indeed to the fact that triples of variables are considered.

With this procedure all the variables are considered to check the dependencies, but one at a time and not all simultaneously as in the FCIG.

What is explored is the set of values of the **pairwise partial correlation coefficients**, that are defined as

$$(3.3) \quad \omega_{ijk} = \frac{\rho_{ij} - \rho_{ik}\rho_{kj}}{\sqrt{(1 - \rho_{ik}^2)(1 - \rho_{kj}^2)}}$$

The pairwise partial correlation coefficient expresses the pairwise values of the correlation between variables X_i , X_j and X_k and its support is the interval $[-1, 1]$. It is also called first order correlation (see [2] and [74]). Thus we can use partial correlation to distinguish between the correlation between two variables due to direct casual relationship from the correlation between the same two variables that originate via intermediate variables or is directly due to other variables. If there is no difference between w_{ijk} and ρ_{ij} then we can infer that the variable X_k has no effect. If the partial correlation approaches to zero, the inference is that the original correlation between X_i and X_j is spurious, i.e. there is no direct causal link between the two original variables because the control variable is either common antecedent cause or intermediate variable. If the pairwise partial correlation coefficient is equal to zero, then the variable X_k is able to explain the correlation between X_i and X_j . In fact in this case the correlation between X_i and X_j (ρ_{ij}) can be expressed as the product of the correlation between X_i and X_k (ρ_{ik}) and X_j and X_k (ρ_{jk}), i.e. $\rho_{ij} = \rho_{ik}\rho_{jk}$. Hence the variable X_k separates the variable X_i and X_j , i.e. $X_i \perp\!\!\!\perp X_j | X_k$. Partial correlation that remains significantly different from zero, may be taken as indicators of a possible causal link. The definition of the TG can be formalized as:

Definition 3.8 *TG draws an edge $X_i \rightarrow X_j$ if and only if $\rho_{ij} \neq 0$ and $\omega_{ijk} \neq 0$ considering every $k \in \{1 \dots p\} \setminus \{i, j\}$.*

This definition can be re-read as: an edge between two variables X_i and X_j is drawn if and only if there does not exist any other variable X_k that can explain this dependency.

Through the definition of the set $T_{ij} = \{\rho_{ij}\} \cup \{\omega_{ijk} \forall k \in \{1 \dots p\} \setminus \{i, j\}\}$ we can rewrite Definition 3.8 as

Definition 3.9 *TG draws an edge $X_i \rightarrow X_j$ if and only if $\tau = \operatorname{argmin}_{\tau_{ij} \in T_{ij}} |\tau_{ij}| \neq 0$*

Since this method is based on the covariance matrix (like the Covariance Graph), it is possible to use it even if the sample of data is small compared to the number of variables. Moreover it has the advantage to examine more complex correlation structures than the direct correlation.

3.5.4 Some examples

To see how the three methods work, some small examples are introduced here.

Example 3.10

We consider five variables X_i with $i \in \{1 \dots 5\}$. Their multivariate normal distribution is characterized by the following covariance matrix

$$\Sigma = \begin{pmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 2 & 1 & 1 & 1 \\ -1 & 1 & 2 & 1 & 1 \\ -1 & 1 & 1 & 2 & 1 \\ -1 & 1 & 1 & 1 & 2 \end{pmatrix}$$

Since all the nondiagonal elements of the matrix are different from zero, the Covariance Graph will connect every couple of variables.

The Full Conditional Independence Graph will give instead as a result a sort of “fork”, connecting just the node X_1 with all the other variables, since the corresponding precision matrix is:

$$\Omega = \begin{pmatrix} 5 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The CG and the FCIG are represented in Figure 3.3.

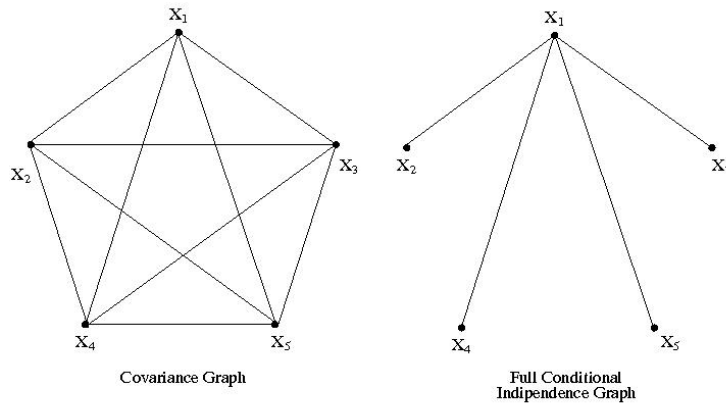


Figure 3.3: Example without cycles in the Full Conditional Independence Graph

To apply the Tri-graph method, we recall Definitions (3.3) and (3.8) and we compute the value of the pairwise partial correlation coefficients:

(X_i, X_j)	X_k	ω_{ijk}	Edge
(1,2)	$\forall k$	$\neq 0$	Yes
(1,3)	$\forall k$	$\neq 0$	Yes
(1,4)	$\forall k$	$\neq 0$	Yes
(1,5)	$\forall k$	$\neq 0$	Yes
(2,3)	1	0	NO

(X_i, X_j)	X_k	ω_{ijk}	Edge
(2,4)	1	0	NO
(2,5)	1	0	NO
(3,4)	1	0	NO
(3,5)	1	0	NO
(4,5)	1	0	NO

Result: The TG coincides with the FCIG.

◁

This is an excellent result, and we would like to get such a one-to-one correspondence between TG and FCIG every time, however this is not always the case. Let us consider another example.

Example 3.11

We slightly change the precision matrix of the previous example:

$$\Omega' = \begin{pmatrix} 5 & 0.5 & 0.5 & 1 & 1 \\ 0.5 & 1.5 & -0.5 & 0 & 0 \\ 0.5 & -0.5 & 1.5 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The covariance matrix becomes

$$\Sigma' = \begin{pmatrix} 0.4 & -0.2 & -0.2 & -0.4 & -0.4 \\ -0.2 & 0.85 & 0.35 & 0.2 & 0.2 \\ 0.4 & 0.35 & 0.85 & 0.2 & 0.2 \\ -0.4 & 0.2 & 0.2 & 1.4 & 0.4 \\ -0.4 & 0.2 & 0.2 & 0.4 & 1.4 \end{pmatrix}$$

The corresponding graphs for the CG and the FCIG are shown in Figure 3.4.

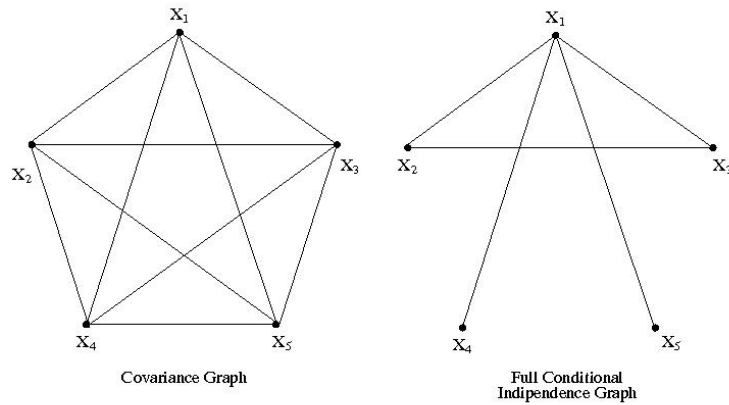


Figure 3.4: Example with cycles in the Full Conditional Independence Graph

Evaluating the pairwise partial correlation coefficients we get:

(X_i, X_j)	X_k	$\omega_{ij k}$	Edge
(1,2)	$\forall k$	$\neq 0$	Yes
(1,3)	$\forall k$	$\neq 0$	Yes
(1,4)	$\forall k$	$\neq 0$	Yes
(1,5)	$\forall k$	$\neq 0$	Yes
(2,3)	$\forall k$	$\neq 0$	Yes

(X_i, X_j)	X_k	$\omega_{ij k}$	Edge
(2,4)	$\forall k$	$\neq 0$	Yes
(2,5)	$\forall k$	$\neq 0$	Yes
(3,4)	$\forall k$	$\neq 0$	Yes
(3,5)	$\forall k$	$\neq 0$	Yes
(4,5)	$\forall k$	$\neq 0$	Yes

Result: The TG coincides with the CG.

◁

In both examples the Full Conditional Independence Graph is a subgraph of the Covariance Graph, but this is not always the case.

Example 3.12

We consider five variables X_i with $i \in \{1 \dots 5\}$. Their multivariate normal distribution is characterized by the following covariance matrix

$$\Sigma = \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 2 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & 2 & 0 \\ -1 & 0 & 0 & 0 & 2 \end{pmatrix}$$

The resulting Covariance Graph is a fork connecting the first variable, X_1 , with all other four variables (see Figure 3.5). The corresponding precision matrix is dense.

$$\Omega = \begin{pmatrix} 0.500 & 0.250 & 0.250 & 0.250 & 0.250 \\ 0.250 & 0.625 & 0.125 & 0.125 & 0.125 \\ 0.250 & 0.125 & 0.625 & 0.125 & 0.125 \\ 0.250 & 0.125 & 0.125 & 0.625 & 0.125 \\ 0.250 & 0.125 & 0.125 & 0.125 & 0.625 \end{pmatrix}$$

Hence in the Full Conditional Independence Graph all possible edges are drawn.

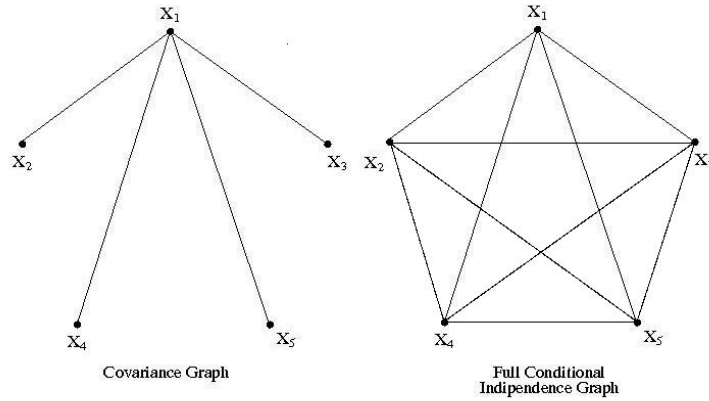


Figure 3.5: Example of Covariance Graph as subgraph of Full Conditional Independence Graph

If we apply the Tri-graph procedure we get the following results:

(X_i, X_j)	X_k	ω_{ijk}	Edge
(1,2)	$\forall k$	$\neq 0$	Yes
(1,3)	$\forall k$	$\neq 0$	Yes
(1,4)	$\forall k$	$\neq 0$	Yes
(1,5)	$\forall k$	$\neq 0$	Yes
(2,3)	1	0	NO

(X_i, X_j)	X_k	ω_{ijk}	Edge
(2,4)	1	0	NO
(2,5)	1	0	NO
(3,4)	1	0	NO
(3,5)	1	0	NO
(4,5)	1	0	NO

Result: the TG does coincide with the CG and both graphs contain less edges than the FCIG. ◁

Remark 3.13 The Tri-graph is always a subgraph of the Covariance Graph.

Before we move on investigating deeply the question in which instances the TG coincides with the FCIG, we examine if these theoretical examples correspond to real cases of the railway network.

Example 3.14

Two stations v and u are connected by a single track line without any overlapping point. Two trains t and s are scheduled to travel in different directions along this line. Figure 3.6 shows the Activity-on-arc Project Network of the problem.

We suppose that train t leaves the station v with some delay (either due to a source delay at the station v , e.g. longer time for the boarding and deboarding of the passengers, or due to previous forced delays). Train s has to wait at station u until the arrival

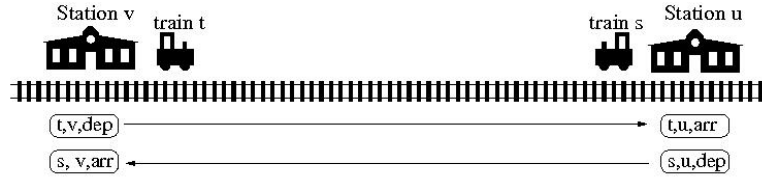


Figure 3.6: single track line

of train t (to receive the green light to proceed) since along the line there are no points where the two trains can travel simultaneously. Train s will have a forced delay, that might spread out along its journey.

We consider four variables corresponding to the delays at:

- Y_1 the departure of train t from station v ;
- Y_2 the arrival of train t at station u ;
- Y_3 the departure of train s from station u ;
- Y_4 the arrival of train s at station v .

The following covariance matrix refers to such a situation with four events corresponding to a single track line (between Seesen and Salz-Ringelheim) in the data file that the Deutsche Bahn put on our disposal to test the Tri-graph procedure. More details about the source file and the considered events are given in Chapter 4.

$$\Sigma = \begin{pmatrix} 3.243477 & 2.967419 & 2.903797 & 1.812586 \\ 2.967419 & 3.133233 & 3.095565 & 1.842816 \\ 2.903797 & 3.095565 & 3.833862 & 2.287471 \\ 1.812586 & 1.842816 & 2.287471 & 1.610345 \end{pmatrix}$$

Accordingly the precision matrix of the problem is:

$$\Omega = \begin{pmatrix} 2.4797356 & -2.4423586 & 0.6007953 & -0.8496433 \\ -2.4423586 & 3.9835325 & -1.8817665 & 0.8635136 \\ 0.6007953 & -1.8817665 & 2.9109165 & -2.6577422 \\ -0.8496433 & 0.8635136 & -2.6577422 & 4.3644464 \end{pmatrix}$$

For the Tri-graph method, we compute the value of the pairwise partial correlation coefficients:

(X_i, X_j)	X_k	ω_{ijkl}	Edge
(1,2)	$\forall k$	$\neq 0$	Yes
(1,3)	2	$= 0$	NO
(1,4)	3	$= 0$	NO

(X_i, X_j)	X_k	ω_{ijkl}	Edge
(2,3)	$\forall k$	$\neq 0$	Yes
(2,4)	3	$= 0$	NO
(3,4)	$\forall k$	$\neq 0$	Yes

Since we are considering real delay measurement, we used a Maximum Likelihood test (see Appendix 7) to check if the elements of the covariance/precision matrix and the pairwise partial correlation coefficient are zero. The quantile was set to 1%.



Figure 3.7: Tri-graph coincides with FCIG

The results of the three methods are shown in Figure 3.7.

Result: The TG coincides with the FCIG.

◁

Example 3.15

We consider three stations v, u, w among which three trains r, s and t are traveling, such that the corresponding Activity-on-arc Project Network is the one shown in Figure 3.8.

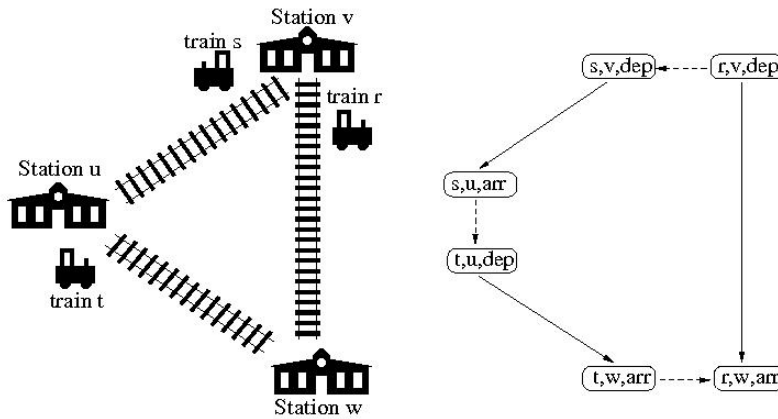


Figure 3.8: Triangular connection

These three trains are pairwise connected at the three stations:

- r and s at station v ;
- s and t at station u ;
- t and r at station w .

Passengers have two possibilities to travel between v and w : either with the direct train or using the connection between the other two trains of the system. E.g. to travel from station v to w a passenger can take either train r or the connection between s and t at station u . In reality such a situation occurs when it is necessary to have a high frequency of trains among some stations but it is not possible to increase the number of vehicles on duty.

We consider six variables corresponding to the delays at:

- Y_1 the departure of train r from station v ;
- Y_2 the departure of train s from station v ;
- Y_3 the arrival of train s at station u ;
- Y_4 the departure of train t from station u ;
- Y_5 the arrival of train t at station w ;
- Y_6 the arrival of train r at station w .

Supposing that the precision matrix Ω obtained from the delay data of the three trains is

$$\Omega = \begin{pmatrix} 1 & 0.5 & 0 & 0 & 0 & 1 \\ 0.5 & 4 & 1 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0.3 \\ 1 & 0 & 0 & 0 & 0.3 & 2 \end{pmatrix}$$

the corresponding covariance matrix will be dense, thus the CG will give the complete set of possible edges among the six variables as result, while the FCIG contains just the edges of a cycle.

Evaluating the pairwise partial correlation coefficients for the Tri-graph we get:

(X_i, X_j)	X_k	ω_{ijk}	Edge
(1,2)	$\forall k$	$\neq 0$	Yes
(1,3)	$\forall k$	$\neq 0$	Yes
(1,4)	$\forall k$	$\neq 0$	Yes
(1,5)	$\forall k$	$\neq 0$	Yes
(1,6)	$\forall k$	$\neq 0$	Yes
(2,3)	$\forall k$	$\neq 0$	Yes
(2,4)	$\forall k$	$\neq 0$	Yes
(2,5)	$\forall k$	$\neq 0$	Yes

(X_i, X_j)	X_k	ω_{ijk}	Edge
(2,6)	$\forall k$	$\neq 0$	Yes
(3,4)	$\forall k$	$\neq 0$	Yes
(3,5)	$\forall k$	$\neq 0$	Yes
(3,6)	$\forall k$	$\neq 0$	Yes
(4,5)	$\forall k$	$\neq 0$	Yes
(4,6)	$\forall k$	$\neq 0$	Yes
(5,6)	$\forall k$	$\neq 0$	Yes

The three corresponding graphs are shown in Figure 3.9.

Result: The TG coincides with the CG.

In this case, the Tri-graph method identifies all possible dependencies without being able to highlight the most important ones, as the FCIG does. \triangleleft

3.5.5 Some theoretical results

In this section we present some theoretical results that explain why we defined the TG as being positioned “somehow” in the middle between the CG and the FCIG.

Proposition 3.16 [76] *If the distribution on $X = (X_1, \dots, X_p)$ is Gaussian and faithful to the Full Conditional Independence Graph, then every edge in the Full Conditional Independence Graph is also an edge of the Tri-graph, $FCIG \subseteq TG$.*

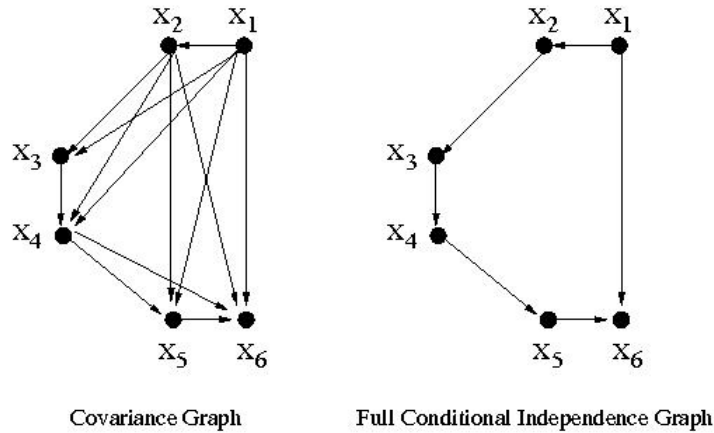


Figure 3.9: Tri-graph coincides with CG

Proof:

We prove the statement by contradiction. Assume that an edge (X_i, X_j) in the FCIG is not in the TG. Then either $\sigma_{ij} = 0$ or $\omega_{ij|k} = 0$ for some $k \in \{1, \dots, p\} \setminus \{i, j\}$.

In the first case ($\sigma_{ij} = 0$), X_i and X_j are (marginally) independent, i.e. they are independent given all the other variables (Thus there can not exist a path P from X_i to X_j (and viceversa) otherwise the variables in the paths could explain the dependence of X_i and X_j , $X_i \perp\!\!\!\perp X_j | \{\text{variables in the path } P\}$. Then we could generalize the set to $X_i \perp\!\!\!\perp X_j | \{X_k : k \in \{1, \dots, n\} \setminus \{i, j\}\}$, hence they will be dependent). Therefore they must be in different connectivity components of FCIG, since X is faithful to the FCIG. In the second case ($\omega_{ij|k} = 0$ for some $k \neq \{i, j\}$), it holds that $X_i \perp\!\!\!\perp X_j | X_k$. That can be generalized to $X_i \perp\!\!\!\perp X_j | \{X_k : k \in \{1, \dots, p\} \setminus \{i, j\}\}$ and since X is faithful to FCIG, it implies that X_i and X_j are separated in the FCIG, i.e. there is no edge (X_i, X_j) in the FCIG. \square

Proposition 3.17 [76] *Assume that the distribution of X is Gaussian. Moreover, assume that if X_i and X_j are not adjacent in the Full Conditional Independence Graph, then X_i and X_j are either in different connectivity components of the Full Conditional Independence Graph or there exists a vertex X_k that separates X_i and X_j in the Full Conditional Independence Graph. Then every edge in the Tri-graph is also an edge in the Full Conditional Independence Graph.*

Proof:

Assume that X_i and X_j are not adjacent in the Full Conditional Independence Graph. Then we either have:

- X_i and X_j are in different connectivity components. X_i and X_j are therefore (marginally) independent, which implies $\rho_{ij} = 0$ and there is no edge between them in the Tri-graph, or
- there exists some X_k with $k \in \{1, \dots, p\} \setminus \{i, j\}$ that separates X_i and X_j . Due to the Markov property, we have $X_i \perp\!\!\!\perp X_j | X_k$ and therefore $\omega_{ij|k} = 0$, which further implies that X_i and X_j are not adjacent in the Tri-graph.

 \square

Due to Propositions 3.16 and 3.17, the TG and FCIG may coincide. In particular all Gaussian distributions corresponding to trees are faithful (see [6]), so that it holds (see [76])

Theorem 3.18 [76] *If the Full Conditional Independence Graph of a Gaussian distribution is a forest of trees (i.e. the graph does not contain any cycle) then the Tri-graph and the Full Conditional Independence Graph coincide.*

Tri-graph and the Full Conditional Independence Graph do also coincide in more complicated scenarios, for example, if the distribution is Gaussian and faithful and if the corresponding Full Conditional Independence Graph consists of sets of cliques that (pairwise) share at most one (common) vertex.

From Proposition 3.16, we expect that sparse Full Conditional Independence Graphs have fewer edges than the Tri-graph. From Theorem 3.18 we also expect that the number of cycles will be an indicator for the difference between the number of edges in the FCIG and in the TG. The larger the number of cycles, the larger the difference in the number of edges.

As distributions are not always faithful, some FCIGs may also contain more edges than the corresponding TGs.

Furthermore we can prove some results concerning the relation between the Tri-graph and the Covariance Graph.

Proposition 3.19 *The Tri-graph is a subgraph of the Covariance Graph, $TG \subseteq CG$.*

Proof:

For every edge (X_i, X_j) in TG we have $\rho_{ij} \neq 0$, hence the edge belongs also to the CG. \square

Theorem 3.20 *If the Covariance Graph of a Gaussian distribution does not contain any cycle, then the Tri-graph coincides with the Covariance Graph*

Proof:

The CG has no cycle if there does not exist any triple of variables (i, j, k) such that $\sigma_{ij} \neq 0$, $\sigma_{ik} \neq 0$ and $\sigma_{jk} \neq 0$ (due to the transitivity of the covariance).

Proposition 3.19 already states that $TG \subseteq CG$. By contraddiction we suppose that $TG \subset CG$, i.e. there exists an edge (i, j) in CG but not in TG. Hence we have that $\sigma_{ij} \neq 0$ but there exists an index k such that $\omega_{ij|k} = 0$ where the pairwise partial correlation coefficient is defined as in (3.3). Writing explicitly the correlation coefficient as function in the variance and in the correlation of the variables, we get:

$$\begin{aligned} \omega_{ij|k} &= \frac{\frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} - \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}\sigma_{kk}}} \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}}}{\sqrt{\left(1 - \frac{\sigma_{ik}^2}{\sigma_{ii}\sigma_{kk}}\right)\left(1 - \frac{\sigma_{jk}^2}{\sigma_{jj}\sigma_{kk}}\right)}} \\ &= \frac{\frac{1}{\sqrt{\sigma_{ii}\sigma_{jj}}}\left(\sigma_{ij} - \frac{\sigma_{ik}\sigma_{jk}}{\sigma_{kk}}\right)}{\frac{1}{\sqrt{\sigma_{ii}\sigma_{jj}}}\sqrt{\frac{(\sigma_{ii}\sigma_{kk} - \sigma_{ik}^2)(\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2)}{\sigma_{kk}}}} \\ &= \frac{\sigma_{ij}\sigma_{kk} - \sigma_{ik}\sigma_{jk}}{\sqrt{(\sigma_{ii}\sigma_{kk} - \sigma_{ik}^2)(\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2)}} \end{aligned}$$

since the variance is a nonnegative quantity. Thus

$$\omega_{ij|k} = 0 \Leftrightarrow \sigma_{ij}\sigma_{kk} = \sigma_{ik}\sigma_{jk}$$

Since $\sigma_{ij} \neq 0$ (the edge exists in the CG) and $\sigma_{kk} \neq 0$ (otherwise the variable X_k would be a constant), it must be $\sigma_{ik}\sigma_{jk} \neq 0$, i.e. $\sigma_{ik} \neq 0$ and $\sigma_{jk} \neq 0$. Consequently the CG contains a cycle. \square

In general it is not possible to say which method between the CG and the FCIG will point out less edges since the presence of null non-diagonal elements in the covariance matrix does not imply the corresponding presence of null elements in the precision matrix.

We still need to decide how to check if the pairwise correlation coefficients corresponding to the different triples (X_i, X_j, X_k) are different than zero, i.e. $\omega_{ijk} = 0 \forall k \in \{1 \dots p\} \setminus \{i, j\}$.

Under the Gaussian assumption given in Section (3.5), this can be done with the Likelihood Ratio test (see Appendix A) based on the hypotheses:

- null hypothesis, $H_0(i, j|k) : \omega_{ijk} = 0$ for $k \in \{1 \dots p\} \setminus \{i, j\}$;
- alternative hypothesis, $H_1(i, j|k) : \omega_{ijk} \neq 0$ for $k \in \{1 \dots p\} \setminus \{i, j\}$.

Under the null hypothesis and the assumption that the data is independent identically distributed (i.i.d.) the log-likelihood test has an asymptotically Chi-squared distribution. We call P-value $P(i, j|k)$ the result of the Log-Likelihood Ratio test of the null hypothesis $H_0(i, j|k)$ versus the alternative hypothesis $H_1(i, j|k)$.

An analogous procedure is defined for the marginal correlation of the pair of variables (X_i, X_j) , so that we will also have:

- null hypothesis, $H_0(i, j|0) : \rho_{ij} = 0$;
- alternative hypothesis, $H_1(i, j|0) : \rho_{ij} \neq 0$;
- P-value, $P(i, j|0)$.

We can now reformulate Definition 3.9.

Definition 3.21 *We draw an edge $X_i \rightarrow X_j$ if and only if all the null hypotheses, $H_0(i, j|0)$ and $H_0(i, j|k) \forall k$, are rejected in all the Log-Likelihood Ratio tests.*

Thus there is evidence for an edge $X_i \rightarrow X_j$ if

$$\max_{k \in \{0\} \cup \{1 \dots p\} \setminus \{i, j\}} P(i, j|k) < \alpha$$

where α is the significance level for the test.

For deciding about a single edge between vertices X_i and X_j it is not necessary to correct α for the $p - 1$ multiple testings over all conditioning vertices k since the following proposition can be proved.

Proposition 3.22 [76] *Consider the single hypothesis (for some fixed pair (X_i, X_j)): $H_0(i, j)$: at least one $H_0(i, j|k^*)$ is true for some $k^* \in \{0\} \cup \{1 \dots p\} \setminus \{i, j\}$. Assume for all $k \in \{0\} \cup \{1 \dots p\} \setminus \{i, j\}$ the individual test satisfies*

$$\mathbb{P}_{\tilde{H}_0(i, j|k)}[H_0(i, j|k) \text{ rejected}] \leq \alpha$$

where $\tilde{H}_0(i, j|k) = \{H_0(i, j|k) \text{ true}\} \cup \{H_0(i, j|k') \text{ true or false (and compatible with } H_0(i, j|k) \text{ for all } k' \neq k)\}$. Then the error of the first type

$$\mathbb{P}_{H_0(i, j)}[H_0(i, j|k) \text{ are rejected for all } k \in \{0\} \cup \{1 \dots p\} \setminus \{i, j\}] \leq \alpha.$$

Proof: Consider the hypothesis:

$$H_0 = H_0(i, j) : \text{at least one } H_0(i, j|k^*) \text{ is true for some } k^*$$

The probability for an error of the first type is:

$$\begin{aligned} \mathbb{P}_{H_0}[H_0(i, j|k) \text{ rejected for all } k] &= \mathbb{P}_{H_0}[\cap_k \{H_0(i, j|k) \text{ rejected}\}] \\ &\leq \min_k \mathbb{P}_{H_0}[H_0(i, j|k) \text{ rejected}] \\ &\leq \mathbb{P}_{H_0}[H_0(i, j|k^*) \text{ rejected}] \\ &\leq \alpha. \end{aligned}$$

□

Remark 3.23 Due to this proposition a lot of calculations will be saved and it is also important to point out that the estimation for a Tri-graph is done in an exhaustively manner. This is a very important difference from the Full Conditional Independence Graph where it is often necessary to apply a non-exhaustive procedure in case of a huge graph space.

3.6 Statistical Multicorrection

In [75] it was suggested to apply a correction to the results of the Likelihood Ratio Test to reduce the error of first type: False Positive discovery rate (see Appendix 7). The recommendation was about the Benjamini-Hochberg (BH) approach as presented in [8] and [67]. The BH procedure for multiplying test is:

1. order the P-values arising from the m tests in a decreasing vector (of dimension m), so that $p[1] \geq p[2] \geq \dots \geq p[m]$;
2. do not correct the first component of the vector (i.e. the maximum P-value resulting from the test);
3. correct the other components by a factor directly proportional to the number of tests and inversionally proportional to the position of the P-value in the array, i.e. the corrected P-value will be $p^*[i] = \frac{m}{i} p[i]$ where $i \in \{2, \dots, m\}$;
4. compare the corrected P-values with the desired quantile α : if $p^*[i] \leq \alpha$ accept the alternative hypothesis (i.e. the estimated parameter is different than zero, thus an edge is drawn) else accept the null hypothesis (i.e. the estimated parameter is equal to zero, thus no edge is drawn).

This procedure is a generalization of the Bonferroni (B) correction, that is based on a fixed correction of the P-values, i.e. $p^*[i] = mp[i] \forall i \in \{1, \dots, m\}$.

In our problem we are more interested in reducing the error of second type (False Negative discovery rate), hence we tested the Tri-graph approach without correction in order to be able to identify as many “virtual” connections as possible. We compared this output with the results obtained by applying one of the previous corrections to confront the size of the sets. Moreover we checked if the increased number of edges (obtained without any corrections) corresponds also to an increased number of “virtual” activities.

Further on in this thesis we will refer to these three different corrections using the following abbreviations:

- – for the absence of a correction;
- *B* for the Bonferroni correction;
- *BH* for the Benjamini-Hochberg correction.

3.7 Tri-graph estimation algorithm

We can schematize the Tri-graph algorithm as follow.

1. compute the P-values $P(i, j|k)$ for all triples of variables (X_i, X_j, X_k) where $i, j \in \{1, \dots, p\}$, $i \neq j$ and $k \in \{0\} \cup \{1, \dots, p\} \setminus \{i, j\}$;
2. compute the maximum P-value for every pair $(X_i, X_j) = (X_j, X_i)$:

$$P_{max}(i, j) = \max_{k \in \{0\} \cup \{1, \dots, p\}} P(i, j|k);$$
3. (in case) correct the maximum P-values over the $\frac{p(1-p)}{2}$ multiple tests for all pairs of vertices, $P_{max,corr}(i, j)$;
4. draw an edge between X_i and X_j if and only if $P_{max,corr}(i, j) < \alpha$, where α is the pre-specified significance level.

3.8 Complexity of the Algorithm

To calculate the complexity of the Tri-graph method we start evaluating the complexity of the Mean of the Covariance (hence of the Variance) and of the correlation coefficients.

Mean

Given a random variable X with n observations, $\{x_1, \dots, x_n\}$, to calculate its mean it is necessary to sum all its observations and to divide this sum by the number of

observations, i.e. $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$.

Hence $n - 1$ additions and 1 division are executed for a total of n operations.

The complexity of the Mean is $O(n)$.

Covariance

Given two variables X and Y , each of them with n observations, to calculate the covariance it is necessary to sum the n products obtained by multiplying the deviation of the i^{th} observation of X from its mean \bar{X} with the corresponding deviation of the i^{th} observation of Y from its mean \bar{Y} and to divide the result by the number of observations we consider, i.e. $\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$.

We already evaluated the operations necessary for the mean, so we can neglect this calculations in the argumentation. $2n$ subtractions, n multiplications, $n - 1$ additions and 1 division (by n) have to be executed, in total $4n$ operations.

The complexity of the Covariance is $O(n)$.

Variance

The variance is a special case of the covariance, i.e. $X = Y$, and hence it necessitates less operations. The difference of the observations from the mean need to be evaluated only once. We have $3n$ operations.

The complexity of the Variance is $O(n)$.

Correlation Coefficient

The correlation coefficient of X and Y is obtained by dividing the covariance of the variables for the squared root of the product of their variances, i.e. $\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$.

Therefore a product, a square root and a division are computed, in total 3 operations (assuming that variance and covariance have already been calculated).

Consequently also the complexity of the Correlation Coefficient is $O(n)$.

If the Tri-graph is applied to p variables each of them provided with n observations (where n is much smaller than p) it is necessary to compute the mean and the variance of p variables (complexity $O(pn)$), and the covariance of $\frac{1}{2}(p-1)p$ pairs (complexity $O(np^2)$). In fact the covariance is symmetric ($\rho_{XY} = \rho_{YX}$), hence we can consider just the pair of variables (i, j) where i is smaller than j . Accordingly, for the first variable we compute its covariance regarding all the other $(p-1)$ variables, but for the second one we need just to consider $(p-2)$ pairs since its covariance with the first variable has already been evaluated. Analogous for the $(p-3)$ calculations regarding the third variable and so on until the second-last variable that can be paired just with the last

one. Hence we have $(p-1) + (p-2) + \dots + 1 = \sum_{r=1}^{p-1} r = \frac{1}{2}(p-1)p$ possible pairs.

To evaluate the covariance matrix we need $np + (3n)p + \frac{1}{2}(4n)(p-1)p = 2np^2 + 2np$ operations.

The resulting complexity is $O(np^2)$.

Likelihood test

To check if some non-diagonal elements of the covariance matrix or if some pairwise partial correlation coefficients are different than zero the Likelihood test is applied. It requires a fixed number of operations that we define as M and that is independent from the number of variables and from the number of their observations. Before proceeding we need to find out how many triples of variables (X_i, X_j, X_k) where $i, j, k \in \{1, \dots, p\}$, $i \neq j$ and $k \notin \{i, j\}$, can be considered starting from $\frac{1}{2}(p-1)p$ possible pairs.

It is important to underline that the triple (X_j, X_i, X_k) is equivalent to (X_i, X_j, X_k) , but both the triples (X_k, X_j, X_i) and (X_i, X_k, X_j) are not, so no matter which pair (i, j) we are considering, we always have $(p-2)$ choices for the third index k . Since the number of possible pairs is $\frac{1}{2}(p-1)p$ we have $\frac{1}{2}p(p-1)(p-2)$ possible triples. The total number of operations is equal to $\frac{M}{2}p(p-1)(p-2)$.

The complexity of the Likelihood test is $O(p^3)$.

Maximum

Given a $(p-1)$ -dim array containing the P-values arising from the test (one for the covariance and $(p-2)$ for all the possible triples that can be built from the considered pair), the maximum among them is calculated through a procedure that compares pairs of values (either the first P-value of the array with the second one or the maximum P-value found up to position i of the array with the $(i+1)$ -th element). The complexity of the Maximum is $O(p)$ in the worst case.

Statistical multi-correction

Finally a statistical multi-correction of the $\frac{1}{2}p(p-1)$ P-values (one for each pair (X_i, X_j)) is applied. This consists in multiplying the P-values by a factor that can be calculated with at most 2 operations. Thus we have a complexity of $O(p^2)$ in this part.

On the whole we have a complexity of $O(np^2)$ to evaluate the covariance matrix, a complexity of $O(p^3)$ for the Likelihood Ratio test and a complexity of $O(p^2)$ for the statistical multi-correction. Since both n and p are bigger than 1, the complexity of the statistical multi-correction is smaller than the previous ones, hence we can neglect it. As result we get that the complexity of the Tri-graph is either $O(p^3)$ or $O(np^2)$.

Since we assumed that p is much larger than n , the Tri-graph overall complexity is $O(p^3)$.

Numerical Results

The value of an idea lies in the using of it.

When I have fully decided that a result is worth getting, I go ahead of it and make trial after trial until it comes.

Just because something doesn't do what you planned it to do doesn't mean it's useless.

THOMAS ALVA EDISON (1847-1931)
American inventor and businessman

Computational issues must be resolved in order to prove the practical value of the Tri-graph approach in the identification of dependencies among train delays. In fact the presented properties of this method would remain incomplete for the practical problem if either it will be too difficult to implement or it will not point out the critical points of the railway network.

4.1 The DisKon Project

Since years the Deutsch Bahn (DB) has been collecting data about disturbances of the timetable, creating a quite huge data set. These information allow the study of the system as a whole (from both a theoretical and a practical point of view). Aim of the research is to find a way to improve the timetable minimizing the impact of the occurring delays on the average delay of the passengers. In particular one of the projects on which the DB embarked upon is the *DisKon - Disposition und Konfliktlösungsmanagement für die beste Bahn*. Challenge of the project is to model a system that is able to recognize the conflicts and to solve them in order to save resources and avoid discriminations. Different universities, among which Georg-August-Universität Göttingen, are collaborating on this project. The optimization group, lead by Prof. Schöbel at the Institute for Numerical and Applied Mathematics (NAM), considers a macro-formulation of the problem as a linear integer program (LIP). Focuses of the research are both a one-criterial formulation based on the average delay of the passenger as well as a bicriterial optimization based on one side on the delay of the vehicles and on the other on the total amount of missed connections. A Micromodel is considered to check the feasibility of the solutions.

Within the framework of this collaboration between the DB and the NAM Institute, the contribution of this thesis to the project is the analysis of the dependencies among train delays to identify the critical points of the railway system. These dependencies are then transformed into constraints of the Macro-formulation of the LIP in order to improve the robustness of the Macro-solution. This gave us the possibility to test the

Tri-graph method using real data.

4.1.1 The raw data

In Autumn 2005, the Deutsch Bahn placed at disposal of the Optimization Group at NAM a set of files containing measurements of real departure and arrival times of regional trains in the following train stations:

- Bad Harzburg;
- Goslar;
- Herzberg;
- Oker;
- Salzgitter-Ringelheim;
- Seesen;
- Vieneburg;
- Wolfenbüttel.

These stations are located in the Harz region, a mountain range in northern Germany that straddles the border between the states of Lower Saxony, Saxony-Anhalt and Thuringia (see Figure 4.1 and Figure 4.2).

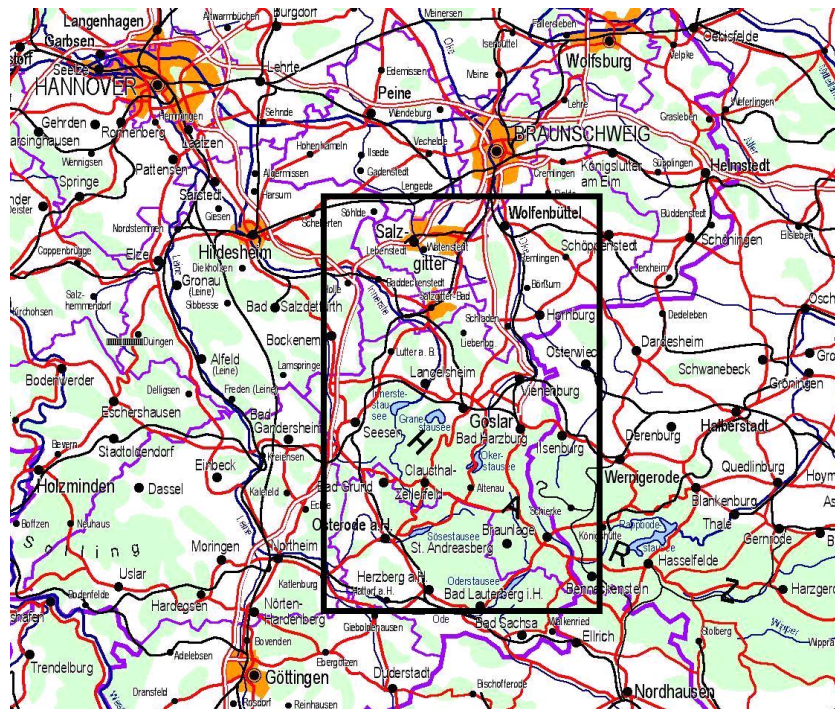


Figure 4.1: Harz area

The data are collected in tabular files each of them referring to arrivals and departures in one of the stations over listed inside a three-month period. Altogether the files

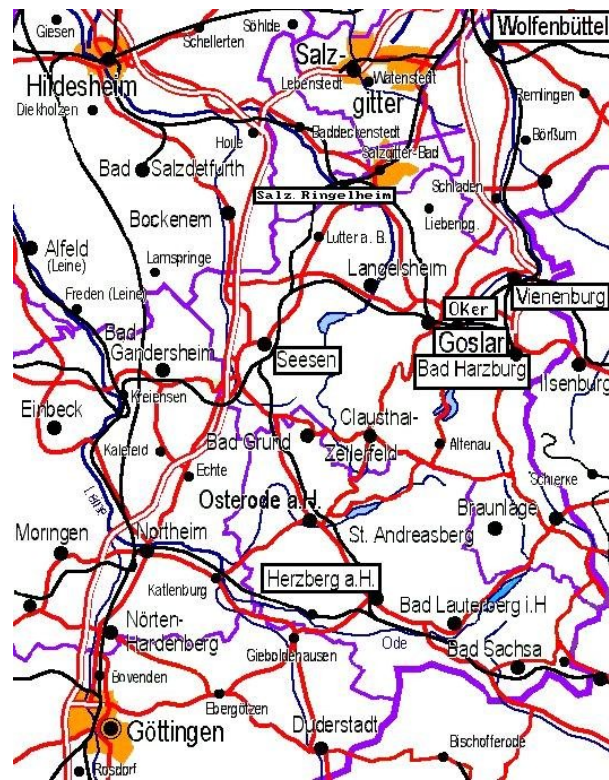


Figure 4.2: Location of the considered stations

correspond to a time window of nine months, between Saturday 1st January and Friday 30th September 2005 (exactly 39 weeks).

The tabular files contain information concerning:

- the identifying abbreviation of the station (e.g. HBHA for “*Hauptbahnhof Bad Harzburg*”);
- the kind of event (i.e. Arrival or Departure);
- the day in which the event has been measured;
- the class of the train (i.e. a number specifying the type of the train: ICE, EC, IC, RB ...);
- the subclass of the train (i.e. a number specifying if the train is regular/special, if it transports passengers or it is empty ...);
- the identification number of the train;
- the scheduled time of the event;
- the measured (“real”) time in which the event took place;
- the delay, in minutes, of the event (i.e. the difference between the measured and the scheduled time of an event);
- the direction in which the train is traveling (i.e. its next scheduled station).

Peculiarity of the file is to contain in its first half all the arrival events and in its second half all the departure events, both groups are chronologically ordered.

4.1.2 Working with the data

As we started to select the events that could be considered as input for the Tri-graph approach, we noticed some unexpected peculiarities of the data files.

Firstly we observed that the registrations contained in the files correspond to different classes of trains (principally local and freight trains) traveling (with or without passengers/cargo) on the considered track system. Freight trains and empty local trains (i.e. local trains without passengers that have to be moved from one station to another due either to logistic needs or maintenance) do not have to follow a fixed timetable but they are allowed to proceed on a track whenever it is free or they are forced to wait longer in a station to avoid disturbances on the regular traffic. The variance corresponding to their traveling times is very large, hence they are very difficult to model. Therefore choosing the sample for the Tri-graph approach, we decided to consider just events corresponding to passenger trains.

Afterwards we noticed that the measurements are precise just up to the minute. In fact the departures or arrivals of some trains are correctly registered in the hour and minute fields, but the field corresponding to the seconds, out of the 60 possible values, assume just a few of them. Many trains always arrive or depart in perfect time, i.e. the difference in seconds between the registered time and the scheduled one is constantly zero. A constant value for the “seconds field” is registered also in case of delay, no matter the magnitude of it in minutes. Nevertheless we decided to calculate the delays in seconds, instead of minutes as given in the files, in order to profit of possible additional information.

Then we noted that in two stations (Oker and Wolfenbüttel) some events were strangely registered: the same scheduled timetable was considered for both the arrival and the departure of some trains. The personal of the Deutsche Bahn confirmed that these sets of trains were not scheduled to stop in these stations. Thus we considered the data related to the arrival of the trains (i.e. the ones reported in the first part of the data files) as intermediate measurements of the delays and marked these events as “special”, defining a (sort of) third kind of events.

Moreover, counting the occurrence of every event in the whole time window we noticed double registrations of the same event. E.g. events corresponding to trains traveling every working day (in the considered time window of 39 weeks) were occurring more than 195 times where $195 = 39 \cdot 5$ (5 working days per week). In that case we decided to consider just the first registration of the events per day, considering the following ones as a repetition, even if their measured delays were not exactly the same.

In contrast with the previous point, for other events we had to face missing registrations either of some fields of the file or of the whole event. We decided not to generate any artificial data to avoid possible influences on the identification of the dependencies and we preferred not to consider these events as suitable samples.

Finally it is important to highlight that the timetable on the weekends and festivities is slightly different than during the week, so we focused our attention principally on the working days (i.e. from Monday to Friday) excluding all the festivities (Saturdays and Sundays have been separately considered, see Section 4.4).

For the record we considered as national holidays (based on the Lower-Saxony calendar) the following days:

- Saturday 1st January 2005 (New Year's Day);
- Friday 25th March 2005 (Good Friday);
- Sunday 27th March 2005 (Easter);
- Monday 28th March 2005 (Easter Monday);
- Sunday 1st May 2005 (Labour Day);
- Thursday 5th May 2005 (Ascension Day);
- Sunday 15th May 2005 (Whit-Sunday);
- Monday 16th May 2005 (Whit-Monday).

Summarizing we considered as possible trains for the sample:

- local trains
- transporting passengers
- traveling on weekdays
- on the Harz area
- in a time window of 9 months (January-September 2005).

In accordance with this list, we decided to proceed with two different strategies: on one side we decided to maximize the number of events we could consider having a reasonable amount of data for each of them, on the other side we maximize the number of occurrences per event keeping a reasonable chain of events.

Concerning the first strategy we consider a set of data pro event with a cardinality equal to 30 reasonable. We found a sample of 928 events corresponding to 229 trains, which consists of 358 waiting activities and 339 “driving” activities (the quotation marks are used since some events in the chain are missing). Since there are no festivity days falling on Tuesday and Wednesday, we mainly focused our attention on those days, choosing this sample. From now on we will refer to this sample as $W - 30$, abbreviation of “weekdays - 30 observations”.

On the contrary, for the second strategy we wanted to have the set of events registered every weekday on the time window we got (i.e. 195 occurrences per event). We found a set of 440 events corresponding to 118 trains, which consists of 161 waiting activities and 152 “driving” activities (again the quotation marks refer to the absence of some events in the chain). From now on we will refer to this sample as $W - 195$, abbreviation of “weekdays - 195 observations”.

	events	trains	# observations
$W - 30$	928	229	30
$W - 195$	440	118	195

Table 4.1: Characteristic of the samples $W - 30$ and $W - 195$

4.2 Outputs of the samples $W - 30$ and $W - 195$

We started applying the Chi-Squared test on a Contingency Table (CT) (see Section 3.3) to the two samples introduced in the previous section, ($W - 30$ and $W - 195$). Firstly we applied the classical version of the test (no multistatistical correction), in which the values of two variables are registered on the single Contingency Table on which the test is based. All the possible pairs (X_i, X_j) in which j is bigger than i have been tested. In fact i and j are the positions of the two events in the sample ordered according to the scheduled timetable. Hence it would have no meaning to check if an event could influence previously happened events. This method can be directly compared with the Covariance Graph (CG) approach.

We tested three different values for the quantile, $\alpha \in \{0.01, 0.05, 0.10\}$. Since the test is thought to check the independence among variables, the values reported in the tables are upper bounds of the possible dependency edges, obtained by subtracting at the amount of possible pairs the cardinality of the pointed out set of independent pairs of variables. We do refer to these highlighted pairs of dependent events as edges since this term will be used for the output of graphical models (as explained in Section 3.4). The results are three huge sets of edges, most of them due to the transitivity property. Since the test has been thought to identify independencies, smaller values of the quantile correspond to a higher number of rejected independence hypotheses. That does not directly imply a higher number of dependencies among the variables but gives us a rough approximation (Upper Bound also for CG) on the possible amount of them.

Reg. Trains - weekdays - 30 data				Reg. Trains - weekdays - 195 data			
Corr.	α	Nr. Edges	%	Corr.	α	Nr. Edges	%
-	0.01	496 500	57.72	-	0.01	124 512	64.46
-	0.05	504 813	58.68	-	0.05	131 767	68.22
-	0.10	555 410	64.56	-	0.10	149 209	77.25
possible edges		860 256	100	possible edges		193 160	100

Table 4.2: Outputs of the Contingency Table test for the samples $W - 30$ and $W - 195$

Afterwards we applied a new version of the test based on triples of variables as explained in Section 3.3. A pair of variables (X_i, X_j) with i smaller than j , will be considered to be dependent not only if the Chi-Squared test on the direct contingency table rejects the null hypothesis, but also if there does not exist any third variable that can explain a dependence between them. Hence moving from the pairs of variables (X_i, X_j) pointed out by the classic contingency table method, we considered a third variable X_k with index $k \in \{1, \dots, p\} \setminus \{i, j\}$ (where p is the total number of variables). The observations of the pair (X_i, X_j) are hence registered in two contingency table P_k and D_k as explained in Section 3.3. The Chi-Squared test is then applied to both tables. The variables X_i, X_j are considered independent given X_k if the null hypothesis is accepted in both the tests (one for P_k and one for D_k), otherwise we do consider them dependent given X_k .

In other words we consider the variables X_i and X_j independent given all other variables if there does not exist any variable $X_k \forall k \in \{1, \dots, n\} \setminus \{i, j\}$, such that the variables X_i and X_j can be defined as dependent given X_k . This method should be directly comparable with the Tri-graph approach. However the output of this procedure was empty, independently from the value of the quantile. Therefore we prefer to neglect this procedure and to concentrate ourselves on the graphical models presented in Section 3.5.

The application of the FCIG method (see Section 3.5.1) was not possible since the

observations corresponding to many pairs of variables, in particular the pairs of variables defining waiting activities, were strongly linearly dependent and hence it was numerically impossible to evaluate the inverse of the covariance matrix. This strong linear dependency is due to the nature of the events (waiting activities are usually characterized by small slack times, hence the value of the departure delay often corresponds to the value of the arrival delay of the train) and/or to a measurement procedure that is exactly just up to minutes (as already highlighted in the previous section).

Applying the CG method (see Section 3.5.2) we had to face again the problem of the transitivity property even if in a weaker form as in the Contingency Table test, as it can be seen in Table 4.2.

Reg. Trains - weekdays - 30 data				Reg. Trains - weekdays - 195 data			
Corr.	α	Nr. Edges	%	Corr.	α	Nr. Edges	%
-	0.01	27 385	3.18	-	0.01	7 390	3.83
	0.05	45 019	5.23		0.05	11 521	5.96
	0.10	59 911	6.96		0.10	14 981	7.76
possible edges		860 256	100	possible edges		193 160	100

Table 4.3: Output of the Covariance Graph for the samples $W - 30$ and $W - 195$

Finally we applied the TG approach. Its output, i.e. the identified dependent pairs of variables, has been subdivided into five groups, corresponding to the kind of activity they were representing:

- wait - waiting activities;
- drive - driving activities;
- drive2 - activities identified by pairs of events corresponding to the same train, that are not exactly waiting or driving activities due to the loss of some events of the journey sequence in the considered list of events;
- virtual - virtual activities, characterized by a (scheduled) time difference smaller than 45 minutes;
- error - unclear activities characterized by a (scheduled) time difference bigger than 45 minutes.

The considered Harz area is quite small, 45 minutes are a little bit more than the scheduled traveling time between the two remotest stations, hence we can consider this value as a period for our system. It is also important to remark that we do not have any information about the scheduled connections between trains, thus we may have included some change activities as “virtual” activities.

The results of the Tri-graph for the two samples are schematized in Tables 4.4 and 4.5, in which the percentages of edges corresponding to the five groups defined above are also reported. It is remarkable that the Tri-graph applied to the second sample, i.e. the one with more observations, is able to identify more than 85% of the waiting activities (with a peak of 95.8 in the case of no correction and a quantile $\alpha = 0.10$). Concerning the driving activities the percentage is not so high also because the lack of (data for) some (arrival/departure) events in the train journey reduces the effect of the spread of the delays (moreover there are some edges due to transitivity). The lowest peak corresponds to 34.6% of identified driving activities and the highest peak to 90.8%.

Regional Trains - weekdays - 30 data												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		virtual		error	
			#	%	#	%	#	%	#	%	#	%
-	0.01	463	203	43.9	77	16.6	40	8.7	50	10.8	93	20.1
	0.05	753	216	28.7	104	13.8	82	10.9	84	11.2	267	35.4
	0.10	1 085	228	21.1	131	12.2	118	11.0	135	12.1	473	43.6
B	0.01	94	77	82.0	11	11.7	0	0.0	2	2.1	4	4.2
	0.05	106	84	79.2	14	13.2	0	0.0	4	3.8	4	3.8
	0.10	115	91	79.1	15	13.0	0	0.0	5	4.4	4	3.5
BH	0.01	150	115	76.7	20	13.3	1	0.7	9	6.0	5	3.3
	0.05	182	131	72.0	25	13.7	4	2.2	10	5.5	12	6.6
	0.10	190	135	71.1	25	13.2	5	2.6	11	5.8	14	7.3
possible edges		860 256										

Table 4.4: Output of the Tri-graph for the sample $W - 30$

Regional Trains - weekdays - 195 data												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		virtual		error	
			#	%	#	%	#	%	#	%	#	%
-	0.01	491	152	31.0	113	23.0	77	15.7	41	8.3	108	22.0
	0.05	745	156	21.0	131	17.6	111	14.9	67	9.0	280	37.5
	0.10	1 002	159	15.8	139	13.9	136	13.6	112	11.2	456	45.5
B	0.01	236	142	60.2	56	23.8	19	8.0	12	5.0	7	3.0
	0.05	249	144	57.8	62	24.9	22	8.9	12	4.8	9	3.6
	0.10	256	144	56.2	65	25.4	23	9.0	13	5.1	11	4.3
BH	0.01	286	146	51.0	76	26.6	34	11.9	17	6.0	13	4.5
	0.05	321	147	45.8	85	26.5	40	12.5	21	6.5	28	8.7
	0.10	342	148	43.2	91	26.6	45	13.2	24	7.0	34	10.0
possible edges		193 160										

Table 4.5: Output of the Tri-graph for the sample $W - 195$

When applied to the first sample, the Tri-graph is able to catch in the worst case 21.5% of the waiting activities and 3.2% of the driving activities, and in the best case 63.7% and 38.6% respectively.

In the tables it is clearly shown that the multiple statistical corrections reduce considerably the percentage of “errors” of the procedure, but they also noticeably reduce the amount of “virtual” connections pointed out.

This problem arises directly from the definition of multistatistical correction and it is an open question to decide in which cases it is necessary to apply this procedure and in which it will be better to avoid it (see [28]). Some researchers recommend adjusting the P-values to prevent the findings of falsely claiming “statistical significance” (e.g. [27]), others disagree with this strategy, because it is inappropriate and may cause incorrect conclusions from the study (e.g. [58]).

Adjustments to the P-values are found on the logic that if a null hypothesis is true, a significant difference may still be observed by chance, since just a sample and not the entire population is observed. The chance of finding at least one test statistically significant due to chance and of incorrectly declaring a difference increases proportionally with the number of statistically independent tests performed simultaneously. The opponents of P-value adjustments raise as objection that the significance of each test will be interpreted according to how many output measures are considered in the family-wise hypothesis, which has been defined ambiguously, arbitrarily and inconsistently by its advocates. The debate over the need for P-value adjustments focuses upon our ability to make distinctions between different results. To date, the issues that separate these two statistical fields remain unresolved. Moreover, other strategies may be used in lieu of P-value adjustment (see [53]). In conclusion we can only quote the word of [28] and [65]: *disagreements over the use of various approaches should not cause us to waver from our aim to produce valid and reliable research findings. There are no “royal” roads to good research, because in science we are never absolutely sure of anything.*

Therefore willing to improve the quality of the Macrosolution of our linear program, we prefer not to apply any statistical correction, getting in such a way the highest number of “virtual” constraints.

Table 4.6 summarizes the results of the four procedures without multistatistical correction.

-	RB - week - 30 data				RB - week - 195 data			
α	CT	CG	TG	FCIG	CT	CG	TG	FCIG
0.01	126.210	27.385	463	NA	52.626	7.390	491	NA
0.05	75.613	45.019	753	NA	35.344	11.521	745	NA
0.10	67.300	59.911	1085	NA	27.932	14.981	1002	NA

Table 4.6: Comparison of different methods on the samples $W - 30$ and $W - 195$

Even if it was not possible to do compare TG versus FCIG (the most interesting case from our point of view since FCIG points out the dependencies between couples of variables given all the other variables of the systems), it clearly results that the TG is much more efficient than the CT and the CG, pointing out just the most relevant dependencies.

Regional Trains - Mon-Wed-Fri - 117 data													
Correction	Quantile α	Nr. Edges #	activity										
			wait		drive		drive2		virtual		error		
			#	%	#	%	#	%	#	%	#	%	
-	0.01	443	146	33.0	91	20.5	55	12.4	40	9.0	111	25.1	
	0.05	656	152	23.2	111	16.9	79	12.0	66	10.1	248	37.8	
	0.10	893	153	17.1	121	13.5	101	11.3	91	10.2	427	47.8	
B	0.01	192	132	68.8	32	16.7	12	6.3	10	5.2	6	3.1	
	0.05	203	134	66.0	38	18.7	13	6.4	11	5.4	7	3.4	
	0.10	206	134	65.0	40	19.4	13	6.3	12	5.8	7	3.4	
BH	0.01	242	141	58.3	49	20.2	20	8.3	16	6.6	16	6.6	
	0.05	272	145	53.3	59	21.7	28	10.3	18	6.6	22	8.1	
	0.10	287	146	50.9	65	22.6	31	10.8	19	6.6	16	9.1	
possible edges		193 160											

Regional Trains - Tue-Thu - 78 data													
Correction	Quantile α	Nr. Edges #	activity										
			wait		drive		drive2		virtual		error		
			#	%	#	%	#	%	#	%	#	%	
-	0.01	344	143	41.6	72	20.9	48	14.0	21	6.1	60	17.4	
	0.05	502	146	29.1	88	17.5	76	15.1	41	8.2	151	30.1	
	0.10	679	148	21.8	99	14.6	91	13.4	63	9.3	278	40.9	
B	0.01	156	120	76.9	27	17.3	5	3.2	2	1.3	2	1.3	
	0.05	167	125	74.9	29	17.4	8	4.8	2	1.2	3	1.8	
	0.10	171	126	73.7	31	18.1	8	4.7	2	1.2	4	2.3	
BH	0.01	195	133	68.2	37	19.0	14	7.2	4	2.1	7	3.6	
	0.05	211	135	64.0	40	19.0	18	8.5	7	3.3	11	5.2	
	0.10	218	136	62.4	43	19.7	19	8.7	9	4.1	11	5.0	
possible edges		193 160											

Regional Trains - Monday - 39 data													
Correction	Quantile α	Nr. Edges #	activity										
			wait		drive		drive2		virtual		error		
			#	%	#	%	#	%	#	%	#	%	
-	0.01	253	128	50.6	39	15.4	27	10.7	17	6.7	42	16.6	
	0.05	390	134	34.4	59	15.1	46	11.8	32	8.2	119	30.5	
	0.10	545	136	25.0	67	12.3	64	11.7	52	9.5	226	41.5	
B	0.01	74	68	91.9	3	4.1	0	0.0	1	1.4	2	2.7	
	0.05	85	75	88.2	5	5.9	1	1.2	1	1.2	3	3.5	
	0.10	91	78	85.7	6	6.6	2	2.2	1	1.1	4	4.4	
BH	0.01	107	89	83.2	7	6.5	1	5.6	1	0.9	4	3.7	
	0.05	128	100	78.1	10	7.8	3	6.3	3	2.3	7	5.5	
	0.10	136	105	77.2	12	8.8	3	6.6	3	2.2	7	5.1	
possible edges		193 160											

Table 4.7: Outputs of the Tri-graph for the samples $MWF - 117$, $TT - 78$ and $Mo - 39$

4.3 Outputs of some subsamples of $W - 195$

Moving from the dataset of the second strategy (Regional Train traveling every weekday of the year, i.e. W_{195} data), we decided to test the Tri-graph approach on data-subsets, to see how the number of observations can influence the outputs.

Initially we wanted to test the Tri-graph considering half and one quarter of the original dataset, however, since the data are obtained by collecting the observations of five weekdays, we further decided to consider three fifths, two fifths and one fifth of them. In such a way we could have an easier rule to select the samples, namely we could consider the observations corresponding to three, two or a single weekday(s). We chose the following set of days to define the subsets:

- {Monday,Wednesday,Friday} (referred as sample $MWF - 117$, abbreviation of “Monday Wednesday Friday - 117 observations”);
- {Tuesday,Thursday} (referred as sample $TT - 78$, abbreviation of “Tuesday Thursday - 78 observations”);
- {Monday} (referred as sample $Mo - 39$, abbreviation of “Monday - 39 observations”).

The comparison between the outputs of the sample $W - 195$ and its three subsamples is schematized in Table 4.8. The percentage values show how many percent of the edges pointed out using the dataset $W - 195$ (“reference edges”) have been found. It is not possible to prove that all the edges pointed out by a subset belong also to the reference set (since a connection between two trains can be guaranteed just on few weekdays, e.g. Monday and Friday to satisfy the need of weekly commuters) but in our checks we did not find any counter example to this statement.

Regional Trains - weekdays								
Corr.	α	$W - 195$	Mon-Wed-Fri		Tue-Thu		Monday	
		# Edges	# Edges	%	# Edges	%	# Edges	%
-	0.01	491	443	98.7	344	70.1	253	51.5
	0.05	745	656	88.1	502	67.4	390	52.3
	0.10	1002	893	89.1	679	89.1	545	54.5
B	0.01	236	192	81.3	156	66.1	74	31.4
	0.05	249	203	81.5	167	67.1	85	34.1
	0.10	256	206	80.5	171	66.8	91	35.5
BH	0.01	286	242	84.6	195	68.2	107	37.4
	0.05	321	272	84.7	211	65.7	128	39.9
	0.10	342	287	83.9	218	63.7	136	39.8
possible edges		193 160						

Table 4.8: Comparison of Tri-graph outputs for the samples $W - 195$, $MWF - 117$, $TT - 78$ and $Mo - 39$

However we are not interested in the number of edges that the Tri-graph highlights. Principally we want to consider the virtual connections that will be pointed out, since these edges are the ones that will be transformed into new constraints for the timetable problem. In Table 4.15 we summarize the results concerning the “virtual” activities. The percentage values show how many percent of the “virtual” edges pointed out using the dataset $W - 195$ (“reference virtual edges”) have been found.

Regional Trains - Tuesday - 39 data												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		virtual		error	
			#	%	#	%	#	%	#	%	#	%
-	0.01	231	128	55.4	40	17.3	16	6.9	9	3.9	38	16.5
	0.05	358	133	37.2	52	14.5	49	13.7	23	6.4	101	28.2
	0.10	493	139	28.3	64	13.0	68	13.8	39	7.9	183	37.1
B	0.01	93	82	88.2	10	10.8	0	0.0	0	0.0	1	1.1
	0.05	106	91	85.8	13	12.3	0	0.0	0	0.0	2	1.9
	0.10	110	92	83.6	16	14.5	0	0.0	0	0.0	2	1.8
BH	0.01	124	95	76.6	20	16.1	5	4.0	1	0.8	3	2.4
	0.05	143	107	74.8	24	16.8	7	4.9	1	0.7	4	2.8
	0.10	156	114	73.1	25	16.0	8	5.1	1	0.6	8	5.1
possible edges		193 160										

Table 4.9: Outputs of the Tri-graph for the sample $Tu - 39$

Regional Trains - Wednesday - 78 data												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		virtual		error	
			#	%	#	%	#	%	#	%	#	%
-	0.01	248	122	49.2	39	15.7	34	13.7	17	6.9	36	14.5
	0.05	365	128	35.1	54	14.8	55	15.1	34	9.3	94	25.8
	0.10	499	131	26.3	73	14.6	77	15.4	47	9.4	171	34.3
B	0.01	93	75	80.6	10	10.8	2	2.2	2	2.2	4	4.3
	0.05	101	79	78.2	12	11.9	3	3.0	2	2.0	5	5.0
	0.10	106	82	77.4	14	13.2	3	2.8	2	1.9	5	4.7
BH	0.01	120	91	75.8	16	13.3	3	2.5	5	4.2	5	4.2
	0.05	134	98	73.1	17	12.7	6	4.5	7	5.2	6	4.5
	0.10	139	99	71.2	20	14.4	6	4.3	7	5.0	7	5.0
possible edges		193 160										

Table 4.10: Output of the Tri-graph for the sample $We - 39$

Regional Trains - Thursday - 39 data													
Correction	Quantile α	Nr. Edges #	activity										
			wait		drive		drive2		virtual		error		
			#	%	#	%	#	%	#	%	#	%	
-	0.01	254	123	48.4	38	15.0	31	12.2	17	6.7	45	17.7	
	0.05	396	133	33.6	54	13.6	47	11.9	37	9.3	125	31.6	
	0.10	552	136	24.6	61	11.1	64	11.6	55	10.0	236	42.8	
B	0.01	70	63	90.0	4	5.7	1	1.4	1	1.4	1	1.4	
	0.05	86	77	89.5	5	5.8	1	1.2	1	1.2	2	2.3	
	0.10	93	82	88.2	7	7.5	1	1.1	1	1.1	2	2.2	
BH	0.01	108	93	86.1	8	7.4	2	1.9	1	0.9	4	3.7	
	0.05	129	103	79.8	14	10.9	3	2.3	3	2.3	6	4.7	
	0.10	136	104	76.5	18	13.2	5	3.7	3	2.2	6	4.4	
possible edges		193 160											

Table 4.11: Output of the Tri-graph for the sample $Th - 39$

Regional Trains - Friday - 39 data													
Correction	Quantile α	Nr. Edges #	activity										
			wait		drive		drive2		virtual		error		
			#	%	#	%	#	%	#	%	#	%	
-	0.01	257	124	48.2	45	17.5	28	10.9	21	8.2	39	15.2	
	0.05	367	129	35.1	64	17.4	43	11.7	25	6.8	106	28.9	
	0.10	510	130	25.5	72	14.1	60	11.8	44	8.6	204	40.0	
B	0.01	88	79	89.8	2	2.3	3	3.4	3	3.4	1	1.1	
	0.05	94	81	86.2	4	4.3	4	4.3	4	4.3	1	1.1	
	0.10	100	86	86.0	4	4.0	4	4.0	5	5.0	1	1.0	
BH	0.01	120	97	80.8	8	6.7	7	5.8	6	5.0	2	1.7	
	0.05	137	103	75.2	15	10.9	7	5.1	7	5.1	5	3.6	
	0.10	142	103	72.5	16	11.3	9	6.3	7	4.9	7	4.9	
possible edges		193 160											

Table 4.12: Output of the Tri-graph for the sample $Fr - 39$

Regional Trains - weekdays - 39 data												
Correction	Quantile α	Nr. Edges 195	total number of edges									
			Monday		Tuesday		Wednesday		Thursday		Friday	
			#	%	#	%	#	%	#	%	#	%
-	0.01	491	128	51.5	231	47.0	248	50.5	256	51.7	257	52.3
	0.05	745	134	53.3	358	48.1	365	49.0	396	53.2	367	49.3
	0.10	1 002	136	54.4	493	49.2	499	49.8	552	55.1	510	51.0
B	0.01	236	68	31.4	93	34.3	93	34.4	70	29.7	88	37.3
	0.05	249	75	34.1	106	42.6	101	40.1	89	35.7	94	37.8
	0.10	256	78	35.5	110	43.0	106	41.4	93	36.3	100	39.1
BH	0.01	286	89	37.4	124	43.4	120	42.0	108	37.8	120	42.0
	0.05	321	100	39.9	143	44.5	134	41.7	129	40.2	137	42.7
	0.10	342	105	39.8	156	45.6	139	40.6	136	39.8	142	41.5
possible edges		193 160										

Table 4.13: Comparison of the outputs of the Tri-graph for samples $W - 195$, $Mo - 39$, $Tu - 39$, $We - 39$, $Th - 39$ and $Fr - 39$

Regional Trains - weekdays - 39 data												
Correction	Quantile α	virtual 195	Virtual activities									
			Monday		Tuesday		Wednesday		Thursday		Friday	
			#	%	#	%	#	%	#	%	#	%
-	0.01	41	17	41.5	9	22.0	17	41.5	17	41.5	21	51.2
	0.05	67	32	47.8	23	34.3	34	50.7	37	55.2	25	37.3
	0.10	112	52	46.4	39	34.8	47	42.0	55	49.1	44	39.3
B	0.01	12	1	8.3	0	0.0	2	16.7	1	8.3	3	25.0
	0.05	12	1	8.3	0	0.0	2	16.7	1	8.3	4	33.3
	0.10	13	1	7.7	0	0.0	2	15.4	1	7.7	5	38.5
BH	0.01	17	1	5.9	1	5.9	5	29.4	1	5.9	6	35.3
	0.05	21	3	14.3	1	4.8	7	33.3	3	14.3	7	33.3
	0.10	24	3	12.5	1	4.2	7	29.2	3	12.5	7	29.2
possible edges		193 160										

Table 4.14: Comparison of the virtual edges for the samples $W - 195$, $Mo - 39$, $Tu - 39$, $We - 39$, $Th - 39$ and $Fr - 39$

Regional Trains - weekdays								
Corr.	α	<i>W</i> – 195	Mon-Wed-Fri		Tue-Thu		Monday	
		# Virtual	# Virtual	%	# Virtual	%	# Virtual	%
-	0.01	41	40	97.6	21	51.2	17	41.5
	0.05	67	66	98.5	41	61.2	32	47.8
	0.10	112	91	81.3	63	56.3	52	46.4
B	0.01	12	10	83.3	2	16.7	1	8.3
	0.05	12	11	91.7	2	16.7	1	8.3
	0.10	13	12	92.3	2	15.4	1	7.7
BH	0.01	17	16	84.1	4	23.5	1	5.9
	0.05	21	18	85.7	7	33.3	3	14.3
	0.10	24	19	79.2	9	37.5	3	12.5

Table 4.15: Comparison of virtual edges for the samples $W - 195$, $MWF - 117$, $TT - 78$ and $Mo - 39$

When no statistical correction is applied, the percentage of highlighted edges is not so regular as in other cases but it is remarkable that with just one fifth of the original data, the Tri-graph is able to point out half of the “reference virtual edges”. To check if this result is a general one or if it is just due to the chosen sample, we tested the Tri-graph also on the other weekdays (Tuesday to Friday). We refer to these samples as $Tu - 39$, $We - 39$, $Th - 39$ and $Fr - 39$. This was the easiest rule to select other samples with the same size of the “Monday-sample”.

The comparison among the different “day samples” has been summarized in Table 4.13. The percentages of pointed edges by the five samples are approximately fifty percent in the case of no statistical correction.

However we are not interested in the number of edges that the Tri-graph highlights. Principally we want to consider the “virtual” connections that will be pointed out (these edges are in fact the ones that will be transformed into new constraints for the timetable problem). Hence we compared also the results concerning the “virtual” activities in Table 4.14.

With the only exception of Tuesday, that also gave the smallest percentage regarding the total number of edges, the percentage of virtual connections that have been highlighted in the single weekdays is around 40%. This is a really good outcome since we are using just the 20% of the original data.

These results allow us to state, with cautious optimism, that also the output obtained in the first strategy, i.e. $W - 30$, is close to the half of what we expect to have.

4.4 Outputs of the Saturdays and Sundays samples

For the record we include the outputs of the Tri-graph for the holidays timetable (Saturdays and Sundays) where we considered:

- regional trains
- traveling every holiday (39 data)
- in the Harz area.

Regional Trains - Saturday - 39 data												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		virtual		error	
			#	%	#	%	#	%	#	%	#	%
-	0.01	327	149	45.6	58	17.7	36	11.0	25	7.6	59	18.0
	0.05	517	161	31.1	88	17.0	60	11.6	45	8.7	163	31.5
	0.10	665	164	24.7	97	14.6	86	12.9	61	9.2	257	38.6
B	0.01	111	87	78.4	19	17.1	2	1.8	3	2.7	0	0.0
	0.05	121	93	76.9	21	17.4	3	2.5	3	2.5	1	0.8
	0.10	126	97	77.0	21	16.7	3	2.4	3	2.4	2	1.6
BH	0.01	146	106	72.6	24	16.4	4	2.7	7	4.8	5	3.4
	0.05	166	112	67.5	28	16.9	6	3.6	10	6.0	10	6.0
	0.10	179	117	65.4	31	17.3	8	4.5	11	6.1	12	6.7
possible edges		152 076										

Table 4.16: Output of the Tri-graph for the sample $Sa - 39$

Regional Trains - Sunday - 78 data												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		virtual		error	
			#	%	#	%	#	%	#	%	#	%
-	0.01	366	154	42.1	77	21.0	39	10.7	27	7.4	69	18.9
	0.05	558	165	29.6	107	19.2	62	11.1	53	9.5	171	30.6
	0.10	755	170	22.5	119	15.8	90	11.9	74	9.8	302	40.0
B	0.01	116	87	75.0	19	16.4	2	1.7	3	2.6	5	4.3
	0.05	130	97	74.6	20	15.4	4	3.1	3	2.3	6	4.6
	0.10	137	102	74.5	22	16.1	4	2.9	3	2.2	9	4.4
BH	0.01	164	113	68.9	28	17.1	7	4.3	7	4.3	9	5.5
	0.05	190	122	64.2	36	18.9	10	5.3	10	5.3	12	6.3
	0.10	200	125	62.5	40	20.0	11	5.5	10	5.0	14	7.0
possible edges		176 715										

Table 4.17: Output of the Tri-graph for the sample $Su - 39$

We have two sets of events with cardinality:

- 551 for Saturday;
- 594 for Sunday.

The sample corresponding to Saturday ($Sa - 39$) is characterized by:

- 206 waiting activities;
- 206 “driving activities”.

and the one corresponding to Sunday ($Su - 39$) is characterized by:

- 221 waiting activities;
- 223 “driving activities”.

We use again the quotation marks for the driving activities since some events of the chain are missing.

	events	waiting	# driving
$Sa - 39$	551	206	206
$Su - 39$	594	221	223

Table 4.18: Characteristic of the samples $Sa - 39$ and $Su - 39$

Concerning these two samples, the Tri-graph method is capable to catch in the best cases around 76% of the waiting activities, but not more than 50% of the driving activities.

4.5 Conclusion

In this chapter we considered different samples of data built up from the files provided by the Deutsch Bahn within the frame of the DisKon Project to the AG-optimization group at NAM. We tested them with the stochastic approaches presented in Chapter 3: Contingency Table, Full Conditional Independence Graph, Covariance Graph and Tri-graph. These approaches were generally tested without a multistatistical correction, only in the case of the Tri-graph two possible corrections were considered: Bonferroni and Benjamini-Hochberg.

The outputs of the different methods (in absence of multistatistical corrections) were compared to get a first impression of their powers.

The Contingency Table test has been designed to identify independencies between variables and it seems weak in identifying (more complex) dependencies, principally because it strongly suffers from the transitivity property. The “new” version suggested in Chapter 3 was not able to strengthen the property of the classical formulation. Consequently a direct comparison with Tri-graph was not possible. Only its classical formulation could be compared with the Covariance Graph.

The Full Conditional Independence Graph could also not be applied, since the correlation matrices corresponding to the samples are strongly linearly dependent. The inaccurate measurement of the delays (exactly just up to the minute and not up to the second) and the small slack times associated to the waiting activities make it impossible to evaluate the precision matrix on which FCIG is based.

The Covariance Graph was easily applicable to the samples but the procedure suffers, as the Contingency Table test, from the transitivity property (even if in a weaker form), hence its output results into huge sets of edges. If we transformed all these edges into virtual activities for the capacity model (as explained in Section 2.11), the problem would contain a large amount of dominated constraints, i.e. unnecessary constraints.

The Tri-graph procedure is capable to do a selective identification of edges and to avoid most of the edges that can be referred to the transitivity property. Moreover if the samples contain enough data (e.g. $W = 195$) it catches almost all the waiting activities and most of the driving activities. In case the amount of available data is reduced to one fifth of the original size, it is still capable to catch 40% of the original virtual activities.

Modeling delays and virtual constraints

Liebe ist so unproblematisch wie ein Fahrzeug. Problematisch sind nur die Lenker, die Fahrgäste und die Straße.

FRANK KAFKA (1883-1924)
German writer

In this chapter, as the title suggests, we discuss how to represent the delays and how to model the “virtual” constraints introduced in Chapter 2. In particular we look for a definition of delay which is consistent with the assumption of multivariate distribution required by the Tri-graph method (described in Chapter 3).

5.1 Distribution of the delays

The knowledge of delay distribution allows a better modeling of delay propagation and consequently a better timetable (re-)scheduling. In fact a stable timetable should be able to absorb small delays, whereas it has no sense to demand it to be stable for larger delays. In other words, a timetable should not be stable for any possible delay but the “common” cases of delay should be successfully resolvable most of the time. To this end one needs information about the distribution of delays. Many studies have been carried out but the results are not uniform.

It is necessary to emphasize that a direct comparison of the results in the literature is not possible since the investigated variables are not always the same. However, we can highlight some reasons for the discrepancies among them:

- different data sets, corresponding to different areas and/or countries;
- different sizes of data;
- outliers, i.e. very large values of the delays, that can deeply influence the results having different effects depending on the size of the sample. Unfortunately there is no (exact) standard method to detect them;
- absence of a standardized measuring method: the question is *where* and *when* delays should be measured. When the engine passes a certain point or when the last wagon passes? At the beginning or at the end of a platform, or at a point in the middle?

Another aspect that we must consider is that:

a probability distribution will never fit empirical data exactly.

We present here a small panorama of studies on delay distribution that can be found in the literature.

Schwanhäusser (see [63]) developed in 1974 a stochastic approach for the estimation of the mean queue length as function of the distribution of primary delays, buffer time, mean headway, train sequence and priority. In his model he considered a negative exponential distribution for the forced delays. It has the drawback that negative arrival delays cannot be sampled.

A negative exponential distribution was also observed by **Meng** (see [50]) regarding the excess of scheduled waiting times, i.e. the difference between the real waiting time of a train at a platform and the scheduled one, and by **Wendler** and **Naehring** (see [78]) for the (arrival and departure) delays of delayed trains recorded by the Deutsche Bahn in the area of Nürnberg.

Hermann (see [35]) analyzed the development of delays of long distance passenger trains in the traffic control area of Frankfurt am Main and developed a model to forecast the secondary delays in the network. He found a good fit of the Chi-square and of the Weibull distributions with the observed primary delays and most of the consecutive delays, while Gamma, negative exponential and Erlang distributions were rejected for either primary or consecutive delays.

Negative exponential and Erlang distributions were instead pointed out by **Higgins** and **Kozan** (see [36] and [37]) for the primary delays and the expected secondary delays (which are determined from specified primary delays, implicitly assuming independence among them) in the Queensland Rail city train network, Australia.

Yuan (see [80]) considered the running times, the waiting times and the constraints at signaling systems corresponding to empirical measurement at The Hague HS Station, Holland as stochastic variables. His results pointed out:

- located shifted log-normal distribution for the arrival times at platforms and at approach signals;
- Weibull distribution for nonnegative arrival delays, departure delays and free waiting times in absence of hindrance from other trains (e.g. in absence of late release of a block/platform due to delayed trains);
- Normal or Weibull distribution for occupancy times (i.e. the variation of the train speed in case of different signal aspects).

In her diploma thesis **Güttler** (see [32]) analyzed a data set corresponding to some ICE lines in Germany. Her work found out a mixture distribution of normal and log-normal as best approximation for the distribution of the delay differences for the driving activities (i.e. variables corresponding to the difference between the departure delays and the arrival delays of the considered trains at two consecutive stations of their journeys).

The complexity of the studies arises from the (undefined) relationships existing between causes and effects: the railway traffic follows deterministic rules, whose selection and application depend on the choices of human operators. These choices

may be different in similar situations and produce different effects on the circulation. Therefore the identification of parameters depending upon human behavior (driver behaviors, passenger boarding ...) has been proved to be complex to be taken into account.

Accordingly, we state that the hypothesis of Normal distribution for the arrival and departure delays, even if it can not be proved by testing historical data (in some cases the distribution of the delays does not fit any known distribution), can not be rejected.

5.2 How to represent delays

Given the source delays d_i as parameter of the railway system, we introduced in Section 2.1 a possible way to define the secondary delays. We do repeat in this section the argumentation that brought us to formulate Equation (2.3) but basing our thinking on the forced delay variables y rather than on the re-scheduled time variables x .

Initially we do consider a single train $t \in \mathcal{T}$ traveling along its scheduled route represented by the sequence of events $(i_1, i_2 \dots i_n)$.

If t is on time, then its delay at every event is zero, i.e. $y_i = 0 \forall i \in \{1 \dots n_t\}$.

Instead if during its journey (i.e. between events $i - 2$ and $i - 1$) it has a source delay ($d_{i-1} > 0$) then its forced delay will be initialized to the value of the corresponding source delay, a nonnegative quantity (i.e. $y_{i-1} = d_{i-1}$).

The propagation of the delay can be limited by scheduled slack times. This is a common technique used in the timetable scheduling as a remedy to minimize the effects of small irregularities. Slack times are usually added to running, waiting and turn-overs times to increase the ability of the timetable to face the unpredictable disturbances arising in operation. According to the notation introduced in Section 2.1, we can define the slack time between two events i and j (such that $a = (i, j) \in \mathcal{A}$) as $c_{ij} = \pi_j - \pi_i - L_a$, where π_i is the scheduled timetable of event i and L_a is the minimal duration of activity $a = (i, j)$. If we assume that the slack time of train t between events $i - 1$ and i is $c_{i-1,i}$ and we initialize its forced delay at event $i - 1$ with the corresponding source delay $y_{i-1} = d_{i-1}$, then the secondary delay corresponding to the next event of its journey is

$$(5.1) \quad y_i = y_{i-1} - c_{i-1,i}$$

We want to avoid negative delays, i.e. earlier departures/arrivals of trains, because they are neither allowed (departure) nor desired (earlier arrivals can generate bigger problems than small delays since blocks that can be required by other trains will be "unexpectedly" occupied and the passengers service time will be smaller, see [50] and [23]). Therefore we include the function $\max\{., 0\}$ in Equation (5.1)

$$(5.2) \quad y_i = \max\{y_{i-1} - c_{i-1,i}, 0\}.$$

Other source delays may also occur, hence $y_i = \max\{d_i, y_{i-1} - c_{i-1,i}, 0\}$. In words, if train t is capable to recover more time than the current delay, then it will be on time.

Now we want to introduce the effects of other trains on the delay of train t . The delay of a train can depend on more than one previous event: e.g. the delay of a train in case of fixed connections depends not only on its delay at the previous event of its journey but also on the delays of the connected trains, if the connections are kept. In order to extend Equation (5.2) including all these dependencies, we define a set of events that can influence the timetable of a train t at a given event i

$$pa(i) = \{k \in \mathcal{E} : a = (k, i) \in \mathcal{A} \cup \mathcal{A}^{virtual}\}$$

where \mathcal{A} and $\mathcal{A}^{virtual}$ are defined as in (2.2) and (2.9f) respectively. In Graph theory this set corresponds to the parent of the node i given a graph $G = (\mathcal{E}, \mathcal{A} \cup \mathcal{A}^{virtual})$ (see Section 3.4.1).

The delay at event i is defined as in (5.2) plus possibly larger delays originating from the connected events k :

$$(5.3) \quad y_i = \max\{y_{i-1} - c_{i-1,i}, 0\} + \max_{k \in pa(i)} \{(y_k - c_{k,i}) - (y_{i-1} - c_{i-1,i}), 0\}$$

that can be rewritten as

$$(5.4) \quad y_i = \max\{\max_{k \in pa(i)} (y_k - c_{k,i}), 0\}.$$

since event $i - 1$ belongs certainly to the set $pa(i)$. Including other possible source delays we get $y_i = \max\{d_i, \max_{k \in pa(i)} (y_k - c_{k,i}), 0\}$.

Remark 5.1 In case of large delays and/or small slack times, these definitions of forced delays imply linear dependencies among the variables.

Remark 5.2 In particular it has been empirically checked that the slack time for the waiting activities is relatively small even if compared to “common” delays, hence we do expect strong limitations in the application of the Full Conditional Independence Graph, due to the linear dependencies between variables.

Definitions (5.2) and (5.4) are quite intuitive and they have already been applied in the prediction of delay, as shown in Section 5.3.

5.3 Modeling train delays in Urban Networks

Higgins and Kozan suggested in their article [37] a model to estimate expected positive delays. We report here their formulation of the model, writing in bold type between brackets the corresponding notation according to our formulation of the problem (see Section 2.1, in particular model [CTM-2]). We prefer not to change the notation since their formulation is mainly based on two index sets (one for the trains, the second for the blocks of the system) while our formulation considers only one index set (the set of events of the system).

In their model Higgins and Kozan consider the following specifications:

- delay: just positive delays are considered since early departures are not allowed for passenger trains (**secondary delay** y);
- source delay: the delay of a train that is not caused by another train (**source delay** d);
- minimum headway: the minimum length of time separating two trains traveling on the same edge, which is determined by the length of the blocks in the edge (**headway** h);
- link: a section of track on which only one train is permitted at any time (**block** m);
- siding: a track link (**a block**) that can be used for the crossing and passing of trains.

Moreover they distinguish among three kinds of delays:

- direct delays of trains, i.e. source delays;
- knock-on delays to other trains, i.e. forced delays due to driving/waiting activities or to occupation of the scheduled block from a delayed train;
- delays due to late connections, where train connections include:
 1. fixed connections;
 2. fixed departure orders in a station;
 3. commencements of a new service after arrival at the destination using the same physical train, i.e. turn-overs of trains.

They consider also the following assumptions:

- scheduled times contain some slack time;
- if a train arrives earlier, it will wait until the scheduled departure;
- trains may have different upper average velocity;
- waiting times at a station are included in the scheduled traveling times;
- trains can bypass a delayed one if there is a free siding or parallel track available;
- conflicts are solved on a “first come first serve” basis (i.e. FIFO);
- the capacity at a station is determined by the number of track links in it.

The following parameters are considered in the model:

I	= set of train services for one cycle of the schedule, usually daily, (set of trains $t \in \mathcal{T}$);
J	= set of links for the whole suburban network (set of blocks $m \in \mathcal{M}$);
Q_j^i	= set of remaining links from link $j \in J$ along which train $i \in I$ is scheduled to travel;
R_j^i	= set of links prior to link $j \in J$ along which train $i \in I$ is scheduled to travel;
Q^i	= set of links along which train $i \in I$ is scheduled to travel from origin to destination, i.e. $Q^i = (R_j^i, j, Q_j^i)$;
V_j^i	= the link immediately prior to link $j \in J$ on which train $i \in I$ is scheduled to travel, $V_j^i \in R_j^i$;
Y_j^i	= scheduled departure time of train $i \in I$ from link $j \in Q_j^i$ (scheduled time π);
T	= duration of a source delay (value of the source delay variable d);
$PRO(T, i, j)$	= probability that train $i \in I$ will have a source delay of duration T on link $j \in J$ (in [37] neg. exponential);
E_{kj}^i	= amount of time train $i \in I$ is able to recover between links $k, j \in Q^i$ (scheduled slack time c_{kj}). The shortest possible travel time for train $i \in I$ between these two links is $Y_k^i - Y_j^i - E_{kj}^i$ i.e. scheduled travel time minus slack time (minimal traveling duration L_a where $a = (k, j)$);
$CONN$	= set of train connections. A train connection takes place between train $i \in I$ at link $j \in J$ and train $l \in I$ at link $h \in J$. Links $i, j \in J$ are located at the same station. $(i, j, l, h) \in CONN$ (set of connections \mathcal{A}^{change}).

The output of the model corresponds to the following information:

$PRS(TL, i, j)$	= probability that train $i \in I$ has a current delay TL at the departure of link $j \in J$;
${}^1TC_j^i$	= source delay of train $i \in I$ at link $j \in Q^i$ (source delay d);
${}^2TC_j^i$	= delay of train $i \in I$ at link $j \in Q^i$ as a result of knock-on delays;
${}^3TC_j^i$	= delay of train $i \in I$ at link $j \in Q^i$ as a result of late connections;
TC_j^i	= delay of train $i \in I$ at link $j \in Q^i$, $TC_j^i = {}^1TC_j^i + {}^2TC_j^i + {}^3TC_j^i$ (secondary delay y).

Direct delays of trains

A train $i \in I$ with a current delay TL has a source delay T at link $k \in J$, therefore its delay at link $j \in Q_j^i$ is

$$\max\{TL - E_{kj}^i + T, 0\}$$

where E_{jk}^i is the scheduled slack time from link $k \in R_j^i$ to $j \in Q_j^i$. Taking into account all possible source delays T and current delays TL , the expected train delay of train $i \in I$ at link $j \in Q_j^i$ due to source delays is

$$(5.5) \quad E(^1TC_j^i) = \sum_{k \in R_j^i} \sum_T \sum_{TL} PRS(TL, i, V_j^i) PRO(T, i, k) g_1(i, k, j, T, TL)$$

where $g_1(i, k, j, T, TL) = \max\{TL + T - \max(TL, E_{jk}^i), 0\}$.

Knock-on delay to other trains

We can distinguish two different cases: unidirectional and bidirectional track.

Unidirectional track

A train $l \in I$ with a current delay TL suffers a source delay T on link $k \in Q_j^l$. We say that a train $i \in I$ with a current delay of TM is in ${}_{TL TM}^T S_k^l$ if it suffers a knock-on delay due to train l . The set ${}_{TL TM}^T S_k^l$ is ordered according to Y_k^i so that every train $i \in {}_{TL TM}^T S_k^l$ is characterized by its position in the set ${}_{TL TM}^T P_k^{lj}$.

Note that a train i will not belong to set ${}_{TL TM}^T S_k^l$ if it can bypass the train l .

Furthermore M_j is defined as the minimum average time for a train to travel along link $j \in J$.

Considering the scheduled slack time E_{kj}^i and just positive delays, the knock-on delay to train $i \in I$ at link $j \in D^i$ due to train $l \in I$ is

$$g_2(i, k, j, l, T, TL, TM) = \begin{cases} \max\{Y_k^l + T + TL + {}_{TL TM}^T P_k^{lj} M_k - TM - Y_k^i - E_{kj}^i, 0\} & \text{if } i \in {}_{TL TM}^T S_k^l \\ 0 & \text{otherwise} \end{cases}$$

Extending the formulation in order to include all possible values of source delays T and current delays TL and TM we get

$$(5.6) \quad E(^2TC_j^i) = \sum_k \sum_T \sum_l \sum_{TL} \sum_{TM} PRO(T, l, k) PRS(TM, i, k) PRS(TL, l, k) g_2(i, k, j, l, T, TL, TM)$$

Bidirectional track

Under the assumption that in a delayed situation trains traveling in the same direction have priority over opposite trains, the knock-on delay on a bidirectional track can be defined as in (5.6) by adding the further condition that the set ${}_{TL TM}^T S_k^l$ contains all "same-direction" trains first (in ascending order with respect to Y_k^i), followed by all "opposite direction" trains.

Delay because of late connections

A scheduled connection between train $i \in I$ and $l \in I$ departing from links $k \in Q^i$ and $h \in Q^l$ located in the same station is given, so that i cannot depart until l arrives, i.e. $Y_k^i \geq Y_h^l$. Given the scheduled slack time E_{jk}^i and supposing that i and l have current delays TM and TL respectively and that there are just positive delays, the arrival delay of train $i \in I$ at link $j \in Q_j^i$ can be defined as

$$g_3(i, k, l, h, j, TL, TM) = \max\{Y_h^l + TM - Y_k^i - TL - E_{jk}^i, 0\}$$

Taking into account all possible connections of train i at link j we get

$$(5.7) \quad E(^3TC_j^i) = \sum_{\substack{k,l,h: \\ (i,k,l,h) \in \text{CONN}}} \sum_{TL} PRS(TL, i, V_k^i) \sum_{TM} PRS(TM, l, V_h^l) g_3(i, k, l, h, j, TL, TM)$$

Expected delay

Adding together the three delay components we obtain the overall expected delay of train $i \in I$ at link $j \in Q^i$

$$(5.8) \quad E(TC_j^i) = E(^1TC_j^i) + E(^2TC_j^i) + E(^3TC_j^i)$$

This model to estimate expected delays (to trains or track links) has been tested on the suburban railway network in Queensland, Australia. The model estimations of $E(TC)$ were compared to stochastic simulations, results of which were used to reflect actual conditions.

The small model inaccuracy was reported to be due to the assessed distribution of current delay, being slightly different from the actual distribution for some trains.

Principal aim of this model is to evaluate the average delay of every train of the system. In future works, this average delay should be then added to the scheduled timetable to get a new timetable that at the current state of reseach is not guaranteed to be feasible, since no further checks are done concerning the capacity constraints of the model. This model seems more suitable for a program which purpose is to reduce the average delay of every train. Our formulation of the problem instead is based on the minimization of the total delay of the system. Moreover our formulation guarantees the feasibility of the “new timetable” for the problem.

5.4 Some considerations about the normal distribution.

In Section 5.2 we introduced a quite intuitive way to represent the delays. Before applying it to our model, it is necessary to check if it satisfies the hypothesis of (multivariate) normal distribution as required by the Tri-graph method.

We consider a train t characterized by a sequence of event (t_1, \dots, t_n) . At event i it suffers a source delay d_i , whit d_i normally distributed, i.e. $d_i \sim N(\mu_{d_i}, \sigma_{d_i}^2)$.

The hypothesis of normal distribution on a source delay is quite strong and is tightly dependent on the concept of “delay”. If we consider just technical problems or faulty signals as source delays, then these variables will always be nonnegative. In this case an exponential distribution will be a more suitable hypothesis. But if we also consider earlier arrivals in a station (earlier departures are not allowed) as source delays, then the negative values become meaningful, especially observing that the effects of an earlier arrival are usually more problematic than the ones of a positive delay (e.g. they occupy a platform longer than expected [23]). An early arrival is an “uncommon” situation and the dispatchers are not trained to face it up. Instead late arrivals or departures are “common” situations and dispatchers know by experience which decisions should be applied.

Constant slack time

Firstly we consider the slack time c as a constant .

Thus the variable $y_i = d_i - c$ is also normally distributed, i.e. $y_i \sim N(\mu_{d_i} - c, \sigma_{d_i}^2)$, since it can be proved that if $T \sim N(\mu_T, \sigma_T^2)$ then every linear function of T , i.e. $S = aT + b$ where $a \in \mathbb{R} \setminus \{0\}$ and $b \in \mathbb{R}$, is normally distributed, i.e.

$$S \sim N(a\mu_T + b, a^2\sigma_T^2) \quad (\diamond).$$

Variable slack time

We defined the slack time as $c_a = \pi_j - \pi_i - L_a$ where $a = (i, j) \in \mathcal{A}$, π_i is the scheduled timetable of event i and L_a is the minimal duration of activity $a = (i, j)$. If we now define the slack time as $c_a = \pi_j - \pi_i - \tilde{L}_a$ where \tilde{L}_a is the real duration of activity a , then the slack time is not any longer a constant. We assume it is normally distributed, i.e. $c \sim N(\nu_c, \sigma_c^2)$, and independent from the source delay variable, i.e. $\text{cov}(d_i, c) = 0$ (see Theorem 3.5). Then y_i as sum of independent normal variables ($-c$ is normally distributed due to (\diamond) considering $(a, b) = (-1, 0)$) is normally distributed, i.e. $y \sim N(\mu_{d_i} + \nu_c, \sigma_{d_i}^2 + \sigma_c^2)$.

In general it has been proved (see [69]) that

Theorem 5.3 (Cramér) *The set of proper components of a gaussian distribution with variance $\sigma^2 > 0$ consists precisely of the gaussian distributions with variance σ_1^2 where $0 < \sigma_1^2 < \sigma^2$.*

Where the components of a variable are defined as follows:

Definition 5.4 *Given a random variable X , a random variable X_1 is called a component of X if there exists a random variable X_2 such that X_1 and X_2 are independent and $X = X_1 + X_2$.*

In particular, for all $a \in \mathbb{R}$ the deterministic random variable $X_a = a$ is a component of every random variable, because $X = X_a + (X - X_a)$ and X_a is independent from all other random variables. X_a is called *improper component*. All other components are said *proper*.

Therefore if we want to decompose the forced delay as sum of delays assuming a normal distribution, we need to require that all its proper components (i.e. the source delays and the previous forced delays) are normally distributed.

Introducing the maximum

Before introducing the case corresponding to the Definition (5.2), we discuss briefly the general case of the maximum of two normally distributed variables, i.e.

$$U = \max\{S, T\} \text{ where } S \sim N(\mu_S, \sigma_S^2) \text{ and } T \sim N(\mu_T, \sigma_T^2).$$

If $F_{ST}(S, T)$ is the joint distribution function of S and T , then

$$F_U(u) = F_{ST}(u, u)$$

since the corresponding region in the plane is (see Figure 5.1)

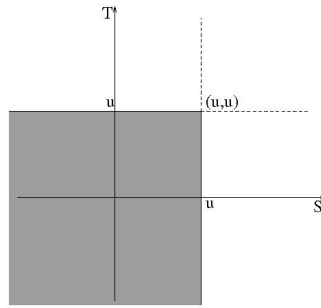
$$\{U \leq u\} = \{\max(S, T) \leq u\} = \{S \leq u\} \cap \{T \leq u\}$$

.

Then

$$F_U(u) = \mathbb{P}(\{U \leq u\}) = \mathbb{P}(\{S \leq u\} \cap \{T \leq u\}) = F_{ST}(u, u).$$

In general if S_1, \dots, S_n are normally distributed variables and $U = \max\{S_1, \dots, S_n\}$, we have $F_U(u) = F_{S_1, \dots, S_n}(u, \dots, u)$.

Figure 5.1: The variable $\text{Max}(S, T)$

If S and T are independent, then $F_U(u) = F_S(u)F_T(u)$. Differentiating with respect to u we get the density $f_U(u) = F_S(u)f_T(u) + f_S(u)F_T(u)$.

We can conclude that U is normally distributed if the joint distribution of S and T is a multivariate normal distribution (see [3]).

Obviously in Definitions (5.2) and (5.4) the constant variable zero is not normally distributed, thus y_i has not a normal distribution. However we can roughly approximate it through a normal distribution by considering it as a normally distributed variable with missing data.

5.5 About the covariance

In general zero correlation between two random variables does not imply independence, but in the case of normal distribution (see. [46], [42], [43]). This result can not be extended to other distributions, since in the general case the necessary and sufficient condition for independence is that every standardized function of the first variable should have zero correlation with any standardized function of the second variable (see [44]). Without the hypothesis of multivariate normal distribution, the resulting Covariance Graph will have very little interest for us. The entailment “independence implies null correlation” will still be valid, but we could not any more conclude that a null covariance implies independence. Thus we could not state anything about the real meaning of the pointed edges.

It still remains to check if it is possible to recognize pairs of variables that have a correlation different than zero independently from the distribution of the data, assuming the definition of delay as given in (5.2). We can positively answer this question since we can relate the covariance of two delay variables with the probability that one of them will be bigger than a certain constant value (i.e. in our case the slack time corresponding to the activity between the selected pair of variables).

According to Definition (5.2), given a random variable X representing the delay of one event, the forced delay on a connected event is defined as $Y = \max\{0, X - c\}$ where c is the slack time.

$$Y = \max\{X - c, 0\} = \begin{cases} X - c & \text{if } X > c \\ 0 & \text{if } X \leq c \end{cases}$$

Lemma 5.5 *Given two variables X and Y , such that X is a nonnegative variable and $Y = \max\{X - c, 0\}$, then $\text{cov}(X, Y) \geq \text{Var}(X) \mathbb{P}(X > c)$*

Proof: We consider two cases: $\mathbb{P}(X > c) = 0$ and $\mathbb{P}(X > c) > 0$.

First case: $\mathbb{P}(X > c) = 0$, i.e. the propability that X is bigger than c is null. Then Y is everywhere zero, hence the variables are uncorrelated, i.e. $\rho_{XY} = 0$.

Second case: $\mathbb{P}(X > c) > 0$.

We define a binary random variable Z

$$Z = \begin{cases} 1 & \text{if } X > c \\ 0 & \text{if } X \leq c \end{cases}$$

so that

$$\text{cov}(X, Y|Z) = \begin{cases} 1 & \text{if } Z = 1 \\ 0 & \text{if } Z = 0 \end{cases}$$

Through this quantity we can estimate the “direct” covariance of (X, Y)

$$\begin{aligned} \mathbb{E}[\text{cov}(X, Y|Z)] &= \mathbb{E}[\mathbb{E}[XY|Z] - \mathbb{E}[X|Z]\mathbb{E}[Y|Z]] \\ &= \mathbb{E}[XY|Z] - \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] \end{aligned}$$

We want now to show that $\mathbb{E}[XY|Z] = \mathbb{E}[XY]$. The first term is equal to

$$\begin{aligned} \mathbb{E}[XY|Z] &= \mathbb{E}[XY|Z = 1]\mathbb{P}(Z = 1) + \mathbb{E}[XY|Z = 0]\mathbb{P}(Z = 0) \\ &= \mathbb{E}[XY|Z = 1]\mathbb{P}(Z = 1) \end{aligned}$$

since when $Z = 0$ we have $Y = 0$ and hence $XY = 0$, while the second term is equal to

$$\begin{aligned} \mathbb{E}[XY] &= \mathbb{E}[XY|X > c]\mathbb{P}(X > c) + \mathbb{E}[XY|X \leq c]\mathbb{P}(X \leq c) \\ &= \mathbb{E}[XY|X > c]\mathbb{P}(X > c) \end{aligned}$$

By definition $Z = 1 \Leftrightarrow X > c$, we can state that

$$\mathbb{E}[XY|Z] = \mathbb{E}[XY]$$

and furthermore that

$$\mathbb{E}[\text{cov}(X, Y|Z)] = \mathbb{E}[XY] - \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]]$$

Applying the formula $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ we get

$$\begin{aligned} \mathbb{E}[\text{cov}(X, Y|Z)] &= \text{cov}(X, Y) + \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] \\ &= \text{cov}(X, Y) - \left[\mathbb{E}[\mathbb{E}[X|Z]\mathbb{E}[Y|Z]] - \mathbb{E}[X]\mathbb{E}[Y] \right] \end{aligned}$$

It is equivalent to

$$\text{cov}(X, Y) = \mathbb{E}[\text{cov}(X, Y|Z)] - \text{cov}(\mathbb{E}[X|Z]\mathbb{E}[Y|Z])$$

since in general $\mathbb{E}[\mathbb{E}[X|Z]] = \mathbb{E}[X]$.

It remains to show that $\text{cov}(\mathbb{E}[X|Z]\mathbb{E}[Y|Z]) \geq 0$.

$$\begin{aligned} \mathbb{E}[X|Z] &= \mathbb{E}[X|Z = 1]\mathbb{P}(Z = 1) + \mathbb{E}[X|Z = 0]\mathbb{P}(Z = 0) \\ \mathbb{E}[Y|Z] &= \mathbb{E}[Y|Z = 1]\mathbb{P}(Z = 1) + \mathbb{E}[Y|Z = 0]\mathbb{P}(Z = 0) \\ &= \mathbb{E}[X - c|Z = 1]\mathbb{P}(Z = 1) \\ &= (\mathbb{E}[X|Z = 1] - c)\mathbb{P}(Z = 1) \\ &= \mathbb{E}[X|Z = 1]\mathbb{P}(Z = 1) - c\mathbb{P}(Z = 1) \\ &= \mathbb{E}[X|Z] - \mathbb{E}[X|Z = 0]\mathbb{P}(Z = 0) - c\mathbb{P}(Z = 1) \end{aligned}$$

As $\mathbb{E}[X|Z = 1] > \mathbb{E}[X|Z = 0]$ and $\mathbb{E}[Y|Z = 1] > \mathbb{E}[Y|Z = 0]$, we have

$$\text{cov}(\mathbb{E}[X|Z]\mathbb{E}[Y|Z]) \geq 0$$

and

$$\text{cov}(X, Y) \geq \mathbb{E}[\text{cov}(X, Y|Z)]$$

In conclusion

$$\text{cov}(X, Y) \geq \text{Var}(X) \mathbb{P}(X > c)$$

since $\mathbb{E}[\text{cov}(X, Y|Z)] = 0 \mathbb{P}(Z = 0) + \text{var}(X) \mathbb{P}(Z = 1) = \text{Var}(X) \mathbb{P}(Z = 1) = \text{var}(X) \mathbb{P}(X > c)$. \square

In words, if the probability that the delay X is bigger than the slack time c , is strictly positive and its variance is strictly positive (i.e. X is not a constant), we can recognize values of the covariance between X and Y that are different than zero.

5.6 Writing the “virtual” constraints

In Section 2.11 we introduced our methodology to include the “virtual” activities arising from the Tri-graph (Equation (2.9f)) into our model. Our idea is to approximate the random variable corresponding to the second component of the edge $(i, j) \in \mathcal{A}^{\text{virtual}}$ pointed out by the Tri-graph, Y_j , by a linear regression based on the random variable corresponding to the first component of the edge, Y_i , i.e. $Y_j = \alpha_{ij}Y_i + \beta_{ij}$. This can be done by applying the linear least squares fitting technique, which is the simplest and most commonly applied form of linear regression and provides a solution to the problem of finding a fitting straight line through a set of points. In this section we define the coefficients α_{ij} and β_{ij} .

Remark 5.6 To simplify the notation we will further on consider the linear regression between the variables X and Y , $\tilde{Y} = aX + b$, so that our thinking will not weighted down by too many indices.

In practice, to calculate the coefficients a and b the vertical offsets from the (straight) line are minimized rather than the perpendicular offsets (see Figure 5.2). This provides a fitting function for the variable Y that estimates y for a given x , allows uncertainties of the data points along the x - and y -axes to be easily incorporated, and also provides a much simpler analytic form for the fitting parameters than the one obtained by using a fit based on perpendicular offsets. In addition, the fitting technique can be easily generalized from a best-fit line to a best-fit polynomial when sums of vertical distances are used. In any case, for a reasonable number of noisy data points, the difference between vertical and perpendicular fits is quite small (as done in [29]).

Vertical least squares fitting proceeds by finding the sum of the squares of the vertical deviations in order to minimize it according to parameters a and b (see [29]).

We will follow the same procedure but working with the mean (see [79]), i.e.

$$\sum_{i=1}^n \rho_{LSE} = \mathbb{E}[Y - (aX + b)]^2 = \sum_{i=1}^n \frac{[y_i - ax_i - b]^2}{n}$$

Note that this procedure does not minimize the perpendicular offsets. In addition, although the absolute sum of distances might be a more appropriate quantity to minimize, use of the absolute value results in non-differentiability which requires a great effort to be treated analytically. The square deviations from each point are therefore

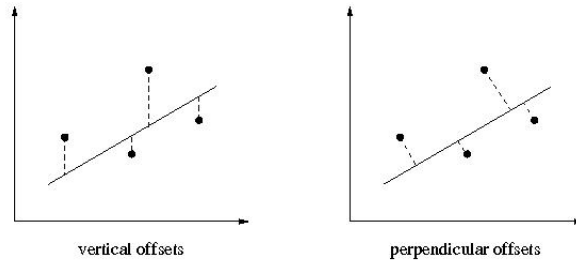


Figure 5.2: Offsets of the linear regression

summed, and the resulting residual is then minimized to find the best fit line.

We report here the calculation presented in [29].

The necessary conditions to minimize $\mathbb{E}[Y - (aX + b)]^2$ are

$$\begin{cases} \frac{\partial}{\partial a} \mathbb{E}[Y - (aX + b)]^2 = 0 \\ \frac{\partial}{\partial b} \mathbb{E}[Y - (aX + b)]^2 = 0 \end{cases}$$

Since

$$\begin{aligned} \mathbb{E}[Y - (aX + b)]^2 &= \mathbb{E}[Y^2 - 2aXY - 2bY + a^2X^2 + 2abX + b^2] \\ &= \mathbb{E}[Y^2] - 2a\mathbb{E}[XY] - 2b\mathbb{E}[Y] + a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 \end{aligned}$$

we have

$$\begin{cases} \frac{\partial}{\partial a} \mathbb{E}[Y - (aX + b)]^2 = -2\mathbb{E}[XY] + 2a\mathbb{E}[X^2] + 2b\mathbb{E}[X] = 0 \\ \frac{\partial}{\partial b} \mathbb{E}[Y - (aX + b)]^2 = -2\mathbb{E}[Y] + 2a\mathbb{E}[X] + 2b = 0 \end{cases}$$

This leads to the system

$$\begin{cases} -\mathbb{E}[XY] + a\mathbb{E}[X^2] + (\mathbb{E}[Y] - a\mathbb{E}[X])\mathbb{E}[X] = 0 \\ b = \mathbb{E}[Y] - a\mathbb{E}[X] \end{cases}$$

and to the solution

$$\begin{cases} a = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\mathbb{E}[X^2] - \mathbb{E}^2[X]} \\ b = \mathbb{E}[Y] - a\mathbb{E}[X] \end{cases}$$

Recalling that:

- $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$;
- $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X]$

we can rewrite the coefficients as

$$\begin{cases} a = \frac{\text{cov}(X, Y)}{\text{Var}(X)} \\ b = \mathbb{E}[Y] - \frac{\text{cov}(X, Y)}{\text{Var}(X)}\mathbb{E}[X] \end{cases}$$

Hence the linear regression can be written as

$$\begin{aligned} \tilde{Y}_{LSE} &= \frac{\text{cov}(X, Y)}{\text{Var}(X)}X + \left(\mathbb{E}[Y] - \frac{\text{cov}(X, Y)}{\text{Var}(X)}\mathbb{E}[X] \right) \\ (5.9) \quad &= \mathbb{E}[Y] + \frac{\text{cov}(X, Y)}{\text{Var}(X)}(X - \mathbb{E}[X]) \end{aligned}$$

where the slope a represents the average charge in Y associated with a one-unit increase in X . The intercept b is the fitted value of Y when $X = 0$ (see [29] and [79]). The resulting value of the mean of the vertical deviations is

$$\begin{aligned}
\mathbb{E}[Y - (aX + b)]^2 &= \mathbb{E}\left[Y - \frac{\text{cov}(X,Y)}{\text{Var}(X)}X - \left(\mathbb{E}[Y] - \frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]\right)\right]^2 \\
&= \mathbb{E}\left[Y^2 - 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}XY - 2\mathbb{E}[Y]Y + 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]Y\right. \\
&\quad \left. + \frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}X^2 + 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[Y]X - 2\frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}\mathbb{E}[X]X\right. \\
&\quad \left. + \mathbb{E}^2[Y] - 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]\mathbb{E}[Y] + \frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}\mathbb{E}^2[X]\right] \\
&= \mathbb{E}[Y^2] - 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[XY] - 2\mathbb{E}^2[Y] + 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad + \frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}\mathbb{E}[X^2] + 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]\mathbb{E}[Y] - 2\frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}\mathbb{E}^2[X] \\
&\quad + \mathbb{E}^2[Y] - 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\mathbb{E}[X]\mathbb{E}[Y] + \frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}\mathbb{E}^2[X] \\
&= (\mathbb{E}[Y^2] - \mathbb{E}^2[Y]) - 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&\quad + \frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}(\mathbb{E}[X^2] - \mathbb{E}^2[X]) \\
&= \text{Var}(Y) - 2\frac{\text{cov}(X,Y)}{\text{Var}(X)}\text{cov}(X,Y) + \frac{\text{cov}^2(X,Y)}{\text{Var}^2(X)}\text{Var}(X) \\
&= \text{Var}(Y) - \frac{\text{cov}^2(X,Y)}{\text{Var}(X)}
\end{aligned}$$

If we compare (5.1), i.e. $Y = X - c$ according to the notation used in this section, with Equation (5.9) we recognize a good agreement.

Our previous line of reasoning was based on the assumption of a positive correlation between the two (delay) variables, i.e. $\text{cov}(X, Y) = \text{Var}(X)$ and if we now calculate $\text{cov}(X, \tilde{Y})$ we get exactly:

$$\text{cov}(X, \tilde{Y}) = \text{cov}(X, aX + b) = a \text{cov}(X, X) = \text{cov}(X, Y)$$

But $\text{cov}(Y, \tilde{Y}) = a \text{cov}(Y, X) = \text{Var}(X) \geq 0$ that can be different than 1 according to the value of $\text{Var}(X)$.

According to Definition (5.2), we want a slope as close as possible to one, but we also want a robust linear regression (and LSE is not). Hence we looked for alternative possibilities to represent the linear regression of the delays, and we consider then the increment of the “error” to judge if it would be reasonably acceptable.

5.6.1 Four alternatives

We considered four different alternative ways to calculate the coefficients a and b of the linear regression $\tilde{Y} = aX + b$ and we compared them with the least squares (LSE). We prefer not to use directly LSE regression since we do expect an arbitrary number of large delays in the data and that can exceed the tolerance of LSE, i.e. the number of “bad data” that LSE can tolerate without itself being affected by an arbitrary large extent.

The four alternatives correspond to four different choices of the slope coefficient a in the linear regression. Given a , the intercept is determined as $b = E(Y) - aE(X)$.

This choice of b assures that Y and its approximation $aX + b$ have the same mean. Moreover it minimizes the corresponding least squares error in case of fixed slope. In fact if we fix the slope a and minimize the corresponding least square error, i.e. $\mathbb{E}[Y - aX - b]^2$ with a fixed and constant, we can determine the intercept b as follows. We consider the first and second derivatives of the least squares error in b , then get the system of equations

$$\begin{cases} \frac{\partial}{\partial b} \mathbb{E}[Y - (aX - b)]^2 = -2\mathbb{E}[Y] + 2a\mathbb{E}[X] + 2b = 0 \\ \frac{\partial^2}{\partial b^2} \mathbb{E}[Y - (aX - b)]^2 = 2 \end{cases}$$

Thus the value

$$(5.10) \quad b = \mathbb{E}[Y] - a\mathbb{E}[X]$$

corresponds to a minimum of the error function.

Our four alternatives for the slope are:

- **(A)** we require that Y and its approximation $aX + b$ will have not only the same mean, but also the same variance;
- **(B)** we define $a = \rho$, the correlation coefficient of Y and X ;
- **(C)** we chose $a = \sigma$ the covariance of Y and X ;
- **(C)** we set $a = 1$.

Alternative A

We calculate a_A and b_A in such a way that the mean and the variance of Y and \tilde{Y} will be the same, i.e.

$$\begin{cases} \mathbb{E}[Y] = \mathbb{E}(\tilde{Y}) \\ \text{Var}(Y) = \text{Var}(\tilde{Y}) \end{cases}$$

This leads to the system

$$\begin{cases} \mathbb{E}[Y] = a_A\mathbb{E}[X] + b_A \\ \text{Var}(Y) = a_A^2\text{Var}(X) \end{cases}$$

The solution is

$$\begin{cases} a_A = \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}} \\ b_A = \mathbb{E}[Y] - \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X] \end{cases}$$

Thus the first alternative for the linear regression approximation is

$$\begin{aligned} \tilde{Y}_A &= \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}X + \left(\mathbb{E}[Y] - \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]\right) \\ (5.11) \quad &= \mathbb{E}[Y] + \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}(X - \mathbb{E}[X]) \end{aligned}$$

Alternative B

The value of a is fixed to the correlation of the variables, $a_B = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$, i.e. $a_B = \rho_{XY}$. According to (5.10) we get the coefficients

$$\begin{cases} a_B = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \\ b_B = \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X] \end{cases}$$

corresponding to the linear regression

$$\begin{aligned} \tilde{Y}_B &= \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}X + \left(\mathbb{E}[Y] - \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X] \right) \\ (5.12) \quad &= \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}(X - \mathbb{E}[X]) \end{aligned}$$

We notice that the variance of \tilde{Y}_B is given by $\text{Var}(\tilde{Y}_B) = \frac{\text{cov}^2(X,Y)}{\text{Var}(X)\text{Var}(Y)}\text{Var}(X)$, therefore it coincides with $\text{Var}(Y)$ if and only if $\text{Var}^2(Y) = \text{cov}^2(X,Y)$.

Alternative C

The value of a is fixed to the covariance of the variables, i.e. $a_C = \text{cov}(X,Y)$. According to (5.10) we get the coefficients

$$\begin{cases} a_C = \text{cov}(X,Y) \\ b_C = \mathbb{E}[Y] - \text{cov}(X,Y)\mathbb{E}[X] \end{cases}$$

corresponding to the linear regression

$$\begin{aligned} \tilde{Y}_C &= \text{cov}(X,Y)X + \left(\mathbb{E}[Y] - \text{cov}(X,Y)\mathbb{E}[X] \right) \\ (5.13) \quad &= \mathbb{E}[Y] + \text{cov}(X,Y)(X - \mathbb{E}[X]) \end{aligned}$$

the variance of \tilde{Y}_C is given by $\text{Var}(\tilde{Y}_C) = \text{cov}^2(X,Y)\text{Var}(X)$ that coincides with $\text{Var}(Y)$ if and only if $\text{cov}^2(X,Y) = 1$.

Alternative D

The value of a is fixed to 1. According to (5.10) we get the coefficients

$$\begin{cases} a_D = 1 \\ b_D = \mathbb{E}[Y] - \mathbb{E}[X] \end{cases}$$

corresponding to the linear regression

$$\begin{aligned} \tilde{Y}_D &= X + \left(\mathbb{E}[Y] - \mathbb{E}[X] \right) \\ (5.14) \quad &= \mathbb{E}[Y] + (X - \mathbb{E}[X]) \end{aligned}$$

The variance of \tilde{Y}_D is always equal to the variance of Y .

Remark 5.7 It is easy to see that (5.14) is a special case of all the previous ones: of (5.11) and (5.13) if $\text{cov}(X,Y) = 1$ (since that implies $\text{Var}(Y) = \text{Var}(X)$), and of (5.12) if $\rho_{XY} = 1$.

5.6.2 Least squares error

Firstly we calculate the values of the mean of the sum of the squares of the vertical deviations to see the error we commit with every alternative.

$$\begin{aligned}
\mathbb{E}[Y - \tilde{Y}_A]^2 &= \mathbb{E}\left[Y - \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}X - \left(\mathbb{E}[Y] - \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]\right)\right]^2 \\
&= \mathbb{E}\left[Y^2 - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}XY - 2\mathbb{E}[Y]Y + 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]Y + \frac{\text{Var}(Y)}{\text{Var}(X)}X^2\right. \\
&\quad + 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[Y]X - 2\frac{\text{Var}(Y)}{\text{Var}(X)}\mathbb{E}[X]X + \mathbb{E}^2[Y] \\
&\quad \left. - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]\mathbb{E}[Y] + \frac{\text{Var}(Y)}{\text{Var}(X)}\mathbb{E}^2[X]\right] \\
&= \mathbb{E}[Y^2] - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[XY] - 2\mathbb{E}^2[Y] + 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad + \frac{\text{Var}(Y)}{\text{Var}(X)}\mathbb{E}[X^2] + 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]\mathbb{E}[Y] - 2\frac{\text{Var}(Y)}{\text{Var}(X)}\mathbb{E}^2[X] + \mathbb{E}^2(Y) \\
&\quad - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\mathbb{E}[X]\mathbb{E}[Y] + \frac{\text{Var}(Y)}{\text{Var}(X)}\mathbb{E}^2[X] \\
&= (\mathbb{E}[Y^2] - \mathbb{E}^2[Y]) + \frac{\text{Var}(Y)}{\text{Var}(X)}(\mathbb{E}[X^2] - \mathbb{E}^2[X]) \\
&= -2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= \text{Var}(Y) + \frac{\text{Var}(Y)}{\text{Var}(X)}\text{Var}(X) - \sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\text{cov}(X, Y) \\
&= 2\text{Var}(Y) - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\text{cov}(X, Y). \\
\mathbb{E}[Y - \tilde{Y}_B]^2 &= \mathbb{E}\left[Y - \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}X - \left(\mathbb{E}[Y] - \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X]\right)\right]^2 \\
&= \mathbb{E}\left[Y^2 - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}XY - 2\mathbb{E}[Y]Y + 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X]Y\right. \\
&\quad + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}X^2 - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[Y]X - 2\frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}\mathbb{E}[X]X \\
&\quad \left. + \mathbb{E}^2[Y] - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X]\mathbb{E}[Y] + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}\mathbb{E}^2[X]\right] \\
&= \mathbb{E}[Y^2] - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[XY] + 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad - 2\mathbb{E}^2[Y] + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}\mathbb{E}^2[X] - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad + \mathbb{E}^2[Y] - 2\frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}\mathbb{E}^2[X] - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}\mathbb{E}^2[X] \\
&= (\mathbb{E}[Y^2] - \mathbb{E}^2[Y]) + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}(\mathbb{E}[X^2] - \mathbb{E}^2[X]) \\
&\quad - 2\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&= \text{Var}(Y) + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}\text{Var}(X) - 2\frac{\text{cov}^2(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \\
&= \text{Var}(Y) + \frac{\text{cov}^2(X, Y)}{\text{Var}(Y)} - 2\frac{\text{cov}^2(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y - \tilde{Y}_C]^2 &= \mathbb{E}\left[Y - \text{cov}(X, Y) X - \left(\mathbb{E}[Y] - \text{cov}(X, Y)\mathbb{E}[X]\right)\right]^2 \\
&= \mathbb{E}\left[Y^2 - 2\text{cov}(X, Y) XY - 2\mathbb{E}[Y] Y + 2\text{cov}(X, Y)\mathbb{E}[X] Y \right. \\
&\quad \left. + \text{cov}^2(X, Y) X^2 - 2\text{cov}(X, Y)\mathbb{E}[Y] X + 2\text{cov}^2(X, Y)\mathbb{E}[X] X \right. \\
&\quad \left. + \mathbb{E}^2[Y] - 2\text{cov}(X, Y)\mathbb{E}[X]\mathbb{E}[Y] + \text{cov}^2(X, Y)\mathbb{E}^2[X]\right] \\
&= \mathbb{E}[Y^2] - 2\text{cov}(X, Y)\mathbb{E}[XY] - 2\mathbb{E}^2[Y] + 2\text{cov}(X, Y)\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad + \text{cov}^2(X, Y)\mathbb{E}[X^2] - 2\text{cov}(X, Y)\mathbb{E}[X]\mathbb{E}[Y] + 2\text{cov}^2(X, Y)\mathbb{E}^2[X] \\
&\quad + \mathbb{E}^2[Y] - 2\text{cov}(X, Y)\mathbb{E}[X]\mathbb{E}[Y] + \text{cov}^2(X, Y)\mathbb{E}^2[X] \\
&= (\mathbb{E}[Y^2] - \mathbb{E}^2[Y]) - 2\text{cov}(X, Y)(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \\
&\quad + \text{cov}^2(X, Y)(\mathbb{E}[X^2] - \mathbb{E}^2[X]) \\
&= \text{Var}(Y) - 2\text{cov}^2(X, Y) + \text{cov}^2(X, Y)\text{Var}(X) \\
&= \text{Var}(Y) + \text{cov}^2(X, Y)(\text{Var}(X) - 2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y - \tilde{Y}_D]^2 &= \mathbb{E}\left[Y - X - (\mathbb{E}[Y] - \mathbb{E}[X])\right]^2 \\
&= \mathbb{E}\left[Y^2 - 2XY - 2\mathbb{E}[Y] Y + 2\mathbb{E}[X] Y + X^2 + 2\mathbb{E}[Y] X \right. \\
&\quad \left. - 2\mathbb{E}[X] X + \mathbb{E}^2[Y] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}^2[X]\right] \\
&= \mathbb{E}[Y^2] - 2\mathbb{E}[XY] - 2\mathbb{E}^2[Y] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] \\
&\quad - 2\mathbb{E}^2[X] + \mathbb{E}^2[Y] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}^2[X] \\
&= (\mathbb{E}[Y^2] - \mathbb{E}^2[Y]) - 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) + (\mathbb{E}[X^2] - \mathbb{E}^2[X]) \\
&= \text{Var}(Y) - 2\text{cov}(X, Y) + \text{Var}(X) \\
&= \text{Var}(Y - X)
\end{aligned}$$

since $\text{Var}(aX + bY) = a^2 \text{Var}(X) + 2ab \text{cov}(X, Y) + b^2 \text{Var}(Y)$.

Furthermore we check how much each alternative differs from the “optimal” approximation given by (5.9).

$$\begin{aligned}
\mathbb{E}[Y - \tilde{Y}_A]^2 - \mathbb{E}[Y - \tilde{Y}_{LSE}]^2 &= \\
&= \left(2\text{Var}(Y) - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\text{cov}(X, Y)\right) - \left(\text{Var}(Y) - \frac{\text{cov}^2(X, Y)}{\text{Var}(X)}\right) \\
&= \text{Var}(Y) - 2\sqrt{\frac{\text{Var}(Y)}{\text{Var}(X)}}\text{cov}(X, Y) + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)} \\
&= \left(\sqrt{\text{Var}(Y)} - \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)}}\right)^2 \geq 0
\end{aligned}$$

that is equal to zero if the variables X and Y are positively correlated, i.e. $\rho_{XY} = 1$.

$$\begin{aligned} \mathbb{E}[Y - \tilde{Y}_B]^2 - \mathbb{E}[Y - \tilde{Y}_{LSE}]^2 &= \\ &= \left(\text{Var}(Y) + \frac{\text{cov}^2(X, Y)}{\text{Var}(Y)} - 2 \frac{\text{cov}^2(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \right) - \left(\text{Var}(Y) - \frac{\text{cov}^2(X, Y)}{\text{Var}(X)} \right) \\ &= \frac{\text{cov}^2(X, Y)}{\text{Var}(Y)} - 2 \frac{\text{cov}^2(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} + \frac{\text{cov}^2(X, Y)}{\text{Var}(X)} \\ &= \left(\frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(Y)}} - \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)}} \right)^2 \geq 0 \end{aligned}$$

that is equal to zero either when the variables are uncorrelated, i.e. $\text{cov}(X, Y) = 0$ (that is not our case since the Tri-graph points out exactly the dependencies among variables), or when the two variables have the same variance, i.e. $\text{Var}(X) = \text{Var}(Y)$.

$$\begin{aligned} \mathbb{E}[Y - \tilde{Y}_C]^2 - \mathbb{E}[Y - \tilde{Y}_{LSE}]^2 &= \\ &= \left(\text{Var}(Y) - 2\text{cov}^2(X, Y) + \text{cov}^2(X, Y)\text{Var}(X) \right) - \left(\text{Var}(Y) - \frac{\text{cov}^2(X, Y)}{\text{Var}(X)} \right) \\ &= \text{cov}^2(X, Y) \left(\text{Var}(X) + 2 - \frac{1}{\text{Var}(X)} \right) \\ &= \text{cov}^2(X, Y) \left(\sqrt{\text{Var}(X)} + \frac{1}{\sqrt{\text{Var}(X)}} \right) \geq 0 \end{aligned}$$

that is equal to zero only if the variables are uncorrelated, i.e. $\text{cov}(X, Y) = 0$, since the variance of a variable is always positive, but this is exactly the case we do not want by definition of alternative C. Therefore we do expect a (large) positive error if we decide to apply this alternative.

$$\begin{aligned} \mathbb{E}[Y - \tilde{Y}_D]^2 - \mathbb{E}[Y - \tilde{Y}_{LSE}]^2 &= \\ &= \left(\text{Var}(Y) - 2\text{cov}(X, Y) + \text{Var}(X) \right) - \left(\text{Var}(Y) - \frac{\text{cov}^2(X, Y)}{\text{Var}(X)} \right) \\ &= \frac{1}{\text{Var}(X)} (\text{Var}^2(X) - 2\text{cov}(X, Y)\text{Var}(X) + \text{cov}^2(X, Y)) \\ &= \frac{1}{\text{Var}(X)} (\text{Var}(X) - \text{cov}(X, Y))^2 \geq 0 \end{aligned}$$

that is equal to zero if $\text{Var}(X) = \text{cov}(X, Y)$.

5.6.3 Empirical comparison

To understand which alternative for the linear regression (A to D) is more suitable to be applied in our model, we applied them to the total number of edges (and to the set of “virtual connections”) pointed out from the Tri-graph for the samples $W - 30$, $W - 195$, $Sa - 39$, $Su - 39$, and the subsamples $MoWeFr - 117$, $TuTh - 78$ and $Mo - 39$ (see Sections 4.2, 4.3 and 4.4).

The following tables report either the number of edges (or “virtual” connections) that have minimal LSE in one of the four alternatives (A to D) with the corresponding percentage on the total amount of edges considered, or the average value of the LSE corresponding to the set of edges in the four alternatives (A to D) (in this table we also include the mean of the least squares approximation given by Equation (5.9), denominated as alternative LSE).

Quantile α	Nr. Edges TG	Regional Trains - weekdays - 30 data							
		A	%	B	%	C	%	D	%
0.01	463	269	58.1	101	21.8	13	2.8	80	17.3
0.05	753	413	54.8	170	22.6	38	5.0	132	17.5
0.10	1085	544	50.1	259	23.9	76	7.0	206	19.0
Quantile α	Nr. Edges TG	Regional Trains - weekdays - 195 data							
		A	%	B	%	C	%	D	%
0.01	491	155	31.6	209	42.6	24	4.9	103	21.0
0.05	745	204	27.4	348	46.7	49	6.6	144	19.3
0.10	1002	241	24.1	494	49.3	85	8.5	182	18.2
Quantile α	Nr. Edges TG	Regional Trains - Saturday - 39 data							
		A	%	B	%	C	%	D	%
0.01	327	180	55.0	73	22.3	11	3.4	63	19.3
0.05	517	271	52.4	117	22.6	27	5.2	102	19.7
0.10	665	323	48.6	162	24.4	46	6.9	134	20.2
Quantile α	Nr. Edges TG	Regional Trains - Sunday - 39 data							
		A	%	B	%	C	%	D	%
0.01	366	201	54.9	80	21.9	17	4.6	68	18.6
0.05	558	279	50.0	131	23.5	44	7.9	104	18.6
0.10	755	358	47.4	170	22.5	80	10.6	147	19.5

Table 5.1: Comparison of LSE on edges TG (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$

Quantile α	Regional Trains - weekdays - 30 data				
	LSE	A	B	C	D
0.01	0.526	1.806	1.625	1.10E+04	1.690
0.05	0.646	1.839	1.638	8.66E+03	1.821
0.10	0.800	1.752	1.732	9.05E+03	2.028
Quantile α	Regional Trains - weekdays - 195 data				
	LSE	A	B	C	D
0.01	17.328	46.807	21.872	8.40E+05	38.537
0.05	12.713	32.920	15.900	5.54E+05	28.065
0.10	18.580	47.083	21.469	4.12E+05	31.378
Quantile α	Regional Trains - Saturday - 39 data				
	LSE	A	B	C	D
0.01	0.654	0.704	1.048	4.44E+11	1.121
0.05	0.957	1.075	1.537	5.61E+11	1.678
0.10	1.177	1.367	1.804	5.45E+11	2.000
Quantile α	Regional Trains - Sunday - 39 data				
	LSE	A	B	C	D
0.01	0.94	0.99	2.05	1.32E+04	2.41
0.05	2.38	2.64	4.43	1.29E+05	6.74
0.10	2.27	2.60	3.94	9.22E+04	5.85

Table 5.2: Comparison of LSE mean on edges TG (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$

Quantile α	Nr. Virtual TG	Regional Trains - weekdays - 30 data							
		A	%	B	%	C	%	D	%
0.01	58	30	60.0	17	34.0	0	0.0	3	6.0
0.05	84	50	59.5	24	28.6	2	2.4	8	9.5
0.10	135	72	53.3	34	25.2	8	5.9	21	15.6
Quantile α	Nr. Virtual TG	Regional Trains - weekdays - 195 data							
		A	%	B	%	C	%	D	%
0.01	41	18	43.9	19	46.3	2	4.9	2	4.9
0.05	67	25	37.3	30	44.8	2	3.0	10	14.9
0.10	112	35	31.3	57	50.9	4	3.6	16	14.3
Quantile α	Nr. Virtual TG	Regional Trains - Saturday - 39 data							
		A	%	B	%	C	%	D	%
0.01	25	18	72.0	5	20.0	1	4.0	1	4.0
0.05	45	30	66.7	8	17.8	2	4.4	5	11.1
0.10	61	36	59.0	16	26.2	2	3.3	7	11.5
Quantile α	Nr. Virtual TG	Regional Trains - Sunday - 39 data							
		A	%	B	%	C	%	D	%
0.01	27	23	85.2	0	0.0	2	7.4	2	7.4
0.05	53	34	64.2	4	7.5	3	5.7	11	22.6
0.10	74	40	54.1	5	6.8	11	14.9	18	24.3

Table 5.3: Comparison of LSE on the virtual connections (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$

Quantile α	Regional Trains - weekdays - 30 data				
	LSE	A	B	C	D
0.01	1.24	1.35	2.74	6.31E+02	2.83
0.05	0.98	1.08	2.01	4.08E+02	2.19
0.10	1.09	1.23	1.98	7.49E+02	2.35
Quantile α	Regional Trains - weekdays - 195 data				
	LSE	A	B	C	D
0.01	2.18	4.25	34.31	9.96E+06	219.29
0.05	2.91	4.82	22.84	6.09E+06	137.85
0.10	2.77	4.44	14.84	3.65E+06	85.06
Quantile α	Regional Trains - Saturday - 39 data				
	LSE	A	B	C	D
0.01	0.89	0.96	1.23	1.97E+04	1.59
0.05	1.75	1.96	2.76	1.13E+04	3.15
0.10	1.60	1.93	2.72	8.35E+03	3.23
Quantile α	Regional Trains - Sundays - 39 data				
	LSE	A	B	C	D
0.01	0.94	0.99	2.05	1.32E+04	2.41
0.05	2.38	2.64	4.43	1.29E+05	6.74
0.10	2.27	2.60	3.94	9.22E+04	5.85

Table 5.4: Comparison of LSE mean on the virtual connections (no corr.) for the samples $W - 30$, $W - 195$, $Sa - 39$ and $Su - 39$

Quantile α	Nr. Edges TG	Regional Trains - Mon-Wed-Fri - 117 data							
		A	%	B	%	C	%	D	%
0.01	443	160	36.1	154	34.8	21	4.7	108	24.4
0.05	656	212	32.3	246	37.5	53	8.1	145	22.1
0.10	893	249	27.9	353	39.5	101	11.3	190	21.3
Quantile α	Nr. Edges TG	Regional Trains - Tue-Thu - 78 data							
		A	%	B	%	C	%	D	%
0.01	344	141	41.0	101	29.4	18	5.2	84	24.4
0.05	502	191	38.0	157	31.3	47	9.4	107	21.3
0.10	679	233	34.3	226	33.3	79	11.6	141	20.8
Quantile α	Nr. Edges TG	Regional Trains - Monday - 39 data							
		A	%	B	%	C	%	D	%
0.01	253	113	44.7	73	28.9	11	4.3	56	22.1
0.05	390	166	42.6	112	28.7	32	8.2	80	20.5
0.10	545	212	38.9	170	31.2	57	10.5	106	19.4

Table 5.5: Comparison of LSE on edges TG (no corr.) for the samples *MoWeFr* – 117, *TuTh* – 78 and *Mo* – 39

Quantile α	Regional Trains - Mon-Wed-Fri - 117 data				
	LSE	A	B	C	D
0.01	29.74	87.03	42.63	4.22E+06	69.48
0.05	22.24	64.05	31.59	2.85E+06	50.75
0.10	17.16	48.39	24.13	2.10E+06	38.86
Quantile α	Regional Trains - Tue-Thu - 78 data				
	LSE	A	B	C	D
0.01	0.99	1.77	1.44	7.85E+03	1.96
0.05	1.54	3.10	2.22	5.84E+03	3.29
0.10	1.67	3.05	2.36	4.39E+03	3.72
Quantile α	Regional Trains - Monday - 39 data				
	LSE	A	B	C	D
0.01	0.46	1.48	1.02	3.87E+05	1.14
0.05	0.70	1.44	1.25	3.28E+05	1.47
0.10	0.85	1.50	1.66	2.95E+05	2.06

Table 5.6: Comparison of LSE mean on edges TG (no corr.) for the samples *MoWeFr* – 117, *TuTh* – 78 and *Mo* – 39

Quantile α	Nr. Virtual TG	Regional Trains - Mon-Wed-Fri - 30 data							
		A	%	B	%	C	%	D	%
0.01	40	18	45.0	17	42.5	1	2.5	4	10.0
0.05	66	26	39.4	30	45.5	3	4.5	7	10.6
0.10	91	32	35.2	39	42.9	7	7.7	13	14.3
Quantile α	Nr. Virtual TG	Regional Trains - Tue-Thu - 39 data							
		A	%	B	%	C	%	D	%
0.01	21	8	38.1	11	52.4	1	4.8	1	4.8
0.05	41	12	29.3	22	53.7	5	12.2	2	4.9
0.10	63	16	25.4	32	50.8	7	11.1	8	12.7
Quantile α	Nr. Virtual TG	Regional Trains - Monday - 39 data							
		A	%	B	%	C	%	D	%
0.01	17	9	52.9	5	29.4	2	11.8	1	5.9
0.05	32	14	43.8	9	28.1	4	12.5	5	15.6
0.10	52	20	38.5	20	35.8	7	13.5	5	9.6

Table 5.7: Comparison of LSE on the virtual connections (no corr.) for the samples *MoWeFr* – 117, *TuTh* – 78 and *Mo* – 39

Quantile α	Regional Trains - Mon-Wed-Fri - 117 data				
	LSE	A	B	C	D
0.01	2.59	8.39	91.78	4.66E+07	373.44
0.05	2.41	6.22	56.61	2.82E+07	228.25
0.10	3.03	6.65	42.52	2.05E+07	167.97
Quantile α	Regional Trains - Tue-Thu - 78 data				
	LSE	A	B	C	D
0.01	0.89	1.10	2.09	18.93	3.33
0.05	1.63	2.18	2.64	42.54	4.20
0.10	1.94	2.73	2.80	65.19	4.20
Quantile α	Regional Trains - Monday - 39 data				
	LSE	A	B	C	D
0.01	0.64	1.06	3.37	14143.59	3.92
0.05	0.93	1.27	2.79	7531.67	3.44
0.10	1.10	1.46	3.91	14693.54	4.71

Table 5.8: Comparison of LSE mean on the virtual connections (no corr.) for the samples *MoWeFr* – 117, *TuTh* – 78 and *Mo* – 39

Alternative B seems to be preferable when there are many data at hand. To check this assertion we consider three subsamples of the data set corresponding to the second strategy, i.e. the three subsets already considered in Section 4.3: *MoWeFr* – 117, *TuTh* – 78 and *Mo* – 39.

Diminishing the number of observations for the sample, alternative A seems to be preferable to alternative B. However it is not clear how many data we should have to favor one of these alternatives. Moreover, observing the outputs there seems to be a “dislike” of alternative C. This can be explained remembering that the covariance is a measure of the linear dependency between two variables, hence it tends to smooth the irregularities. With the least squares procedure we weighted a lot the outliers of the data set (since we summed the squares of the vertical distances of the observations from the linear regression line).

Outliers and influential data should not be ignored (at least without any further investigation), but they are problematic in linear models fit by least squares because they can substantially influence the results of the analysis, and because they may indicate that the model fails to capture important features of the data.

Unexpected was the position of alternative D, as penultimate. According to the definitions given in Section 5.2, we expected it to be the most promising alternative among the four. To check the correctness of Definition 5.4 and to make a survey of the various approaches, we evaluate the mean of the values of the intercept and of the slope of the four linear regressions and we compare that with the one of the Least Squares Error regression. Moreover to evaluate the impact of outliers we also considered two further linear regressions (Huber’s method and Least Trimmed Squares) as explained in the following section.

The results are to be found in tabular form at the end of this chapter (see Tables 5.9 - 5.16).

5.6.4 Robust Estimators

Robust estimation is an alternative approach to outliers in heavy-tailed error distributions. Properly formulated, robust estimators are almost as efficient as the least squares error when the error is normally distributed, and much more efficient when the error is heavy-tailed. In other words, these methods have a higher breakdown point, i.e. the fraction of arbitrary “bad” data that the estimator can tolerate without being itself affected to an arbitrarily large extent.

The simplest robust estimator is the *least absolute values* (LAV) objective function, that corresponds to the sum of the absolute values

$$\sum_{i=1}^n \rho_{LAV} = \mathbb{E}|Y - (aX - b)| = \sum_{i=1}^n \frac{|y_i - (ax_i + b_i)|}{n}$$

This function is minimized by the median, which is much more resistant to outliers than the mean (its breakdown point is 50% against the 0 breakdown point of the mean), but it is less powerful than the mean if the errors are normally distributed. Moreover it has an asymptote in 0, which makes a weighting approach difficult, especially in the railway problem where we do expect a lot of observations being (close to) zero, i.e. punctuality.

Therefore we preferred to consider other objective functions which combine resistance to outliers with greater robustness of efficiency.

Huber’s method (H)

The Huber objective function is a compromise between least squares and least absolute values, behaving like the least squares in the center and like least absolute values

in the tails:

$$\rho_H = \begin{cases} \frac{1}{2}(Y - (aX - b))^2 & \text{for } |Y - (aX - b)| \leq k \\ k|Y - (aX - b)| - \frac{1}{2}k^2 & \text{for } |Y - (aX - b)| > k \end{cases}$$

The problem consists in minimizing the sum of these errors $\min_{i \in \{1, \dots, n\}} \rho_H(i)$.

The value k , which defines the center and the tails, is called tuning constant. A natural choice is the double of the median absolute deviation, i.e. $k \simeq 2 \text{median}_i |y_i - (ax_i - b)|$ that corresponds to a 95% efficiency relative to the sample mean when the population is normally distributed.

Least Trimmed Squares (LTS)

We report the method as explained in [29]. Given any possible choice of the parameters (a, b) , the squared residuals $[y_i - (ax_i + b)]^2$ are ordered from the smallest to the largest, i.e. $[y_{(1)} - (ax_{(1)} + b)]^2 \leq [y_{(2)} - (ax_{(2)} + b)]^2 \leq \dots \leq [y_{(n)} - (ax_{(n)} + b)]^2$. Then the coefficients a and b are selected in such a way to minimize the sum of the smallest q squared residuals, that is

$$\sum_{i=1}^q [y_{(i)} - (ax_{(i)} + b)]^2$$

where q is some number smaller than n . The (high) breakdown point depends on how q is chosen. Usually $q = \lfloor \frac{n}{2} \rfloor + 1$ (where the floor brackets indicate rounding down to the next smallest integer, i.e. floor function). For this choice of q the breakdown point is nearly 50%.

5.6.5 Slope and intercept

We evaluated the slope a and intercept b of the linear regression $\tilde{Y} = aX + b$ according to the methods presented in the previous sections:

- least squares error (LSE);
- Huber's method (H);
- least trimmed squares (LTS);
- alternative A;
- alternative B;
- alternative C;
- alternative D.

Our aim was to check if the slope a was close to one, as required by the delay definitions given in Section 5.2.

We calculated the mean of the slope and intercept coefficients of the linear regression for the samples $W - 30$, $W - 195$, $Sa - 39$, $Su - 39$ and the subsamples $MoWeFr - 117$, $TuTh - 78$ and $Mo - 39$. As in the previous sections we considered the output of the Tri-graph method without multistatistical correction. Furthermore we subdivided the outcome of the Tri-graph into three classes: waiting activities, driving activities and "virtual" connections (neglecting all the edges belonging to error group) as explained in Section 4.2 to see if there were substantial differences in the value of

the slope.

Regarding the waiting activities the (mean of the) slope coefficient is really close to one. This is in accordance with the small values (or the absence) of slack times associated to this activity so that we can explain the deviation from one by measurement errors. Hence, alternative D seems theoretically and empirically a good choice to represent these activities.

Concerning the driving and the virtual activities, the dependence is not so prominent as for the waiting activities. In the mean the coefficient a is smaller than one for both types of activities. This can be due to the slack times used in practice (for driving activities the slack is roughly 17% of the traveling time) and to changes in the priority list during the disposition process: if a train has a huge delay, connections will not be kept and the scheduled order of departures or arrivals will be changed.

To summarize we consider alternative A the most flexible regression for virtual constraints. However, there is a potential also in alternative B that should be further investigated, especially if the number of observations is big “enough”. Alternative C is too much subject to outliers to become really interesting. Alternative D seems more suitable for waiting activities.

According to alternative A, we may represent the virtual activities as

$$Y_j \geq \mathbb{E}(Y_j) + \sqrt{\frac{\text{Var}Y_j}{\text{Var}Y_i}}(Y_i - \mathbb{E}(Y_i))$$

Since $x_i = \pi_i + y_i$ we can rewrite them as

$$(X_j - \pi_j) \geq \mathbb{E}(Y_j) + \sqrt{\frac{\text{Var}Y_j}{\text{Var}Y_i}}((X_i - \pi_i) - \mathbb{E}(Y_i))$$

that is

$$\begin{aligned} X_j &\geq (\pi_j + \mathbb{E}(Y_j)) + \sqrt{\frac{\text{Var}Y_j}{\text{Var}Y_i}}(X_i - (\pi_i + \mathbb{E}(Y_i))) \\ &\geq \mathbb{E}(X_j) + \sqrt{\frac{\text{Var}X_j}{\text{Var}X_i}}(X_i - \mathbb{E}(X_i)) \end{aligned}$$

Hence we write the model [CTM-3] as:

$$\begin{aligned} \min \quad & \sum_{i \in \mathcal{E}} x_i \\ & x_j - x_i \geq L_a && \forall a = (i, j) \in \mathcal{A} \\ & x_i \geq \pi_i + d_i && \forall i \in \mathcal{E} \\ & x_j \geq v_{ij}^1 x_i + v_{ij}^2 && \forall i, j \in \mathcal{E} \text{ such that } a = (i, j) \in \mathcal{A}^{\text{virtual}} \\ & x_i \in \mathbb{Z}^+ && \forall i \in \mathcal{E} \end{aligned}$$

where

- $\mathcal{A}^{\text{virtual}}$ is the output of the Tri-graph method;
- $v_{ij}^1 = \sqrt{\frac{\text{Var}X_j}{\text{Var}X_i}}$ slope of the regression;
- $v_{ij}^2 = \mathbb{E}(X_j) - v_{ij}^1 \mathbb{E}(X_i)$ intercept of the regression.

Mean of slope and intercept

Regional Trains - week - 30 data															
	Edges	LSE		H		LTS		A		B		C		D	
α	TG	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	463	1.5	0.08	1.5	0.06	1.19	0.05	1.91	0.03	0.91	0.07	6.09	-10.32	1	0.01
5%	453	1.1	0.11	1.06	0.08	0.85	0.06	2.17	0.04	0.84	0.1	5.13	-8.61	1	0.01
10%	1085	1.18	0.14	1.1	0.1	0.88	0.05	2.14	0.02	0.78	0.13	4.74	-8.07	1	-0.01
Regional Trains - week - 195 data															
	Edges	LSE		H		LTS		A		B		C		D	
α	TG	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	491	0.72	0.21	0.66	0.1	-0.41	0.08	1.23	-0.13	0.75	0.14	5.77	-10.68	1	-0.03
5%	745	0.65	0.26	0.52	0.12	1.18	0.11	1.26	-0.12	0.65	0.21	4.63	-7.75	1	-0.06
10%	1002	0.56	0.31	0.44	0.15	1.12	0.07	1.41	-0.13	0.57	0.26	3.96	-6.26	1	-0.07
Regional Trains - Saturday - 39 data															
	Edges	LSE		H		LTS		A		B		C		D	
α	TG	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	327	1.1	0.1	1.01	0.07	0.26	0.06	1.25	0.04	0.91	0.1	262.45	8128.51	1	0.04
5%	517	1.15	0.09	0.99	0.06	0.36	0.04	1.42	0	0.85	0.11	328.72	10279.49	1	0.02
10%	665	1.1	0.12	0.94	0.07	0.53	0.04	1.49	0.01	0.82	0.14	319.14	10032.55	1	0.03
Regional Trains - Sunday - 39 data															
	Edges	LSE		H		LTS		A		B		C		D	
α	TG	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	366	-0.11	0.27	0.39	0.1	0.81	0.04	2.55	-0.39	0.89	0.01	39.97	280.03	1	-0.04
5%	558	0.31	0.23	0.57	0.08	0.46	0.05	2.2	-0.26	0.84	0.03	28.23	179.21	1	-0.01
10%	755	0.48	0.22	0.62	0.09	0.35	0.04	2.06	-0.21	0.8	0.08	22.37	129.91	1	0

Mean of slope and intercept															
Regional Trains - MoWeFr - 117 data															
α	Edges TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	443	0.68	0.21	0.69	0.09	0.26	0.01	1.53	-0.2	0.76	0.09	4.98	-16.9	1	-0.04
5%	656	0.62	0.25	0.58	0.11	0.87	0.01	1.59	-0.23	0.66	0.14	4.3	-12.65	1	-0.07
10%	893	0.59	0.29	0.5	0.14	0.6	0.06	1.54	-0.18	0.59	0.22	3.54	-9.49	1	-0.06
Regional Trains - TuTh - 78 data															
α	Edges TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	344	0.81	0.13	0.76	0.08	0.98	0.04	1.17	0.02	0.82	0.11	6.5	-7.53	1	0.01
5%	502	0.72	0.15	0.64	0.09	0.83	0.03	1.31	-0.02	0.73	0.11	5.58	-6.55	1	-0.05
10%	679	0.66	0.2	0.55	0.12	0.67	0.03	1.43	-0.01	0.65	0.17	4.59	-5.2	1	-0.06
Regional Trains - Monday - 39 data															
α	Edges TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	253	0.77	0.14	0.87	0.09	0.91	0.06	1.26	0.05	0.86	0.12	10.43	-29.81	1	0.05
5%	390	0.78	0.12	0.79	0.08	0.66	0.05	1.25	0	0.79	0.1	8.67	-24.71	1	-0.02
10%	545	0.75	0.13	0.71	0.08	0.56	0.04	1.22	-0.02	0.74	0.09	8.03	-22.88	1	-0.08

Mean of slope and intercept

Regional Trains - week - 30 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	203	0.97	0.16	0.98	0.12	1.13	0.08	1.04	0.11	0.93	0.19	6.45	-10.39	1	0.15
5%	216	0.37	0.17	0.42	0.12	0.19	0.08	2	0.12	0.92	0.2	6.44	-10.3	1	0.15
10%	228	0.4	0.17	0.44	0.12	0.22	0.08	1.96	0.13	0.91	0.2	6.25	-9.9	1	0.15
Regional Trains - week - 195 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	152	0.96	0.24	0.94	0.17	1.11	0.13	1.09	0.17	0.9	0.28	7.11	-7.93	1	0.21
5%	156	0.94	0.25	0.93	0.17	1.65	0.12	1.09	0.17	0.88	0.28	7.04	-7.73	1	0.22
10%	159	0.93	0.26	0.93	0.17	1.52	0.12	1.08	0.18	0.88	0.29	7.02	-7.67	1	0.22
Regional Trains - Saturday - 39 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	149	0.99	0.19	0.98	0.13	0.94	0.09	1.06	0.15	0.94	0.21	7.29	-10.7	1	0.17
5%	161	0.98	0.2	0.96	0.13	0.92	0.1	1.05	0.15	0.94	0.21	7.29	-10.7	1	0.17
10%	164	0.97	0.2	0.96	0.13	0.92	0.09	1.05	0.15	0.93	0.21	261.82	8089	1	0.16
Regional Trains - Sunday - 39 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	154	0.99	0.19	0.98	0.13	1.32	0.06	1.05	0.15	0.94	0.22	74.18	745.08	1	0.19
5%	165	0.98	0.19	0.96	0.13	1.28	0.07	1.05	0.15	0.93	0.21	69.8	694.52	1	0.18
10%	170	0.98	0.19	0.96	0.13	1.27	0.08	1.05	0.15	0.93	0.22	68.01	673.85	1	0.18

Mean of slope and intercept															
Regional Trains - MoWeFr - 117 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	146	0.95	0.23	0.94	0.17	1.29	0.13	1.06	0.17	0.9	0.27	7.38	-10.27	1	0.2
5%	152	0.94	0.25	0.93	0.17	1.09	0.13	1.08	0.17	0.89	0.28	7.16	-9.9	1	0.21
10%	153	0.94	0.25	0.93	0.17	1.24	0.13	1.09	0.17	0.88	0.28	7.12	-9.83	1	0.21
Regional Trains - TuTh - 78 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	143	0.96	0.23	0.93	0.18	0.91	0.14	1.07	0.18	0.9	0.26	7.57	-9.05	1	0.22
5%	146	0.96	0.23	0.93	0.18	0.91	0.14	1.07	0.18	0.9	0.27	7.44	-8.89	1	0.21
10%	148	0.95	0.24	0.92	0.18	0.9	0.14	1.07	0.18	0.89	0.27	7.38	-8.79	1	0.22
Regional Trains - Monday - 39 data															
α	Wait TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	128	0.98	0.19	0.97	0.14	1.07	0.11	1.07	0.13	0.92	0.22	13.26	-37.91	1	0.17
5%	134	0.97	0.2	0.97	0.14	1.04	0.1	1.07	0.14	0.91	0.23	12.73	-36.23	1	0.17
10%	136	0.97	0.2	0.96	0.14	1.06	0.1	1.08	0.14	0.9	0.23	12.55	-35.69	1	0.18

Mean of slope and intercept

Regional Trains - week - 30 data

Drive	LSE		H		LTS		A		B		C		D	
	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1% 77	0.9	-0.06	0.87	-0.05	1.04	0.01	0.98	-0.11	0.92	-0.09	8.96	-19.5	1	-0.15
5% 104	0.88	-0.04	0.85	-0.05	1.11	0	0.96	-0.1	0.92	-0.09	8.4	-16.9	1	-0.15
10% 131	0.87	-0.05	0.84	-0.06	0.94	0	0.95	-0.13	0.91	-0.11	8.17	-16.43	1	-0.19

Regional Trains - week - 195 data

Drive	LSE		H		LTS		A		B		C		D	
	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1% 113	0.86	-0.09	0.78	-0.09	-0.04	0.26	1.03	-0.22	0.85	-0.1	8	-12.11	1	-0.22
5% 131	0.85	-0.09	0.76	-0.09	0.17	0.06	1.06	-0.22	0.83	-0.1	7.77	-11.77	1	-0.22
10% 139	0.85	-0.08	0.75	-0.09	0.47	-0.16	1.06	-0.22	0.83	-0.09	7.58	-11.3	1	-0.22

Regional Trains - Saturday - 39 data

Drive	LSE		H		LTS		A		B		C		D	
	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1% 58	0.89	-0.12	0.87	-0.1	0.23	-0.05	0.96	-0.19	0.93	-0.17	1446.45	45879.01	1	-0.24
5% 88	0.89	-0.13	0.86	-0.1	0.27	-0.04	0.97	-0.2	0.93	-0.17	1446.45	45879.01	1	-0.26
10% 97	0.87	-0.12	0.84	-0.1	0.27	-0.04	0.98	-0.19	0.9	-0.17	1295.83	41131.35	1	-0.25

Regional Trains - Sunday - 39 data

Drive	LSE		H		LTS		A		B		C		D	
	a	b	a	b	a	b	a	b	a	b	a	b	a	b
1% 77	0.89	-0.08	0.87	-0.08	0.31	-0.02	0.98	-0.14	0.92	-0.13	22.24	-89.54	1	-0.19
5% 107	0.89	-0.08	0.85	-0.08	0.2	-0.02	1.01	-0.15	0.89	-0.13	17.14	-65.6	1	-0.21
10% 119	0.87	-0.09	0.83	-0.09	0.2	-0.04	1	-0.16	0.88	-0.14	16.35	-61.86	1	-0.22

Mean of slope and intercept																			
Regional Trains - MoWeFr - 117 data																			
α	Virtual TG	LSE		H		LTS		A		B		C		D					
		a	b	a	b	a	b	a	b	a	b	a	b	a	b				
1%	91	0.85	-0.08	0.79	-0.09	0.3	0	1	-0.2	0.85	-0.1	6.76	-10.13	1	-0.22				
5%	11	0.85	-0.09	0.78	-0.1	0.33	-0.06	1.01	-0.22	0.85	-0.12	7.69	-12.86	1	-0.24				
10%	121	0.85	-0.08	0.78	-0.09	0.21	0.04	1.02	-0.22	0.84	-0.1	7.52	-12.35	1	-0.24				
Regional Trains - TuTh - 78 data																			
α	Virtual TG	LSE		H		LTS		A		B		C		D					
		a	b	a	b	a	b	a	b	a	b	a	b	a	b				
1%	72	0.86	-0.05	0.81	-0.07	0.4	-0.01	0.98	-0.13	0.89	-0.09	7.82	-9.14	1	-0.17				
5%	88	0.87	-0.1	0.8	-0.09	0.34	-0.01	1	-0.19	0.88	-0.13	8.21	-10.61	1	-0.22				
10%	99	0.86	-0.09	0.78	-0.09	0.3	-0.01	1.01	-0.19	0.87	-0.12	7.86	-10.17	1	-0.22				
Regional Trains - Monday - 39 data																			
α	Virtual TG	LSE		H		LTS		A		B		C		D					
		a	b	a	b	a	b	a	b	a	b	a	b	a	b				
1%	39	0.85	0.01	0.69	-0.04	0.07	0.03	0.97	-0.08	0.88	-0.02	15.29	-53.57	1	-0.11				
5%	59	0.81	-0.06	0.67	-0.06	0.11	0.02	0.95	-0.16	0.85	-0.1	12.69	-40.41	1	-0.21				
10%	67	0.82	-0.08	0.68	-0.07	0.11	0.01	0.95	-0.17	0.86	-0.12	16.88	-59.18	1	-0.23				

Mean of slope and intercept

Regional Trains - week - 30 data															
α	Virtual TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	50	1.06	-0.01	1.02	-0.01	0.14	0.04	1.2	-0.15	0.88	-0.17	3.89	-4.44	1	-0.33
5%	84	0.91	0.07	0.88	0.06	0.56	0.07	1.07	-0.07	0.85	-0.08	3.1	-3.35	1	-0.25
10%	135	0.89	0.17	0.82	0.14	0.46	0.08	1.1	0.02	0.81	0.04	3.37	-4.07	1	-0.15

Regional Trains - week - 195 data															
α	Virtual TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	41	0.44	0.42	0.38	0.27	0.17	0.16	0.72	0.21	0.55	0.22	-3.3	-32.08	1	0.1
5%	67	0.45	0.5	0.34	0.32	0.5	0.14	0.87	0.22	0.51	0.36	-1.17	-20.18	1	0.09
10%	112	0.43	0.53	0.3	0.33	0.17	0.1	1.01	0.18	0.45	0.41	-0.21	-12.24	1	0.04

Regional Trains - Saturday - 39 data															
α	Virtual TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	25	0.83	0.12	0.7	0.14	0.53	0.23	1.01	0.04	0.84	0	10.17	-22.56	1	-0.1
5%	45	1.07	0.11	0.83	0.08	0.46	0.16	1.48	-0.05	0.84	0	10.17	-22.56	1	0.01
10%	61	0.87	0.13	0.72	0.08	0.35	0.11	1.49	0.02	0.76	0.13	6.99	-12.36	1	-0.01

Regional Trains - Sunday - 39 data															
α	Virtual TG	LSE		H		LTS		A		B		C		D	
		a	b	a	b	a	b	a	b	a	b	a	b	a	b
1%	27	1.07	0.03	0.83	0.07	0.49	0.12	1.28	-0.04	0.86	-0.1	10.9	-21.89	1	-0.19
5%	53	1.09	0.06	0.81	0.03	0.33	0.04	1.4	-0.11	0.78	0.04	10.5	-26.82	1	-0.14
10%	74	0.98	0.13	0.74	0.06	0.23	0.03	1.34	-0.04	0.72	0.09	7.93	-19.35	1	-0.12

Mean of slope and intercept																			
Regional Trains - MoWeFr - 117 data																			
α	Virtual TG	LSE		H		LTS		A		B		C		D					
		a	b	a	b	a	b	a	b	a	b	a	b	a	b				
1%	40	0.52	0.39	0.49	0.22	0.39	0.15	0.91	0.15	0.59	0.13	-7.06	-96.37	1	0.35				
5%	66	0.53	0.38	0.41	0.22	0.33	0.13	0.97	0.1	0.55	0.21	-3.55	-58.74	1	0.17				
10%	91	0.53	0.44	0.4	0.26	0.27	0.11	1.09	0.09	0.51	0.31	-2.29	-42.74	1	0.18				
Regional Trains - TuTh - 78 data																			
α	Virtual TG	LSE		H		LTS		A		B		C		D					
		a	b	a	b	a	b	a	b	a	b	a	b	a	b				
1%	21	0.62	0.3	0.59	0.19	-0.16	0.18	0.99	0.14	0.69	0.15	1.58	-0.58	1	-0.06				
5%	41	0.52	0.31	0.42	0.2	-0.05	0.15	1.02	0.09	0.57	0.18	1.6	-0.87	1	-0.16				
10%	63	0.3	0.36	0.12	0.23	-0.31	0.11	2.27	0.11	0.5	0.28	1.62	-0.86	1	-0.07				
Regional Trains - Monday - 39 data																			
α	Virtual TG	LSE		H		LTS		A		B		C		D					
		a	b	a	b	a	b	a	b	a	b	a	b	a	b				
1%	17	0.76	0.18	0.12	0.22	0.13	0.08	1.16	0.09	0.65	0.09	4.56	-9.44	1	0				
5%	32	0.79	0.3	0.15	0.31	0.33	0.1	1.13	0.16	0.68	0.19	3.37	-5.56	1	0.02				
10%	52	0.67	0.25	0.15	0.28	0.22	0.07	1.02	0.07	0.64	0.12	5.18	-10.3	1	-0.11				

About the Robustness

*The investigator should have a robust faith,
and yet not believe.*

CLAUDE BERNARD (1813-1878)
French physiologist

As introduced in Chapter 2, we are looking for a new way to identify the (most important) capacity constraints in order to improve the robustness of the (uncapacitated) disposition timetable.

Robustness is the quality of a timetable of being able to withstand (small) delays. A rescheduling system may be said to be “robust” if it is capable of coping well with (unpredictable) variations in trains operating environment with minimal loss of functionality and punctuality.

In this chapter we consider three different criteria to test the robustness of model [CTM-3] presented in Section 5.6.5. Due to lack of information about the system (especially about the considered headways) to whom are referring the delays reported on the files of the Deutsche Bahn presented in Chapter 4, we consider in this chapter another system which has been provided by the Deutsche Bahn within the frame of the ARRIVAL Project (Algorithm for Robust and online Railway optimization: Improving the Validity and reliability of Large scale systems). This system is also focused on the Harz area, but it considers a larger site.

The tests presented in this chapter have been carried out in collaboration with the PhD project of Michael Schachtebeck (AG Optimization, University of Göttingen), who applied different (heuristic) algorithms as well as suggested different ideas on how to check the robustness of the solutions.

6.1 Robustness

It is not trivial to define the concept of robustness for optimization problems. Several definitions have been proposed (see for instance [4], [10] and [21]) to capture all the peculiarities of the problem. Nevertheless the concept of robustness is much broader to be easily enclosed in a definition.

In the most restricted sense, a robust timetable stays unchanged in every likely scenario, that is: the solution provided for one instance (i.e. a particular set of input parameters) should hold even if a perturbation of the parameters is registered. In our case, the solution for the capacitated model without any source delay, i.e. $d_i = 0 \forall i \in \mathcal{E}$, should hold even when some positive source delays are measured, e.g. $\bar{d}_i > 0 \forall i \in \bar{\mathcal{E}} \subseteq \mathcal{E}$. Clearly this definition is really strong and it requires the introduction of huge slack times to be realizable. This would lead to both an “un-

acceptable” increase of the traveling times and a lower level of exploitation of the network (e.g. less trips per day on a line). Hence this definition of robustness is not suitable in our case. A possible compromise is to require the timetable to hold just in case of “small” delays, e.g. delays smaller than three minutes. In fact “small” delays occur very frequently, instead “big” delays are not so common.

Moreover it is reasonable to require that if a perturbation occurs, it will be possible to maintain as much as possible a pre-computed solution taking into account some “soft” recovery strategies. This concept can be expounded through a visual explanation (see [21]).

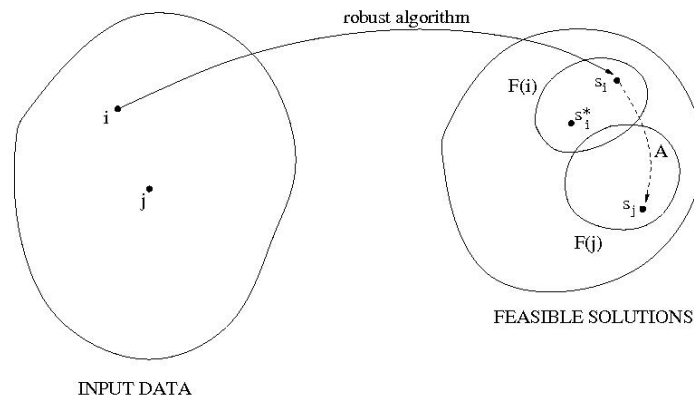


Figure 6.1: Concept of robustness

Figure 6.1 can be read as: given an optimization problem P , a set of possible input parameters and a set of available recovery strategy A_{rec} , we define the corresponding robustness problem R_P . Every instance i of P can be associated to a set of feasible solutions of the problem, $F(i) \ni s_i$. We define s_i^* as the optimal solution for P given the instance i . A robust algorithm takes i as input and outputs a feasible solution to which it is possible to apply a recovery algorithm. That is given a perturbation j of the input parameters, the solution $s_i \in F(i)$ can be transformed into a feasible solution $s_j \in F(j)$ for the input j applying the recovery strategy A . Solution s_j is then called a robust solution. Clearly, robust solutions provided by a robust algorithm can be far from the optimum. The distance between the robust solution and the optimal one is referred as the price of robustness. The aim is to provide the “best” robust solution, the problem is to define what we mean with the word “best”.

6.2 The system

The numerical results that will be further presented are based on data from the Harz Region in the center of Germany. The dataset contains 598 stations, 92 vehicles and 31 lines, each line with two directions, for a period of 2 hours.

The events set is again formed by three kind of events: departures, arrivals and intermediate measurements. Intermediate measurements of the delays are registered in the stations where the trains are passing by without stopping. As in Chapter 4 we marked them as special events.

We considered three cases:

- one period, i.e. two hours (denominated $2HH$);

- two periods, i.e. four hours (denominated $4HH$);
- three periods, i.e. six hours (denominated $6HH$);

that are characterized by

- $2HH$: 3 349 events and 115 trains;
- $4HH$: 6 695 events and 188 trains;
- $6HH$: 10 040 events and 247 trains.

In every period there are a little bit more than one thousand capacity constraints. This number increases to circa 4500 and 9000 for the two-periods samples ($4HH$) and the three-periods samples ($6HH$) respectively, due to new bounds on trains belonging to different periods and turn-overs.

We focused our attention principally on the line between Goslar and Oker both because it is the geographical center of the example presented in Chapter 4 and because it is a double track line (so a quite easy structure). We delayed randomly all the events taking place on this line with source delays between zero and twentyfive minutes:

$$d_i \in [0, 25] \quad \forall i \in \mathcal{E}_{GO}^{\text{dep}} \cup \mathcal{E}_{GO}^{\text{arr}}$$

where

$$\begin{aligned} \mathcal{E}_{GO}^{\text{dep}} &= \{i = (t, v, \text{dep}) \in \mathcal{E}^{\text{dep}} : v \in \{\text{Goslar}, \text{Oker}\}, t \in \mathcal{T}\} \\ \mathcal{E}_{GO}^{\text{arr}} &= \{i = (t, v, \text{arr}) \in \mathcal{E}^{\text{arr}} : v \in \{\text{Goslar}, \text{Oker}\}, t \in \mathcal{T}\} \end{aligned}$$

The cardinality of both sets is equal to 4 in the case of the single period ($2HH$), hence considering the samples $4HH$ and $6HH$ we delayed respectively 16 and 24 events. We generated 133 scenarios with an average delay pro scenario of 643 seconds (minimum source delay one minute, maximum twenty minutes) and applied the same graphical methods presented in Chapter 3 to these datasets. Due to the absence (or to their small values) of slack times for the waiting activities, we had to cope again with the singularity of the covariance matrix, hence the Full Conditional Independence Graph was not applicable. We do not present the results of the Covariance Graph since its output is strictly dependent on the transitivity property. The transformation of the highlighted edges into “virtual” activities of the problem results not only in a large set of constraints but also in a highly dominated set of constraints, i.e. there are many unnecessary constraints. We applied the Tri-graph considering two possible multistatistical corrections or the absence of it (see Section 3.5.5) and three possible values for the quantile, $\alpha \in \{0.01, 0.05, 0.10\}$, as done in Chapter 4. We subdivided the outputs into five groups (see Table 6.1):

- waiting activities;
- driving activities;
- drive2 activities, i.e. edges connecting events of the same trains that can not be classified in one of the previous sets;
- changing activities;
- “virtual” activities.

one period - 2HH												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		change		virtual	
			#	%	#	%	#	%	#	%	#	%
-	0.01	398	26	6.5	50	12.6	291	73.1	0	0.0	31	7.8
	0.05	412	26	6.3	50	12.1	291	70.7	0	0.0	45	10.9
	0.10	429	26	6.1	51	11.9	293	68.3	0	0.0	59	13.8
B	0.01	344	26	7.6	50	14.5	249	72.4	0	0.0	19	5.5
	0.05	356	26	7.3	50	14.0	259	72.8	0	0.0	21	5.9
	0.10	365	26	7.1	50	13.7	267	73.1	0	0.0	22	6.0
BH	0.01	387	26	6.7	50	12.9	285	73.6	0	0.0	26	6.7
	0.05	392	26	6.6	50	12.8	290	74.0	0	0.0	26	6.6
	0.10	394	26	6.6	50	12.7	291	73.9	0	0.0	27	6.9
possible edges		6 555										

two periods - 4HH												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		change		virtual	
			#	%	#	%	#	%	#	%	#	%
-	0.01	1587	99	6.2	192	12.1	1212	76.4	0	0.0	84	5.3
	0.05	1704	99	5.8	193	11.3	1289	75.6	0	0.0	123	7.2
	0.10	1777	99	5.6	194	10.9	1310	73.7	0	0.0	174	9.8
B	0.01	1067	69	6.5	138	12.9	5	0.5	0	0.0	17	1.6
	0.05	1357	99	7.3	192	14.1	8	0.6	0	0.0	26	1.9
	0.10	1368	99	7.2	192	14.0	8	0.6	0	0.0	28	2.0
BH	0.01	1483	99	6.7	192	12.9	14	0.9	0	0.0	48	3.2
	0.05	1514	99	6.5	192	12.7	18	1.2	0	0.0	58	3.8
	0.10	1531	99	6.5	192	12.5	19	1.2	0	0.0	66	4.3
possible edges		101 025										

three periods - 6HH												
Correction	Quantile α	Nr. Edges #	activity									
			wait		drive		drive2		change		virtual	
			#	%	#	%	#	%	#	%	#	%
-	0.01	2963	165	5.6	337	11.4	2190	73.9	4	0.1	267	9.0
	0.05	3270	166	5.1	339	10.4	2376	72.7	6	0.2	383	11.7
	0.10	3423	167	4.9	340	9.9	2456	71.7	6	0.2	454	13.3
B	0.01	1678	123	7.3	245	14.6	1249	74.4	3	0.2	58	3.5
	0.05	2250	163	7.2	336	14.9	1573	69.9	4	0.2	174	7.7
	0.10	2273	163	7.2	337	14.8	1590	70.0	4	0.2	179	7.8
BH	0.01	2257	165	7.3	337	14.9	1850	82.0	4	0.2	201	8.9
	0.05	2671	165	6.2	337	12.6	1951	73.0	4	0.1	214	8.0
	0.10	2721	165	6.1	337	12.4	1993	73.2	4	0.1	222	8.2
possible edges		315 615										

Table 6.1: Outcome of the Tri-graph for the samples 2HH, 4HH and 6HH

It is important to remark that even if all events of the journey of every train are considered (i.e. we do not have any lack of information about the traveling time of the train as it happened in the samples presented in Chapter 4), we still have to face with edges in the set *drive2*. These edges can be referred to the effects of the transitivity property. We did not define a set for “errors”, since it was difficult to formulate a criterion to characterize them. The highlighted edges are spatially and temporally meaningful (i.e. the nodes of every edge refer to events happening in close stations (or in the same station) and with a reasonable time difference according to the distance of the stations). Hence we decided not to portion any further the set of “virtual activities”.

In the case of the sample *6HH*, the “virtual activities” have been transformed into constraints using the linear regressions A, B and D suggested in Chapter 5. Regression C has been omitted due to the poor results concerning the average approximation error. The [CTM-3] model has then been solved and three different criteria to check the robustness of the model have been considered:

1. number of violated capacity constraints of the micromodel (headway function);
2. sum of the cost (in seconds) of the violations (second function);
3. sum of the cost (in seconds) to transform the solution into a feasible one for the micromodel (price function).

By cost we mean the difference between the values of the variables involved in the violated capacity constraints, in the optimal Micromodel solution and in the optimal “virtual model” solution of the problem. Tables 6.2, 6.4 and 6.6 report the number of scenarios in which the corresponding set of “virtual” constraints (characterized by the kind of multistatistical test and by the value of the quantile) gives a better, equal or worse value than the uncapacitated model for the corresponding robustness criterion. Tables 6.3, 6.5 and 6.7 report the minimal, average and maximal percentage of improvement for the scenarios in which the sets of “virtual” constraints give a better solution with respect to the value of the uncapacitated model.

Figures 6.2, 6.3 and 6.4 represent the three robustness criteria as functions (headway, second and price functions). Each of them is composed of twelve pictures having as abscissa values the considered scenarios and as ordinate values the numerical value of the functions. Every column of pictures is characterized by a different linear regression (alternative A, B and D (i.e. constant slope equals to one)) as explained in the previous chapter). The first three rows show the comparison of the function obtained from the Bonferroni (BO), Benjamini-Hochberg (BH) and no multistatistical correction (NC) sets of “virtual constraints” for the three chosen values of the quantile $\alpha \in \{1\%, 5\%, 10\%\}$. Last row presents the comparison of the functions obtained from the sets NC 1%, NC 5% and NC 10% for the three regressions (A, B and D). The scenarios have been ordered according to the values of the functions obtained from the NC sets with $\alpha = 10\%$. We considered three sorted (ascending) criteria corresponding to the three linear regressions (A, B and D). All the twelve pictures in a figure have the same order of scenarios, but the three figures have different sorted scenarios.

Figure 6.5 represents the three robustness criteria as functions (headway, second and price functions) for the uncapacitated model, UM, and for the set NC 5% with the three possible linear regressions (A, B and D). Concerning the second and third criteria it can be easily seen that the values of the functions obtained from the uncapacitated model are bigger than the one obtained from the other sets. Instead for the first criterion it is not possible to show any predominance of the “virtual” set against the uncapacitated model.

slope α_A									
Nr. violated Headways	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	0	0	0	72	72	71	83	29	29
equal	129	129	129	34	34	34	20	28	28
worse	4	4	4	27	27	28	30	76	76

slope α_B									
Nr. violated Headways	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	0	0	0	50	50	62	77	40	37
equal	122	122	122	46	46	38	20	24	25
worse	11	11	11	37	37	33	36	69	71

slope α_D									
Nr. violated Headways	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	0	0	0	65	65	66	77	36	16
equal	118	118	118	30	30	33	17	24	24
worse	15	15	15	38	38	34	39	73	93

Table 6.2: Number of violated headways

slope α_A									
% violated Headways	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	0	0	0	5.56	5.56	5.88	5.56	5.56	5.56
mean %	0	0	0	10.33	10.33	11.92	14.34	13.37	13.37
max %	0	0	0	30.00	30.33	30.00	30.77	31.25	31.25

slope α_B									
% violated Headways	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	0	0	0	5.56	5.56	5.88	5.56	5.88	5.88
mean %	0	0	0	11.70	11.70	12.03	14.37	14.94	15.34
max %	0	0	0	30.00	30.00	30.00	38.46	38.46	38.46

slope α_D									
% violated Headways	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	0	0	0	5.56	5.56	5.88	5.88	5.88	5.88
mean %	0	0	0	10.52	10.52	12.13	14.07	12.82	10.53
max %	0	0	0	30.00	30.00	30.00	30.77	25.00	15.38

Table 6.3: Percentage of violated headways

slope α_A									
Amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	0	0	0	109	109	115	107	116	116
equal	129	129	129	1	1	0	0	0	0
worse	4	4	4	23	23	18	26	17	17

slope α_B									
Amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	0	0	0	111	111	115	116	118	118
equal	122	122	122	0	0	0	0	0	0
worse	11	11	11	18	22	22	17	15	15

slope α_D									
Amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	0	0	0	107	107	113	104	102	94
equal	118	118	118	3	3	1	0	0	0
worse	15	15	15	23	23	19	29	31	39

Table 6.4: Cost in seconds of the violated headways

slope α_A									
% amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	0	0	0	0.55	0.55	0.28	0.06	0.25	0.25
mean %	0	0	0	18.31	18.31	17.59	20.47	25.56	25.56
max %	0	0	0	42.93	42.93	44.47	40.70	56.30	56.30

slope α_B									
% amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	0	0	0	0.40	0.40	0.40	0.04	1.87	1.87
mean %	0	0	0	15.92	15.92	15.97	17.57	22.37	22.30
max %	0	0	0	34.71	34.71	37.75	43.79	46.47	48.03

slope α_D									
% amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	0	0	0	0.39	0.39	0.39	0.41	1.55	1.14
mean %	0	0	0	18.47	18.47	17.79	20.82	22.91	18.97
max %	0	0	0	45.28	45.28	48.32	45.83	53.31	49.43

Table 6.5: Percentage of the cost in seconds of violated headways

slope α_A									
Amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	104	108	108	97	97	98	101	113	113
equal	28	24	24	0	0	0	0	0	0
worse	1	1	1	36	36	35	32	20	20

slope α_B									
Amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	115	125	125	89	89	90	94	116	116
equal	18	0	0	0	0	0	0	0	0
worse	0	0	0	44	44	43	39	17	17

slope α_D									
Amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
better	107	119	119	94	94	95	99	111	107
equal	26	14	14	1	1	0	0	0	0
worse	0	0	0	38	38	38	34	22	26

Table 6.6: Price in seconds of correction

slope α_A									
% amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	2e-04	3e-04	3e-04	0.19	0.19	0.20	0.22	0.04	0.05
mean %	0.13	0.15	0.17	13.69	13.70	13.68	14.67	28.61	28.63
max %	0.52	0.52	0.64	80.41	80.41	80.41	82.96	93.05	93.05

slope α_B									
% amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	1e-04	2e-04	2e-04	0.32	0.32	0.32	0.01	0.53	0.57
mean %	0.29	0.30	0.34	15.00	15.00	15.00	16.53	28.48	28.61
max %	1.41	1.41	1.48	80.72	80.72	80.72	82.22	89.51	89.33

slope α_D									
% amount seconds	BO			BH			NC		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
min %	3e-04	4e-04	4e-04	5e-04	5e-04	0.08	0.08	0.24	1.24
mean %	0.41	0.40	0.44	14.45	14.51	14.50	15.49	28.76	19.08
max %	1.99	1.99	1.99	80.26	80.26	80.26	82.82	85.50	83.56

Table 6.7: Percentage of price in seconds of correction

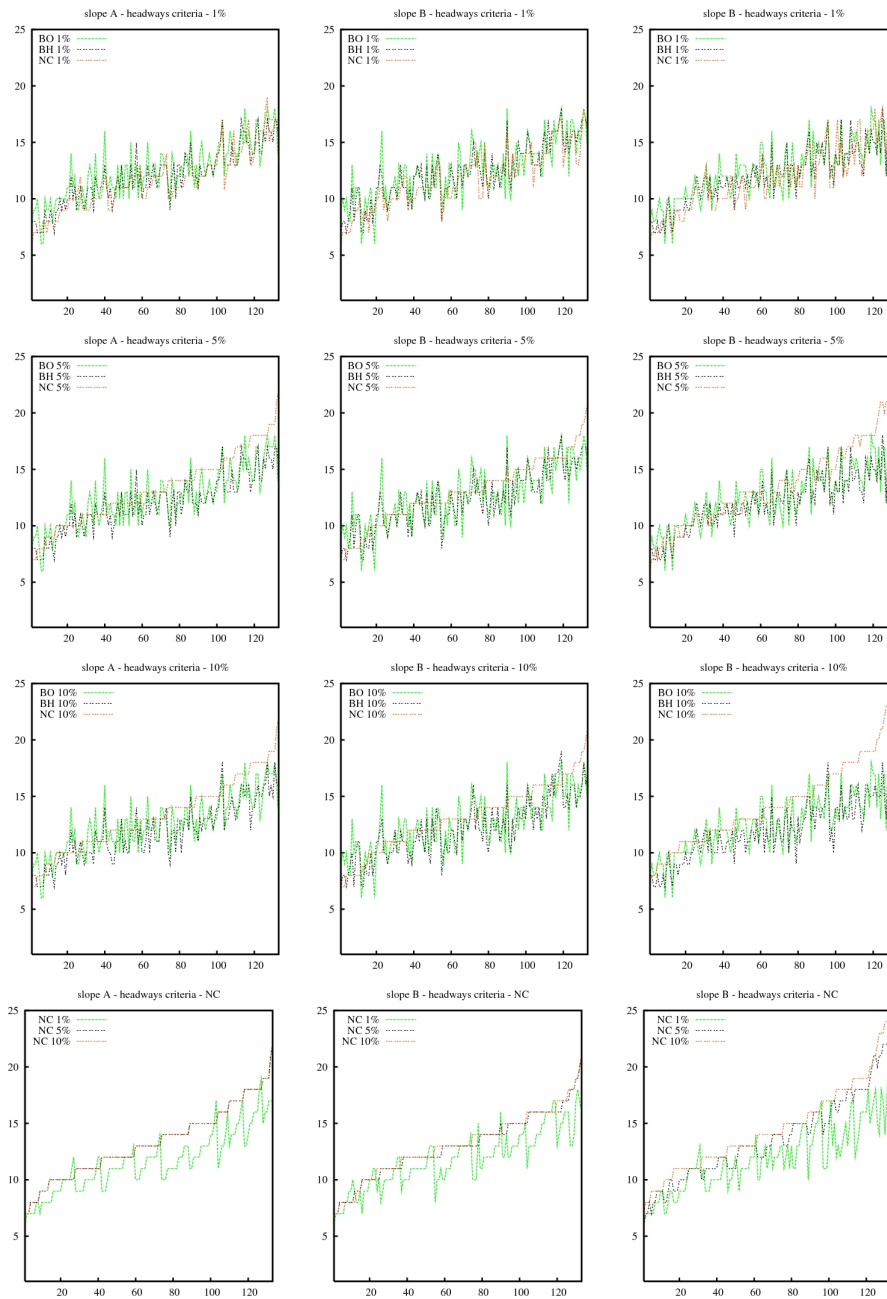


Figure 6.2: Robustness criteria: headways function

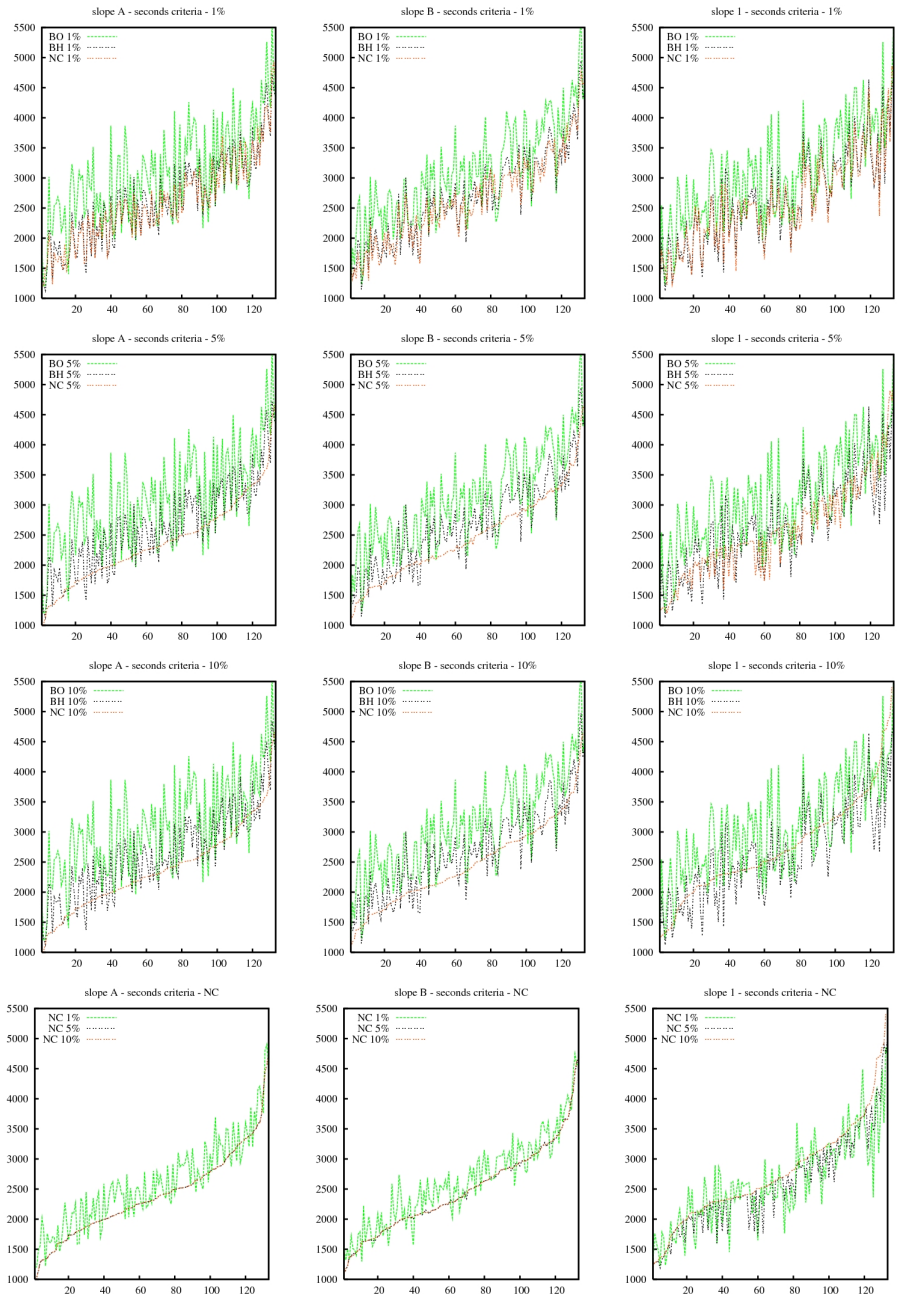


Figure 6.3: Robustness criteria: second function

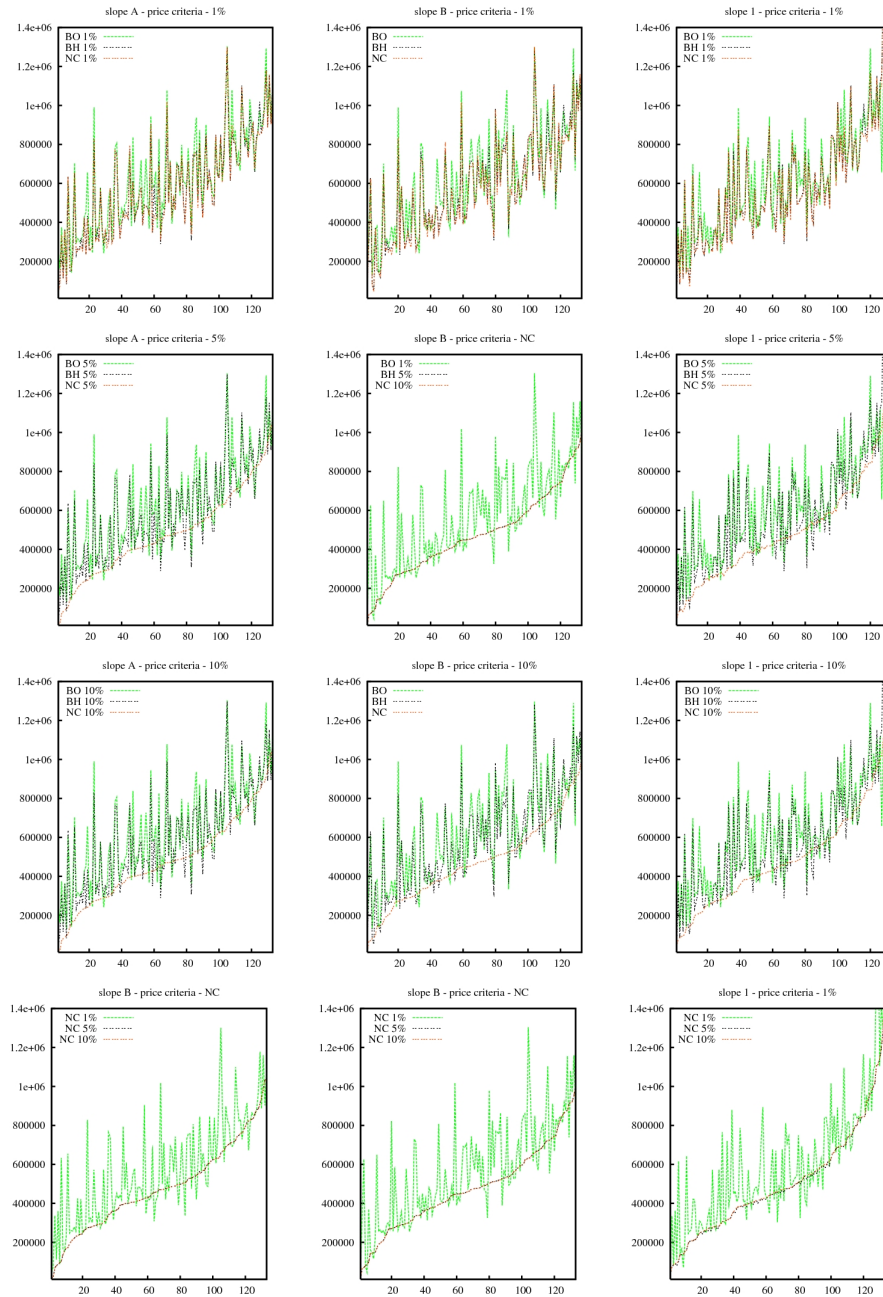


Figure 6.4: Robustness criteria: price function

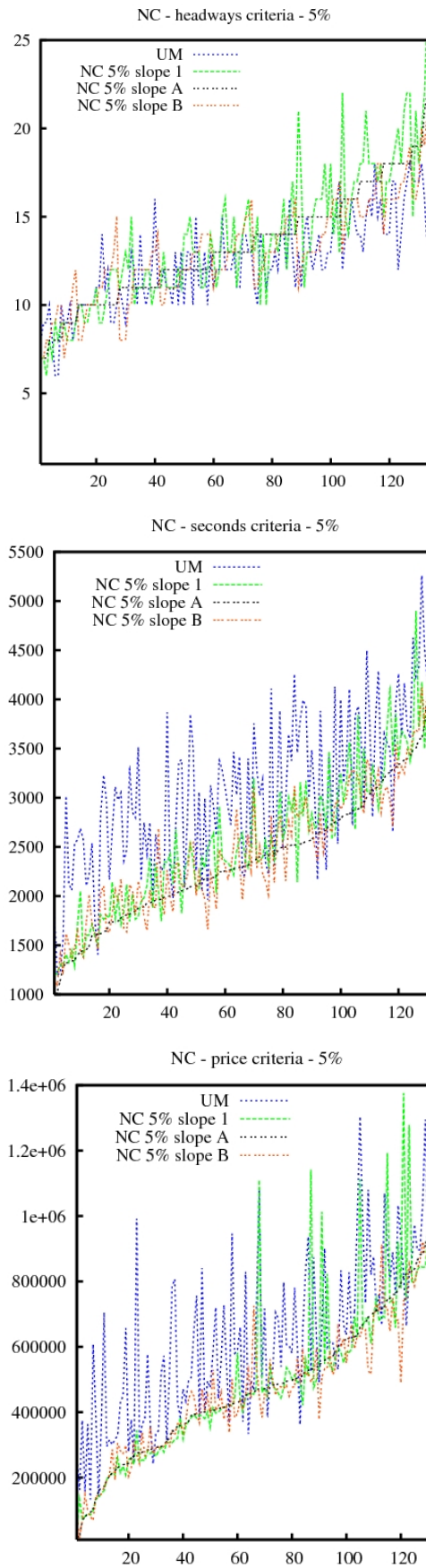


Figure 6.5: Robustness criteria: Comparison for NC 5%

		Headways			Second			Price		
		A	B	D	A	B	D	A	B	D
BO	1%	0	0	0	0	0	0	104	115	107
	5%	0	0	0	0	0	0	108	125	119
	10%	0	0	0	0	0	0	108	125	119
BH	1%	72	50	65	109	111	107	97	89	94
	5%	72	50	65	109	111	107	97	89	94
	10%	71	62	66	115	115	113	98	90	95
NC	1%	83	77	77	107	116	104	101	94	99
	5%	29	40	36	116	118	102	113	116	111
	10%	29	37	16	116	118	94	113	116	107
Total number of scenarios : 133										

Table 6.8: Nr. of scenarios with improvement w.r.t. the uncapacitated model

The first criterion considered for the robustness (i.e. the number of violated capacity constraints) does not show a significant improvement (just in the half of the scenarios there are betterments, also like a fifty-fifty probability). The source delays we generate are not so big that the delay propagation extends itself to the whole considered area, hence the amount of capacity constraints involved in the problem is limited. Nevertheless it is really useful to show the weakness of the Bonferroni correction, which is not able to improve the solution of the uncapacitated model in any test (see Tables 6.2 and 6.3). The improvements related to the other corrections are not so significant to be considered as reasons to state the predominance of a particular set of “virtual” constraints (as it can be seen in Figure 6.2).

Concerning the second criterion (i.e. the cost in seconds of the violations) we can see that both Benjamini-Hochberg and no multistatistical correction improve the uncapacitated solution in more than 80% of the scenarios. The linear regression denoted by B is more effective, improving the results in almost 90% of the scenarios. The Bonferroni correction instead does not show any sign of betterments (see Tables 6.4 and 6.5 and Figure 6.3). Figure 6.3 shows clearly how the BO set performs worse than the other sets, independently from the choice of quantile. The same figure also shows how the NC sets overcome in performance the BH sets in most of the scenarios. Moreover the values of the functions corresponding to the 1% NC sets are evidently higher than the values of the sets 5% NC and 10% NC. It is not possible to highlight any strong difference between the sets NC 5% and NC 10%.

Only considering the third robustness criterion the Bonferroni correction shows some improvements compared to the solution of the uncapacitated model, but they are so small that we can neglect them. Instead both Benjamini-Hochberg and no multistatistical correction enhance the solutions up to 90% betterment (see Tables 6.6 and 6.7 and Figure 6.4). Figure 6.4 shows again a predominance of the NC sets against the BH sets, and the similarity of the performance of NC sets in the case of 5% and 10% quantile.

It is important to remark that the improvement in one criterion does not always imply an improvement in the other two criteria. We found, however, a positive correlation (with values 0.6) between the improvements in the second and third criteria. For the BH and NC sets we observed in 75% of the scenarios an enhancement in both the criteria. In the best case, when the price of correction is improved up to 90%, the value of the second function is reduced of almost 85%. Concerning the capacity constraints we can not identify any significant correlation with the other two robustness criteria.

6.3 Final consideration

In average we could observe an improvement of 10% - 17% (with respect to all three considered criteria) in the case of Benjamini-Hochberg multistatistical correction. Avoiding multistatistical corrections delivers a slightly bigger improvement (15% - 20% for all the considered criteria).

Concerning the linear regression it is not easy to state which one of the suggested models (A, B or D) is better. Their average values are really close to each other.

Alternative B seems to be slightly stronger with respect to the first two criteria. This predominance can be due to the large amount of simulations executed (133 scenarios can be seen as four and a half months of daily registrations), as already observed in Chapter 5.

Alternative D also shows a quite good performance, which we ascribe to the definition of very small slack times. The considered stations are really close to each other, hence the traveling times are seldom bigger than fifteen minutes, especially thanks to the intermediate delay measurements (events marked as special).

Even if the results do not show a clear predominance of our suggested model (no multistatistical correction and regression A), this set of “virtual” constraints provides a good improvement on the robustness of the solutions. The choice of a quantile of 10% does not bring a reasonable enhancement if compared to a 5% quantile.

To summarize, we do suggest to transform the dependencies corresponding to the Tri-graph output without multistatistical correction and with a 5% quantile into “virtual” constraints of the model, by the application of the linear regression A.

Conclusion

Ich verstehe nur Bahnhof.

German idiomatic expression

In this thesis we have been concerned with the analysis of the capacity constraints in the Delay Management problem for the railway system, from both a theoretical and a practical point of view. As the title suggests, our aim was to find a new formulation for the bonds, that could be inserted in the corresponding uncapacitated Macromodel of the problem, increasing the robustness of the solution without requiring a detailed knowledge of the system infrastructure.

The Delay Management problem has been described as a directed graph (Event-Activity-Graph) and it has been proved to be \mathcal{NP} -complete. Our formulation of the problem has been compared to the ones of other research groups to highlight similarities, advantages and drawbacks. Altogether it appears easily readable and adaptable to represent all the aspects of the problem.

Core of the work was the investigation of the delay dataset through a stochastic analysis. The measured arrival and departure delays have been considered as observations of random variables, to which different stochastic algorithms (in particular three graphical models) have been applied, in order to determine the dependencies among the disturbances. Three methods have been used for the first time on the railway problem: two of them are classical procedures (Full Conditional Independence Graph and Covariance Graph) while the third one (Tri-graph) has been introduced only in 2004 from Wille and Bühlmann (ETH Zürich). This method has in this thesis one of its first applications outside the genetic field.

The graphical method procedures for the identification of the constraints have been implemented in R and validated through applications to datasets reporting train delays in the Harz area. These data have been placed at disposal of the Optimization Group (Prof. A. Schöbel) of the NAM Institute in the context of the DisKon project (*Disposition und Konfliktlösungsmanagement für die beste Bahn*). Different samples have been tested to evaluate the influence of the amount of variables and observations on the methods. The results obtained either using a multistatistical test (Bonferroni, Benjamini-Hochberg) or without correction, have been compared for different values of the quantile (1%, 5% and 10%). The edges pointed out by the Tri-graph have been subdivided into four groups corresponding to waiting, driving, virtual activities and errors. Due to the lack of information about the timetable on which the data are based, it has not been possible to make any statement on the changing activities. Further comparisons have been carried out just on the pointed “virtual” activities.

The Full Conditional Independence Graph method could not be applied but in small samples, due to the singularity of the covariance matrix. The Covariance Graph is

penalized by the transitivity property of the covariance, which largely increases the cardinality of the output. The Tri-graph method, although not completely free from the transitivity property, is able to confine its effects. When applied to a “large” data sample the Tri-graph points out up to 90% of the original waiting activities and nearly 80% of the driving activities. Even if the size of the sample is reduced to one fifth of the original, the Tri-graph is still able to capture 40% of the “virtual” activities.

As common in mathematics, we had to face with some challenging problems along the way. Firstly, the assumption on the distribution of the delays. The analysis we carried out could not prove the normal distribution of the variables required by the graphical methods, but up to now it has not been possible to find any distribution that fits the delays data in any possible contest. We can trace this problem back to the unavailability of a standard procedure to measure the delays. Further researches on this direction are recommended, to analyze properly the behavior of the railway system.

Secondly, the transformation of the edges pointed out by the Tri-graph method into time constraints of the uncapacitated model. The solution we considered is a linear regression on the pair of variables corresponding to each edge. This choice is quite satisfactory, since it is coherent with the theoretical definition of the delay and the slope of the approximation results really close to one in case of waiting activities. In general, we should deeply consider the influence of the slack times in these constraints, especially in the case of driving and “virtual” activities.

Additional investigations on the robustness of the new formulation should be conducted. A first step on this directions has already been done in the last chapter of this thesis, where an analytical comparison between the Macromodel and our Meso-model (i.e. Macromodel plus “virtual” constraints) has been performed. The concept of robustness is however too broad to consider our analysis exhaustive. Nevertheless, our model improves by 18% the robustness of the uncapacitated solution according to the three considered criteria (number of violated headways, cost in seconds of the violations and price in seconds to correct the violations).

Though the literature on capacitated problems and delay propagation is abundant, we did not find any other research group that applied stochastic approaches to identify the “important” capacity constraints of the model (whatever important means). The stochastic analysis was limited to identify the distribution of the delays so that this information could be used in the simulation of the delay propagation. This makes difficult to judge the results we obtained so far, since they are the first of their kind. But, as a Chinese proverb says, *the journey of a thousand miles must begin with a single step*.



Figure 7.1: ... just an “ironic” conclusion ...



Maximum Likelihood Estimation

I have no objection to the study of likelihood as such.

Sir HAROLD JEFFREYS
English mathematician and statistician (1891-1989)

A.1 A short introduction

The best description of a phenomenon is that which provides the maximum probability of arriving, as a result of measurements, at exactly those values of the quantities being measured which have been actually obtained.

The concept of likelihood (see [24]) is closely related to the concept of *probability* but while the probability speaks about future events, the likelihood is related to past events, since it moves from the outcome.

When we speak about the probability of an event we are simply assuming some kind of model, to which we can apply some *law of probability* to make inferences or predictions based on probabilistic information.

In the real world few things have fixed probabilities, so the notation of conditional probability allows us to incorporate other potentially important variables into the model:

$\mathbb{P}(X|P)$ = the probability of an event X depending on model parameter P

whereas X is the variable and P is constant. The likelihood is the hypothetical probability that a probabilistic model would yield a specific outcome, i.e. an event that has already occurred:

$L(P|X)$ = likelihood of the model parameter P given the data X .

whereas P is the variable and X is fixed. For sensible models it is possible to see that certain data are more probable than other data. The aim of the *Maximum Likelihood Estimator*, MLE, is to find the hypotheses (e.g the parameter values) that make the observed data most likely.

Definition A.1 *The likelihood, $L(H|X)$, of a hypothesis H given the data X and a specific model, is proportional to the probability $\mathbb{P}(X|H)$, the constant of proportionality being arbitrary: $L(H|X) \approx \mathbb{P}(X|H)$*

Without loss of generality, we will consider $L(H|X) = \mathbb{P}(X|H)$. (♠)

A.2 Computational Problem

There are some features of the previous definition which are worth emphasizing.

- Sometimes it is possible to define the likelihood as a differential function, sometimes not, so we have to evaluate it by points. If the model has a lot of parameters, the space (of the parameters) will grow up very quickly and it will be virtually impossible to evaluate the MLE *exhaustively*.
- If we do a lot of tests, we will have as outcome a large set of data, and we can spend a lot of time just evaluating the coefficients of the likelihood function. Since those coefficients are fixed we can omitted them but then, to keep a *meaningful scale* for the likelihood function, we need to consider the *ratio of the likelihood*.
- It can also happen that in the likelihood function it is necessary to multiply a lot of small numbers together. The result will end up very quickly as a number that is too small to be represented by any calculator. A possible solution is to consider the *likelihood logarithm* instead of the likelihood function. In fact one of the property of the logarithmic function is $\log (a \cdot b) = \log (a) + \log (b)$.
When we optimize the *log-likelihood* with respect to the parameters of the model (note: we minimize the negative of the log-likelihood), we also optimize the likelihood since there is a one to one relationship between numbers and their logarithm. We will indicate the log-likelihood as $LL(H|X)$.
- Occasionally there does not exist just a set of parameters that uniquely optimize the log-likelihood, so we can have *under identification models*.
- If the log-likelihood has local minima, to avoid considering one of them as approximation for the parameters, it is necessary to iterate the procedure: some starting points A_i (first guess values for the parameters) are specified and they are given as input to some optimization algorithm that will find the (local) minima B_i .

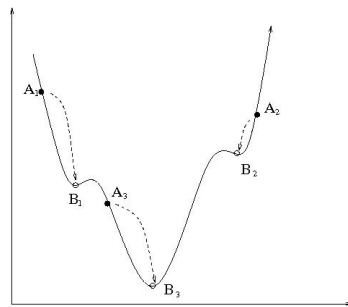


Figure A.1: Likelihood function with local minima

A_i are the first guesses $\forall i$
 B_i (local) optimum found moving from A_i
 We need to compare the B_i in order to choose the (global) optimum, i.e. $\min_i B_i$
 For the success of the procedure we need a well specified model, an appropriate optimization algorithm and sensible starting values.

A.3 The Likelihood Ratio Test

Generally our aim is not to estimate the parameters of the model, we just want to know if they are (or not) different from some specific values (significantly different).

Given a parameter ω , we fix a more restricted null hypothesis $H_0 : \omega \in \Omega_0$ (e.g the parameter ω is equal to zero) and an alternative hypothesis $H_1 : \omega \in \Omega_1$ (e.g the parameter ω is not zero). The hypothesis H_0 corresponds to the “reduced model” and H_1 to the “full model”.

Definition A.2 *The Log-likelihood Ratio of two hypotheses given a sample of data is the ratio of the log-likelihood function evaluated in the two different cases. The ratio is usually quoted as a function:*

$$\lambda(X) = \frac{LL(H_0 | X)}{LL(H_1 | X)}$$

One tends to favor H_0 if the Log-Likelihood Ratio is high, and H_1 if it is low. The Likelihood Ratio indicates in which cases we should favor H_0 over H_1 . The rejection region is

$$R = \{X : \lambda(X) < c\}$$

where c is determined so that $\sup_{\omega \in \Omega_0} \mathbb{P}_\omega(X \in R) = \alpha$. The difficulty is to express c as a function of α , because $\lambda(X)$ might be a complicated function of X . Instead of $\lambda(X)$ we consider the logarithm of $\lambda(X)$.

$$\log(\lambda(X)) = \log\left(\frac{LL(H_0 | X)}{LL(H_1 | X)}\right)$$

Using the logarithm property $\log\left(\frac{a}{b}\right) = \log(a) - \log(b)$, we can rewrite the ratio as

$$\log(\lambda(X)) = LL(H_0 | X) - LL(H_1 | X).$$

or equivalently as $-2 \log(\lambda(X)) = 2[LL(H_1 | X) - LL(H_0 | X)]$. In this case the rejection region is $R = \{X : -2 \log(\lambda(X)) > k\}$ where $k = -2 \log(c)$.

We still need to find an answer to the question: which distribution for λ or $-2 \log(\lambda)$ should we choose in order to compute c or k ?

Theorem A.3 *If $\Omega_0 \subseteq \mathbb{R}^q$ is a q -dimensional space and if $\Omega_0 \subseteq \Omega_1$ is an r -dimensional subspace, then the regularity conditions state*

$$\forall \omega \in \Omega_0 : -2 \log \lambda \rightarrow \chi_{q-r}^2 \quad \text{as } n \rightarrow \infty.$$

The regularity conditions required for this theorem can be found in [34]. They are based on the assumption that the data constitutes an i.i.d. sample (of values in some Euclidean data-space) with the same density (known except for the unknown parameter ω) and that the MLE is locally unique, consistent and asymptotically normal. These conditions include the restriction that Ω_0 contains an open neighborhood of the true value of the parameter ω and that Ω_0 lies in some sufficiently small neighborhood of this true value, not depending upon the size of the data sample. In other words, the main regularity condition assumes that the unknown parameters under the null hypothesis are not on the border of the parameter space of the unknown parameters under the alternative hypothesis.

An asymptotic rejection region can now be given by simply computing the $(1 - \alpha)$ quantile $k = \chi_{1-\alpha, q-r}^2$. The likelihood rejection region is therefore

$$R = \{X : -2 \log(\lambda(X)) > \chi_{1-\alpha, q-r}^2\}$$

The theorem (A.3) gives a general way of building the rejection region, although the result is just asymptotic, i.e. the size of the test is only approximatively equal to α (the approximation becomes better when the sample n increases). A general answer to the question “how large should n be?” is unfortunately not define.

To summarize, if we multiply the Likelihood Ratio test for a factor 2, we get a quantity $2[LL(H_0 | X) - LL(H_1 | X)]$ that has a Chi-squared distribution, χ^2 , with degree of

freedom equal to the difference between the number of parameter being estimated under H_1 and H_0 . (E.g under the hypothesis $H_0 = (p = 0)$ and $H_1 = (p \neq 0)$, the degree of freedom of the Log-Likelihood Ratio is 1) so that the result of the test can be compared to the critical significance level of χ^2 with the calculated degree of freedom. If the value of the Log-Likelihood Ratio test is less then the critical value, we have no reason to reject the null hypothesis.

A.4 Errors

Hypothesis tests are usually done in four steps:

1. formulate the null hypothesis H_0 and the alternative one H_1 ;
2. identify a statistic test (e.g. the Log-Likelihood Ratio test) to say if H_0 is correct or not;
3. compute the P-value, which is the probability of the hypothesis H_0 being true;
4. compare the P-value with a significant value α . If the P-value is less than α , the observed effect is statistically significant, so the alternative hypothesis is valid.

The value of the parameter α is chosen in a four level of significant scale:

- $\alpha = 0.01$ if the P-value $\leq \alpha$ the test is high significant \Rightarrow we reject H_0 ;
- $\alpha = 0.05$ if the P-value $\leq \alpha$ the test is statistically significant \Rightarrow we reject H_0 ;
- $\alpha = 0.10$ if the P-value $\leq \alpha$ the test is is tending toward statistical significance \Rightarrow we usually do not reject H_0 ;
- $\alpha > 0.10$ if the P-value $> \alpha$ the test is not statistically significant \Rightarrow we do not reject H_0 .

Using statistic test to see if a hypothesis is significant or not, we can introduce some errors. We distinguish between two kinds of error:

- Type I: *False Positive*, that is we reject H_0 when it is true.
Possible reason: a not representative sample or presence of outliers.
- Type II: *False Negative*, that is we accept H_0 when it is false.
Possible reason: a small size of the sample or a large variance of the data.

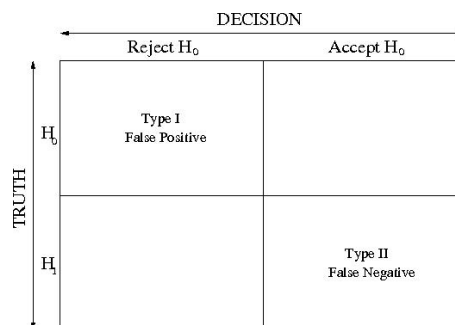


Figure A.2: MLE errors of type I and II

Bibliography

- [1] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela and M. Protasi, *Complexity and approximation. Combinatorial optimization problems and their approximability properties*, Springer, Berlin, Germany, 1999.
- [2] K. Backhaus, B. Erichson, W. Pilke and R. Weiber, *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*, Springer, Berlin, 1986.
- [3] P. Baldi, *Calcolo delle probabilità e statistica*, McGraw-Hill Milano, Italy, 1998.
- [4] H.G. Bayer and B. Sendhoff, *Robust optimization. A comprehensive survey*, to appear in Computer Methods in Applied Mechanics and Engineering, 2007.
- [5] J.C.M. Beaten, *A brief history of Process Algebra*, Rapport CSR 04-02, Vakgroep Informatica, Technische Universiteit Eindhoven, 2004.
- [6] A. Becker, D. Geiger and C. Meek, *Perfect tree-like Markovian distribution*, Uncertainty in Artificial Intelligence, pp. 19-23, 2000.
- [7] A. Ben-Tal, L. El Ghaoui and A. Nemirovki, *Mathematical programming: special issue on robust optimization*, Vol. 107, Springer, Berlin, 2006.
- [8] Y. Benjamini and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiplying testing*, Journal of the Royal Statistical Society, Series B (Methodological), Vol 57, pp. 289-300, 1995.
- [9] M. Bernardo, L. Donatiello and P. Ciancarini, *Stochastic Process Algebra: from an algebraic formalism to an architectural description language*, Springer, Berlin, 2002.
- [10] D. Bertsimas and M. Sim, *The price of robustness*, Operations Research, Vol. 52, pp. 35-53, 2004.
- [11] N. Bissanz, S. Güttler, J. Jacobs, S. Kurby, T. Schaer, A. Schöbel and S. Scholl, *DisKon – Disposition und Konfliktlösungsmanagement für die beste Bahn*, in Eisenbahntechnische Rundschau (ETR), 12\05, pp. 809-821, 2005.
- [12] J.A. Bergstra and C.A. Middelburg, *Process Algebra for hybrid systems*, Theoretical Computer Science, Vol. 335, pp. 215-280, 2005.
- [13] S. Boyles and S.T. Waller, *A stochastic delay prediction model for real-time incident management*, Daniel B. Fambro Student Paper Award, ITE Journal, 2007.
- [14] F.W. Brian, *Railroad's traffic control systems*, Trains Magazine, 1st May 2006, at www.trains.com

- [15] F.W. Brian and R.S. McGonigal, *Railroad signals*, Trains Magazine, 1st May 2006, at www.trains.com
- [16] P. Brucker, *Scheduling algorithms*, Springer, Berlin, Germany, 2004.
- [17] P. Brucker, S. Heitmann and S. Knust, *Scheduling railway traffic at a construction site*, OR Spectrum, Vol. 24, pp. 19-30, 2002.
- [18] P. Brucker and S. Knust, *Lower bounds for resource-constrained project scheduling problems*, European Journal of Operational Research, Vol. 149, Nr. 2, pp. 302-313, 2003.
- [19] P. Brucker, Y.N. Sotskov and F. Werner, *Complexity of shop-scheduling problems with fixed number of jobs: a survey*, Mathematical Methods of Operations Research, Vol. 65, No. 3, pp. 461 - 481, 2007.
- [20] A. Caprara, M. Monaci, P Toth, M. Fischetti, P.L. Guida and G. Sacco, *Solution of real world train timetabling problem*, Proceedings of the 34-th Hawaii International Conference on System Sciences (HICSS 34), IEEE Computer Society Press, 2001.
- [21] S. Cicerone, G d'Angelo, G. di Stefano, D. Frigioni and A. Navarra, *Robust Algorithms and Price of Robustness in Shunting Problems*, to appear in ATMOS 2007 proceedings, ARRIVAL-TR-0072, 2007.
- [22] S.A. Cook, *The complexity of theorem-proving procedures*, Proceeding of the 3rd annual ACM Symposium on Theory of Computing, ACM Press, pp. 151-158, 1971.
- [23] A.F. de Kort, B. Heidergott, R.J. van Edgmond and G. Hooghiemstra, *Train movement analysis at railway stations: procedure and evaluation of the Wakob's approach*, TRAIL Studies in Transportation Science, S99/1, Delft University Press, Delft, The Netherlands, 1999.
- [24] A. W. F. Edwards, *Likelihood. An account of the statistical concept of likelihood and its application to scientific inference*, Cambridge University Press, 1972.
- [25] D. Edwards, *Introduction to Graphical Modeling*, Springer 2nd edition, New York, 2000.
- [26] A. Fay, *A fuzzy knowledge-based system for railway traffic control*, Engineering Applications of Artificial Intelligence, Nr. 13, pp. 719-729, 2000.
- [27] R.J. Feise, *Behavioral-graded activity compared with usual care after first-time disk surgery: considerations of the design of a randomized clinical trial (Letter)*, Journal of Manipulative and Physiological Therapeutics, Vol. 24, pp. 67-68, 2001.
- [28] R.J. Feise, *Do multiple outcome measures require p-value adjustment?*, BMC Medical research Methodology, Vol. 2, pp 1-8, 2002.
- [29] J. Fox, *Applied regression analysis, linear models and related methods*, Sage Publications, Thousand Oaks, California, 1997.
- [30] M.R. Garey and D.S. Johnson, *Computer and Intractability. A guide to the Theory of NP-Completeness*, W.H. Freeman and Company, New York, 1979.
- [31] M.R. Garey, D.S. Johnson and R. Sethi, *The complexity of flow-shop and the job shop scheduling*, Mathematics of Operations Research, Vol 1, Nr. 2, pp. 117-129, 1976.

- [32] S. Güttler, *Statistical Modeling of Railway Data*, Diploma thesis at Georg-August-Universität, Göttingen, 2006.
- [33] NIST/SEMATECH e-Handbook of Statistical Methods, www.itl.nist.gov/div898/handbook/date
- [34] W. Härdle and L. Simar, *Applied Multivariate Statistical Analysis*, Springer, Berlin, 2003.
- [35] U. Herdamm, *Untersuchung der Verspätungsentwicklung von Fernreisezügen auf der Datengrundlage der rechnerunterstützten Zugüberwachung am Frankfurt am Main*, Doctoral Thesis, Technische Hochschule Darmstadt, Germany, 1996.
- [36] A. Higgins, *Optimization of train schedules to minimize transit time and maximize reliability*, Doctoral Thesis, Queensland University of Technology, Brisbane, Australia, 1996.
- [37] A. Higgins and E. Kozan, *Modeling train delays in urban networks*, Transportation Science, Vol 32, Nr. 4, pp. 346-357, 1998.
- [38] L. Ingolotti, P. Tormos, A. Lova, F. Barber, M.A. Salido and M. Abril, *A decision support system (DSS) for the railway scheduling problem*, AIAI 2004, Toulouse, France, pp. 465-474, 2004.
- [39] J. Jacobs, *Rechnerunterstützte Konfliktermittlung und Entscheidungsunterstützung bei der Disposition des Zuglaufs*, Dissertation, Veröffentlichungen des Verkehrswissenschaftlichen Institutes der RWTH Aachen, Heft 61, 2003.
- [40] U. Khadim, *A comparative study of Process Algebra for hybrid systems*, submitted to Science of Computer Programming, 2006.
- [41] O. Kluge, J. Padberg and H. Ehrig, *Modeling Train Control Systems: From Message Sequence Charts to Petri Nets*, citeseer.ist.psu.edu/kluge00modeling.html
- [42] H.O. Lancaster, *The structure of bivariate distributions*, The Annals of Mathematical Statistics, Vol. 29, No. 3, pp 719-736, 1958.
- [43] H.O. Lancaster, *Correlations and canonical forms of bivariate distributions*, he Annals of Mathematical Statistics, Vol. 34, No. 3, pp 719-736, 1958.
- [44] H.O. Lancaster, *Correlation and complete dependence of random variables*, The Annals of Mathematical Statistics, Vol. 34, Nr. 2, pp. 532-538, 1963.
- [45] S.L. Lauritzen, *Graphical Models*, Oxford Science Publication, Great Britain, 1996.
- [46] R. Leipnik, *When does zero correlation imply independence?*, The American Mathematical Monthly, Vol. 68, No. 6, pp. 563-565, 1961.
- [47] S. Leonardi and D. Raz, *Approximating total flow time on parallel machines*, Proc. 29th Ann. ACM Symp. on Theory of Comp., ACM, 110-119, 1997.
- [48] H.W. Lenstra, *Integer Programming with fixed number of variables*, Mathematics of Operations Research, Vol 8, Nr. 4, pp. 538-548, 1983.
- [49] L.G. Mattsson, *Train service reliability. A survey of methods for deriving relationship for train delays*, Swedish Institute for Transport and Communications Analysis (SIKA), Stockholm, Sweden, 2004.

- [50] Y. Meng, *Bemessung von Pufferzeiten in Abschlüssen von Reisezügen*, Veröffentlichungen des Verkehrswissenschaftlichen Instituts der RWTH, Aachen, 1991.
- [51] K. Nachtigall, *Periodic Network Optimization and fixed intervals timetables*, Technical report, Deutsche Zentrum für Luft- und Raumfahrt, Institut für Luft- und Raumfahrt, Braunschweig, 1998.
- [52] K. Neumann and C. Schwindt, *Project Scheduling with inventory constraints*, Karlsruhe University, Germany, 2002.
- [53] P.C. O'Brian, *Procedures for comparing samples with multiple endpoints*, *Biometrics*, Vol. 40, pp. 1079-1087, 1984.
- [54] J. Pachl, *German Block and Interlocking Principles*, at joernpachl.gmxhome.de
- [55] J. Pachl, *Systemtechnik des Schienenverkehrs. Bahnbetrieb planen, steuern und sichern*, Teubner, 2004.
- [56] The program R can be for free downloaded at: www.r-project.org The necessary packages to implement the graphical methods can be found at: cran.r-project.org
- [57] J. Remy, *Resource constrained scheduling on multiple machines*, ETH Zürich, Switzerland, 2004.
- [58] K.J. Rothman, *No adjustments are needed for multiple comparisons*, *Epidemiology*, Vol. 1, pp. 43-46, 1990.
- [59] M. Schachtebeck and A. Schöbel, *Algorithm for delay management with capacity constraints*, ARRIVAL Report TR-0070, August 2007.
- [60] A. Schöbel, *Capacity constraints in delay management*, ARRIVAL Report TR-0017, CASPT 2006, Leeds, UK, June 21-23, 2006.
- [61] A. Schöbel, *Customer-Oriented Optimization in Public Transportation*, Springer, Germany, 2006.
- [62] A. Schöbel, *Customer-Oriented Optimization in Public Transport*, Habilitationsschrift, 2003.
- [63] W. Schwanhäuser, *Estimation of Finite Mixtures of Distributions from the Exponential Family*, Veröffentlichungen des Verkehrswissenschaftlichen Instituts der RWTH Aachen, 1974.
- [64] C. Schwindt, *Resource Allocation in Project Management*, Springer, Heidelberg, Germany, 2005.
- [65] R.D. Small and S.S. Schor, *Bayesian and non-Bayesian methods of inference*, *Annals of Internal Medicine*, Vol. 99, pp. 857-859, 1983.
- [66] W. A. Stahel, *Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler*, Vieweg, 2002
- [67] J.D. Storey, *A direct approach to false discovery rate*, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, Vol 64, pp. 479-498, 2002.
- [68] Y. Takeuchi and N. Tomii, *Robustness indices for train rescheduling*, 1st International Seminar on Railway Operations Modeling and Analysis, Delft, The Netherlands, 2005.

- [69] F.J. Theis and E.W. Lang, *Uniqueness of linear independent component analysis and the Skitovitch-Darmois theorem*, preprint 2003.
- [70] N. Tomii, T. Yoshiaki, T. Noriyuki, H. Chikara and M. Kunimitsu, *Train rescheduling algorithm which minimizes passengers' dissatisfaction*, IEA/AIE 2005, Bari, Italy, 2005.
- [71] J. Törnquist, *Railway traffic disturbance management*, Doctoral Thesis at Blekinge Institute of Technology, Karlskrona, Sweden, 2006.
- [72] J. Törnquist and J. A. Persson, *Train Traffic Deviation handling using Tabu Search and Simulated Annealing*, Department of System and Software Engineering, Blekinge Institute of Technology, Karlskrona, Sweden, 2005.
- [73] UIC Code 406, DB 405.0102 Anhang 3 Tab. 301.
- [74] I. Vincze, *Mathematische statistik mit industriellen Anwendungen*, Akadémiai Kiadó, Budapest, 1971.
- [75] A. Wille and P. Bühlmann, *Tri-graph: a Novel Graphical Model with Application to Genetic Regulatory Network*, ETH Zürich, Switzerland, 2004.
- [76] A. Wille and P. Bühlmann, *Low-Order Conditional Independence Graphs for Inferring Genetic Networks*, ETH Zürich, Switzerland, 2006.
- [77] S. Wegele and E. Schnieder, *Dispatching of train operations using genetic algorithms*, Technical University Braunschweig, Germany, submitted to CASPT 2004, San Diego, California, 2004.
- [78] E. Wendler and M. Naehring, *Statistische Auswertung von Verspätungsdaten*, Eisenbahn Ingenieur Kalender, pp. 321-331, 2004.
- [79] J. Woerner, *Einführung in Stokastic*, note of the course at Georg-August Universität, Göttingen, Winter Semester 2005.
- [80] J. Yuan, *Stochastic modeling of train delays and delay propagation in stations*, Doctoral Thesis at TU Delft, The Netherlands, Eburon Academic Publisher, 2006.
- [81] P. Zwaneveld, L. Kroon, H. Romeijn, M. Salomon, S. Dauzère-Pérès, S. Van Hoesel and H. Ambergen, *Routing trains through railway stations: model formulation and algorithms*, Transportation Science, Vol 30, Nr. 3, pp. 181-194, 1996.

Curriculum Vitæ

Personal Data

Name : Carla
Surname : Conte
Nationality : Italian
Domicile : Bassano del Grappa (VI), Italy

Education

Dec. 2004 **Ph.D. in Mathematics**
Jan. 2008 Operational Research
Georg-August University of Göttingen, Germany

2004 Training course in Production Planning and Logistic
FOREMA, Padova, Italy

Aug. 2002 Research assistant
Jul. 2003 ITWM, Optimization department
Kaiserslautern, Germany

Dec. 2002 **MS Mathematics**
University of Milano-Bicocca, Italy

Jul. 2001 **MS Mathematics**
University of Padova, Italy