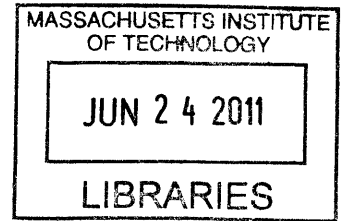


Modeling Cost and Time Uncertainty in Rail Line Construction

by

Yvonne Moret

Dipl.-Ing., Eidgenössische Technische Hochschule Zürich (2006)



ARCHIVES

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of


Doctor of Philosophy
in the field of Construction Engineering and Management

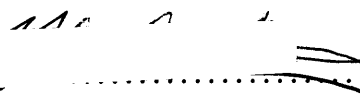
at the

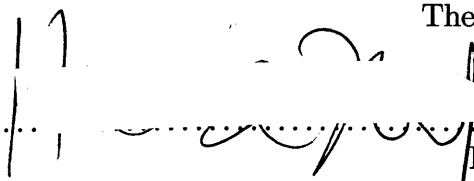
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

© 2011 Massachusetts Institute of Technology. All rights reserved.

Signature of Author 
Department of Civil and Environmental Engineering
29 April 2011

Certified by 
Herbert H. Einstein
Professor of Civil and Environmental Engineering
Thesis Supervisor

Accepted by 
Heidi M. Nepf
Chair, Department Committee for Graduate Students

Modeling Cost and Time Uncertainty in Rail Line Construction

by

Yvonne Moret

Submitted to the Department of Civil and Environmental Engineering
on 29 April 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the field of Construction Engineering and Management

Abstract

Transportation construction projects are often plagued by cost overruns and delays. Technical, economic-political, psychological, and legal causes explain the frequent underestimations. To counteract such underestimations, the author developed an innovative approach to capture cost and time uncertainty in rail line projects, and applied this to the construction of a new high speed rail line in Portugal.

The construction of the four main types of structures in rail lines (tunnels, viaducts, cuts and embankments) is modeled bottom-up from the single activity to the entire rail line. Sub-networks of activities are combined in structure networks to model the rail line structures; in turn, the structure networks are organized in the construction network to represent the rail line. For the first time, three sources of uncertainty (variability in the construction process, correlations between the costs of repeated activities, and disruptive events) are modeled jointly at the level of the single activity. These uncertainties are propagated to the total construction cost and time through the combination of the individual activity costs and times. The Construction and Uncertainty Models are integrated in the Decision Aids for Tunneling (DAT), which have been extended beyond tunneling to consider different structures and different uncertainty types. Based on historical input data and expert estimations, the cost and time uncertainty in the construction of four alignments of the new Portuguese high speed rail line is simulated. The three sources of uncertainty cause different cost and time impacts depending on the type of structure suggesting structure specific mitigation measures. Most importantly, their cumulative impact causes significant increases in construction cost and time compared to the deterministic estimates: 58% in the construction cost of tunnels, and 94% in the construction time of cuts and embankments.

The Construction and Uncertainty Models and their integrated implementation in the DAT provide transportation agencies with a modeling tool to tackle cost and time uncertainty in the construction of rail lines and other linear/networked infrastructure projects.

Thesis Supervisor: Herbert H. Einstein

Title: Professor of Civil and Environmental Engineering

Contents

1	Introduction	47
1.1	Problem Statement and Motivation	47
1.2	Models of Uncertainty for Infrastructure Construction Projects	48
1.3	Objectives, Approach and Contributions	49
1.3.1	Objectives	49
1.3.2	Approach	49
1.3.3	Contributions	50
1.4	Thesis Structure	52
2	Literature Review: Risk in the Construction of Transportation Infrastructure Projects	54
2.1	Big Dig and Swiss Transalpine Base Tunnels: Examples of Cost Underestimation	56
2.2	A Statistical Study of Cost Underestimation in the Construction of Transportation Infrastructure Projects	57
2.3	Causes and Cures of Cost Underestimation according to Flyvbjerg	62
2.3.1	Causes	62
2.3.2	Cures	65
2.3.3	A Specific Cure: Reference Class Forecasting	66
2.4	Causes and Cures according to the Transportation Community	69
2.4.1	Risk Definition	70
2.4.2	Causes: Risk Factors	73
2.4.3	Cures: Risk Strategy and Other Strategies	75
2.4.4	Cures: Risk Identification	78

2.4.5	Cures: Risk Analysis	82
2.4.6	Cures: Other Phases of the Risk Strategy	88
2.4.7	A Specific Cure: Cost Estimate Validation Process	90
2.5	Conclusions	93
3	Construction Model for Rail Lines	95
3.1	Construction Cost and Time	96
3.2	Activity Networks	97
3.2.1	Tunnels	98
3.2.2	Viaducts	99
3.2.3	Cuts and Embankments	105
3.3	Material Handling in Cuts and Embankments	107
3.4	Conclusions	115
4	Uncertainty Model	116
4.1	Uncertainty Definition	116
4.2	Sources of Uncertainty	118
4.2.1	Variability	118
4.2.2	Cost Correlations	120
4.2.3	Disruptive Events	122
4.2.4	Cumulative Impact of the Sources of Uncertainty	123
4.3	Expert Estimation	124
4.4	Effectiveness and Advantages of the Proposed Uncertainty Model	134
4.5	Conclusions	139
5	Correlations in Rail Line Construction	140
5.1	Examples of Cost Underestimation in Building Construction due to Correlation . . .	141
5.2	Correlation Measures	142
5.3	Correlation Models	152
5.4	Correlations in the Construction of a Rail Line	173
5.4.1	Identification of Correlations between Costs	174

5.4.2	Assessment of the Correlations' Impact: Viaduct Case Study (Correlation Type 2 and Type 5)	181
5.4.3	Assessment of the Correlations' Impact: Tunnel Case Study (Correlation Type 1)	199
5.5	Summary and Conclusions	208
6	Simulation Tool	210
6.1	Simulation Tools to Model Construction Projects	211
6.2	The Decision Aids for Tunneling (DAT)	216
6.3	Extension of the DAT to Model Viaducts, Cuts and Embankments	220
6.4	Implementation of the Construction Model in the DAT	223
6.5	Implementation of the Uncertainty Model in the DAT	231
6.5.1	Probability Distributions	231
6.5.2	Cost Correlations	233
6.5.3	Disruptive Events	236
6.6	Conclusions	241
7	Application of the Construction and Uncertainty Models to the New Portuguese High Speed Rail Line	242
7.1	Description of the Project	243
7.2	Input Data	244
7.2.1	Tunnel Data	244
7.2.2	Viaduct Data	257
7.2.3	Cuts and Embankments Data	264
7.3	Application of the Construction Model: Deterministic Total Cost and Total Time . .	270
7.3.1	Tunnels	270
7.3.2	Viaducts	273
7.3.3	Cuts and Embankments	276
7.4	Application of the Uncertainty Model: Total Cost and Total Time Distributions . .	279
7.4.1	Variability	279
7.4.2	Cost Correlations	290

7.4.3	Disruptive Events	293
7.5	Parallel Versus Sequential Construction	301
7.6	Value of the Construction and Uncertainty Models: Capturing the Impact of Uncertainty on Total Cost and Total Time	302
7.7	Summary and Conclusions	309
8	Conclusions and Future Research	311
8.1	Conclusions	311
8.2	Current Limitations and Future Perspectives	313
A	Questionnaire for Expert Estimation	316
B	Probability Concepts	335
B.1	Variance of the Sum of Cost Variables	335
B.2	Independent Versus Uncorrelated Variables	336
B.3	Correlation Methods generating Samples with Desired Pearson Correlation	338
B.4	Positive Definiteness	339
C	Structures of the New Portuguese High Speed Rail Line	342
C.1	List of structures	342
C.2	Tunnels	343
C.2.1	Ground Parameter States and Ground Classes	348
C.2.2	Construction Methods	348
C.3	Viaducts	348
C.3.1	Activity Networks	348
C.3.2	Construction Cost and Time	383
C.4	Cuts and Embankments	384
C.4.1	Volumes of the Cuts and Embankments	384
C.4.2	Construction Cost and Time	408
D	Uncertainty in the Construction of the New Portuguese High Speed Rail Line	416
D.1	Variability	416
D.2	Cost Correlations	420

D.3 Disruptive Events	423
D.4 Value of the Construction and Uncertainty Models	427
E Literature References	437

List of Figures

- 2-1 Cost breakdown structure of the current additional costs for the Gotthard base tunnel: while 22% of the additional costs are caused by geology and the contract awarding process, 78% of the costs are caused by additional investments in the project in the form of measures to protect the environment, new technology and higher safety standards, and delays in the construction and decision process (Teuscher 2007). . . . 58
- 2-2 Distribution of the cost escalation in transportation infrastructure projects. Underestimation (positive cost escalation) was far more common (86%) than overestimation (negative cost escalation) (14%), and the underestimation was also substantially larger (up to 280% compared with -80%). After Flyvbjerg et al. (2002). 60
- 2-3 Distribution of cost escalation for a) rail projects, b) fixed-link projects (tunnels and bridges), and c) road projects. The type of project influences the magnitude of cost escalation, with rail projects suffering from the largest percentages. After Flyvbjerg et al. (2002). 61
- 2-4 The risk breakdown structure is a tool to identify risks. In the figure four types of risk are identified in the case of a hazardous waste remediation project: project, technical, internal and external risks. Examples of technical risk can be mechanical, electrical and nuclear (after Molenaar et al., 2010). 82
- 2-5 Monte Carlo simulation: a) input cost and time distributions, and b) output cost-time scattergram and cost and time histograms. The Monte Carlo simulation generates samples from the input distributions of the cost and time parameters, and sums the generated costs and times to calculate the total cost and the total time for each simulation run. The scattergram shows 1,000 simulation runs. 84

2-6	Three-point estimates: the probability distribution is assumed triangular, where the minimum is equal to the optimistic estimate, the maximum is equal to the pessimistic estimate, and the mode is equal to the most-likely estimate.	85
2-7	A P <i>x</i> I matrix (Probability <i>x</i> Impact matrix) is a tool for qualitative risk analysis: first, the probability of occurrence of a risk is matched to a ranking number (probability ranking table); second, the impact of a risk on the project objective is also matched to a ranking number (impact ranking table); finally, according to the probability ranking number and the impact ranking number, a risk falls into the red (H-high), the yellow (M-moderate), or the green (L-low) area. Depending on this qualitative risk analysis result, the risk is warranted risk management attention. . .	86
2-8	A risk map is a visualization of the probability and impact of the risks in a project. It allows one to visually compare more than one risk.	87
3-1	a) two activities in series (top) and in parallel (bottom). b) the OR node activates the following activity as soon as any of the preceding activities is completed (top), whereas the AND node activates the following activity after both preceding activities are completed (bottom). c) in the bottom sequence, the second and the third activities depend on the first and the second activities of the top sequence, respectively. This is modeled with two dummy activities connecting the end of the activities in the top sequence to the start of the activities in the bottom sequence.	98
3-2	The construction of a tunnel is reproduced with an activity modeling the construction of one unit length of the tunnel. The cost and the time of the activity are calculated with equations 3.3 and 3.4.	99
3-3	Elements of a viaduct. Each element is modeled with one activity characterized by its cost and time.	100
3-4	Activity modeling the construction of a viaduct element. The cost and the time of the activity are calculated with equations 3.5 and 3.6.	100

3-5	a) construction of a viaduct with the balanced cantilever method: the deck section is built after the pile set, the footings and the pier are constructed. b) construction of a viaduct with the span-by-span method: the deck section is built after the preceding and following pile sets, footing and piers are constructed. c) construction of a viaduct with the launching method: similarly to the span-by-span construction method, the deck section is built (launched) after the preceding and following pile sets, footings and piers are constructed.	102
3-6	Balanced cantilever construction method: a) activity network and repeating sub-network of b) a six-span viaduct. The repeating sub-network models the construction of a unit (dashed in b) including the foundation (pile set and footing), the pier and the deck section. In the activity network the sub-network is repeated five times, the number of the units in the viaduct (the first and the fifth repetitions of the sub-network are dashed in a).	103
3-7	Span-by-span and launching construction methods: a) activity network and repeating sub-network of b) a six-span viaduct. The repeating sub-network models the construction of a unit (dashed in b) including the foundation (pile set and footing), the pier and the deck section. The sub-network is repeated five times, the number of the units in the viaduct (the first and the fifth repetitions of the sub-network are dashed in a).	104
3-8	Cross-sections of a) a railroad cut and b) a railroad embankment. They show the construction processes: clearing the soil, placing the capping layer, placing the sub-ballast, the ballast, the railroad ties and the tracks are common processes, while excavating is a process in cut construction, and improving the in situ soil and filling the embankment are processes in embankment construction.	106
3-9	a) parallel and sequential excavation of a cut. b) parallel and sequential fill of an embankment. c) lateral parallel and sequential fill of an embankment. Cuts and embankments can be constructed in several ways, of which all include parallel and sequential processes.	108
3-10	Activity modeling a process of the cut or embankment construction. The cost and the time of the activity are calculated with equations 3.7 and 3.8.	108

3-11 a) sequence of cuts and embankments between a viaduct and a tunnel, modeled with b) an activity network consisting of c) repeated sub-networks. The cut sub-network includes the clearing, the excavating and the capping activities, while the embankment sub-network includes the clearing, the improving, the filling and the capping activities. Each sub-network is repeated two times in the activity network, the number of cuts and embankments in the particular sequence (Figure a). In b) and c), thinner arrows are dummy activities. 109

3-12 The virtual material is delivered by the clearing activity of the embankment to the repository in the quantity equivalent to the filling volume (V_{fi}). It is used by the activity excavating in the quantity equivalent to the excavation volume (V_{ex}). By delivering the virtual resource, the activity clearing triggers the activity excavating. This uses the virtual resource and produces the resource exvated_material in the quantity also equivalent to the excavating volume (V_{ex}). This is used by the activity filling in the quantity equivalent to the filling volume (V_{fi}). This concept, which applies to all cuts and embankments in the alignment, controls the mass balance between cuts and embankments. 113

4-1 Lognormal and triangular probability distributions are used to model the variability in the cost and time variables, respectively. The lognormal distribution is the selected probability distribution to model cost variables since it often underlies the distribution of construction cost variables. Time variables are modeled with the triangular distribution for four reasons: it has a positive minimum, it can be skewed either to the left or to the right, it is common in construction modeling, and its parameters are relatively easily estimated by an expert. 119

4-2 Varying degrees of correlation between variables 1 and 2: from left to right, highly negatively correlated, mildly negatively correlated, uncorrelated, mildly positively correlated, and highly positively correlated. Correlations between cost variables are modeled in the proposed methodology using Spearman correlation coefficients and the correlation model NORTA (Chapter 5). 121

4-3	The cost and the time impacts of a disruptive event are modeled with triangular distributions. Since there are only very few studies on the underlying distributions of cost and time impacts of disruptive events, due to their rare occurrence, the triangular distribution is preferred for three reasons: it has a positive minimum, it can be skewed either to the left or to the right, and its parameters are relatively easily estimated by an expert.	123
4-4	Total cost and total time distributions for the construction of a tunnel a) modeling no uncertainty, b) modeling variability, c) modeling variability and cost correlations, d) modeling variability, cost correlations and one disruptive event. Compared to the single point of the deterministic calculation in a), the clouds of points in b), c) and d) visualize the impact of uncertainty on the total cost and the total time. In b) the cloud of points visually represents the variability of the total cost and total time. In c) the cost correlations cause the cloud of points to expand in the total cost direction. In d) the cloud of points explodes due to the additional cost and time of the disruptive event. The largest impact is produced by the disruptive event for which the range of possible outcomes drastically increases in both total cost and total time. Note that the graphs are shown with same axis scales in order to guarantee comparability.	125
4-5	Concept visualization for a) the probability of occurrence and not-occurrence of event A, and for b) the triangular distribution of a time variable. These two probability concepts were reviewed with the experts in the calibration.	129
5-1	1,000 samples from Gaussian distributed random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ with different Pearson correlation coefficient, ρ . The value of this can vary between -1 and $+1$	144
5-2	1,000 samples from Gaussian distributed random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ with different Spearman correlation coefficient, ρ_r	145
5-3	Ranks of 1,000 samples from Gaussian distributed random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ with different Spearman correlation coefficient, ρ_r	145

5-4 Spearman correlation coefficient, ρ_r (dashed), and Kendall's Tau, τ (dotted), as a function of Pearson correlation coefficient, ρ , for the bivariate normal distribution. For comparison the equality $\rho = \rho_r = \tau$ is shown (solid). 147

5-5 Minimum (dashed) and maximum (solid) Pearson correlation coefficients for $X \sim Lognormal(0, 1)$ and $Y \sim Lognormal(0, \sigma^2)$, $\sigma > 0$. For the given distributions, the actual range of ρ is a subset of the interval $[-1, +1]$ 148

5-6 Minimum (dashed) and maximum (solid) Pearson correlation coefficient as a function of the distribution parameter α_2 , for $\alpha_1 = 1$, for a bivariate gamma distribution. For the given distributions, the actual range of ρ is a subset of the interval $[-1, +1]$. . . 149

5-7 Given random variables $X \sim beta(5, 5)$ and $Y \sim lognormal(0, 1)$, depicted are the samples (x_i, y_i) (top left), the cumulative distributions of the samples $(F_X(x_i), F_Y(y_i))$ (top right), the ranks (r_i, s_i) of the pairs (x_i, y_i) , (bottom left), and the ranks (v_i, t_i) of the pairs $(F_X(x_i), F_Y(y_i))$ (bottom right). Differently from the Pearson correlation coefficient, the Spearman correlation coefficients of the values are the same since $\rho_r(X, Y) = \rho(R, S) = \rho(V, T) = \rho_r(F_X(X), F_Y(Y))$ 151

5-8 Simulation steps using Iman and Conover's correlation model to generate correlated variables with desired correlation matrix and desired marginal probability distributions. Clockwise from top left a) step 1: generate independent samples with beta and lognormal marginal distributions, b) step 2: generate independent samples with normal marginal distributions, c) step 3: generate correlated samples with normal marginal distributions, d) step 4: generate correlated samples with beta and lognormal distributions. 161

5-9 Simulation steps using the correlation model NORTA to generate correlated costs with desired correlation matrix and desired marginal probability distributions. a) step 1: generate correlated samples with normal marginal distributions, b) step 2: calculate the cumulative distributions of the correlated normal samples, that is generate correlated samples with uniform marginal distributions, c) step 3: calculate the inverse cumulative of the correlated uniform samples, that is generate correlated samples with desired marginal distributions. 163

5-10	Copula on the unit square $[0, 1]^2$ with uniform marginal distributions. Copulas allow one to generate samples from correlated variables with the desired Spearman correlation coefficient and with desired marginal distributions.	165
5-11	a) Frank and b) Clayton copulas with marginal uniform distributions (0, 1) and Spearman correlation $\rho_r = 0.7$. Depending on the copula, the dependence structure between variables can be different. Thus, given the marginal distributions and the correlation, the multivariate distribution is not uniquely defined.	166
5-12	Samples from uniform distributions with different Spearman correlations (left column: $\rho_r = -0.7$, center column: $\rho_r = 0$, right column: $\rho_r = 0.7$) generated with different Archimedean copulas (top row: Frank copula, middle row: Clayton copula, bottom row: Gumbel copula). Depending on the copula, the dependence structure between variables changes. The Clayton and the Gumbel copulas exist only for positive correlations.	168
5-13	Samples from uniform distributions with different Spearman correlations (left column: $\rho_r = -0.7$, center column: $\rho_r = 0$, right column: $\rho_r = 0.7$) generated with different elliptical copulas (top row: normal copula, middle row: t-copula with 2 degrees of freedom, bottom row: t-copula with 20 degrees of freedom). Depending on the copula, the dependence structure between variables changes.	169
5-14	Viaduct case study. a) the viaduct consists of several elements, whose construction is modeled with b) the activity network. The total cost is the sum of all activity costs, while the total time is the sum of the activity times on the critical path in the activity network. A dummy activity, which is used to model dependencies between activities, has no cost and no time.	182
5-15	a) example of a lognormal cost distribution (abutment's cost), whose minimum, mode, and high value (exceeded with 2% probability) are determined based on the estimated cost (Table 5.5), b) example of a time triangular distribution (abutment's time), whose minimum, mode, and maximum value are determined based on the estimated time (Table 5.5).	184
5-16	A unit (dashed) of the viaduct consists of a pile set, a footing and a pier. In scenario 2, the costs of these are correlated.	186

5-17	Viaduct case study, scenarios 1, 2 and 3. Scatterplots and frequency plots of the total cost and total time for a) scenario 1 (independent), b) scenario 2 (correlation type 1), and c) scenario 3 (correlation type 5). While correlation type 1 increases the spread of the total cost, correlation type 5 causes the correlation between total cost and total time.	188
5-18	Viaduct case study, scenario 2, sensitivity analysis. Scatterplots and frequency plots of the total cost and total time for a) 2-unit, b) 9-unit, c) 30-unit, d) 50-unit, and e) 100-unit viaduct. With increasing number of units, the spread of the total cost increases. Note that the graphs are plotted with ratio $x - axis$ to $y - axis$ of 15,000 euro to 1 day to ensure comparability.	192
5-19	Normalized difference between the total cost standard deviation for correlated pile set cost, footing cost and pier cost, and the total cost standard deviation for independent pile set cost, Δ , as a function of the viaduct unit, u . Δ increases with increasing number of units, u ; however, the rate of increase of Δ reduces with increasing number of units, u	194
5-20	Viaduct case study, scenario 3, sensitivity analysis. Scatterplots and frequency plots of the total cost and total time for a) 2-unit, b) 9-unit, c) 30-unit, d) 50-unit, and e) 100-unit viaduct. With increasing number of units, the correlation between total cost and total time increases. Note that the graphs are plotted with ratio $x - axis$ to $y - axis$ of 15,000 euro to 1 day to ensure comparability.	196
5-21	Correlation between total cost and total time, ρ , as a function of the viaduct unit, u . The correlation increases with number of units; however, for $u = 100$ it does not reach 0.8, the correlation between activity cost and activity time.	197
5-22	Correlation type 1, scenarios A, B1, B2, and B3. Scatterplots and frequency plots of the total cost and total time for a) scenario A, b) scenario B1, c) scenario B2, and d) scenario B3. When modeling correlation type 1 (scenarios B2 and B3), the spread of the total cost dramatically increases compared to the scenarios modeling independence (scenarios A and B1).	203

5-23	Correlation type 1, scenarios B2 and B3, sensitivity analysis. Normalized difference in total cost standard deviations, Δ , as a function of the tunnel length. Δ increases from scenario B2 to scenario B3. It also increases with tunnel length. However, with increasing tunnel length, the increase rate of Δ decreases faster in scenario B2 than in scenario B3.	207
6-1	Description of the geology component in the DAT. Areas, zones and ground parameters (input) determine the ground class profile (output). The ground parameters are described by their states. From the combination of ground parameter states, the ground class is determined. For example, in area 1, zone 1, the ground parameter states gneiss (lithology), poor (rock mass), and low (overburden) correspond to ground class GC5.	217
6-2	Construction simulation component in the DAT. The superposition of tunnel network and ground class profile produces the construction method profile. A construction method can be assigned to more than one ground class, for instance construction method CM-C is assigned to two ground classes, GC1 and GC2.	218
6-3	Activity network of a construction method. It consists of three activities, each characterized by cost, time and resource (used and produced) equations.	219
6-4	Time-cost scattergram. One point represents the total cost and total time for the construction of a tunnel. The simulation calculates this by proceeding through one construction method profile. The other points are generated by simulating other construction method profiles.	220

6-5 Area, zones, ground parameters, and ground classes in the extended framework of the DAT. The area is alignment 1, and the zones are the structures, e.g. embankment 1, viaduct 1, cut 1, and tunnel 1. While for tunnels the ground classes are geology descriptors, for viaducts, cuts and embankments, ground classes describe a specific characteristic of the structure. For viaducts, the ground class identifies the viaduct: ground class "Moinhos" is the ground class for viaduct 1. For cuts, the ground class indicates the percentage of material excavated by blasting: cut 1 has ground class GCB25 (ground class blasting 25%) and cut 2 has ground class GCB50 (ground class blasting 50%). For embankments, the ground class indicates the need of improving the ground under the embankment: this is improved in embankment 1 (ground class CMI), while it is not improved in embankment 2 (ground class CMNI). 222

6-6 The construction network depicts the structures. These are built with one or more construction methods. The activity network of each construction method determines the order of the activities to be performed. 224

6-7 a) activity network modeling the construction of a unit length of tunnel at the general level, and b) activity network modeling the construction of a unit length of tunnel at the detailed level. The DAT can model the construction of a structure at different levels of detail. 225

6-8 a) a viaduct consists of several elements, whose construction is modeled with b) the activity network. The more spans the viaduct has, the longer the activity network becomes. 227

6-9 a) the uniform distribution, b) the triangular distribution, and c) the triangular distribution with spikes can be modeled in the DAT. 232

6-10 The DAT generate samples from a lognormal distribution (μ, σ) . μ and σ are the parameters of the corresponding normal distribution. Given the minimum, the mode and a percentile of the distribution, they are determined with expression 6.1. 233

6-11	A Markov process models the probability of occurrence of a disruptive event in the construction of a structure. The state "no" indicates that the disruptive event does not occur, while the state "yes" indicates that the disruptive event occurs. The probability of transitioning to the "yes" state is equal to the probability of occurrence of the disruptive event, p . The probability of transitioning to the "no" state is equal to $(1 - p)$	237
6-12	In the modified Markov process, the transition to the same state is replaced with the transition to an equivalent state. For instance, the transition from the "no" state to itself is replaced with the transition from the "no1" state to the "no2" state, or vice versa. Similarly, the transition from the "yes" state to itself is replaced with the transition from the "yes1" state to the "yes2" state, or vice versa.	238
7-1	New Portuguese high speed rail. The three main axes (blue) are Lisbon-Madrid (Spanish border), Lisbon-Porto, and Porto-Vigo. Four alignments of the Porto-Vigo (Spanish border) axis are investigated in this chapter.	243
7-2	New Portuguese high speed rail line, Porto-Vigo axis, Porto-Braga alignments. Alignment A (blue) and alignment C (B (red) - BA (green) - A (blue)) are two of the four alignments studied in this chapter.	246
7-3	New Portuguese high speed rail line, Porto-Vigo axis, Braga-Vigo (Spanish border) alignments. Alignment A (red) and alignment B (blue) are two of the four alignments studied in this chapter.	247
7-4	The probability density function of the lognormal distribution and the probability mass function of the discrete distribution are plotted (left) for the cost of tunnels in a) good, b) medium, and c) poor geology. The cumulative distribution functions of the lognormal distribution and of the discrete distribution are plotted (right) for the cost of tunnels in a) good, b) medium, and c) poor geology. Lognormal and discrete distributions have been fitted by matching the mean and the standard deviation. The lognormal distributions are the input for the uncertainty model.	252

7-5 The probability density function of the triangular distribution and the probability mass function of the discrete distribution are plotted (left) for the time per unit length of tunnels in a) good, b) medium, and c) poor geology. The cumulative distribution functions of the triangular distribution and of the discrete distribution are plotted (right) for the time of tunnels in a) good, b) medium, and c) poor geology. Triangular and discrete distributions have been fitted by matching the mean, the standard deviation, and a point of the cumulative distribution in a) and b), and the mode of the probability mass/density function in c). The triangular distributions are the input for the uncertainty model. 254

7-6 Probability distribution of the cost factor. It is multiplied with the deterministic input costs of viaducts to obtain the probability distributions of the costs of viaducts. 260

7-7 Deterministic input time, time distribution estimated by the expert and calculated time distribution for the construction of a viaduct. The calculated time distribution is obtained by moving the time distribution estimated by the expert along the time axis and matching the mode of the estimated time distribution with the deterministic input time. It is the input for the uncertainty model. 263

7-8 Probability distributions of the input cost per unit length (left) and the time per unit length (right) of tunnels in a) good, b) medium, and c) poor geology. Note that the mean input cost per unit length and the mean input time per unit length are equal to the deterministic input cost per unit length and the deterministic input time per unit length (Table 7.41). 281

7-9 Probability distribution of the input cost factor of viaducts. Note that the mean input cost factor (1.04) is larger than the deterministic input cost factor of 1 (Table 7.42). 282

7-10 Probability distributions of input times of viaducts. Note that the mean input time is larger than the deterministic input time (Table 7.42). 283

7-11 Probability distributions of the input costs per unit volume of cuts and embankments. Note that the mean input cost per unit volume is larger than the deterministic input cost per unit volume (Table 7.43). 284

7-12	Probability distributions of the input production rates of cuts and embankments. Note that the mean input production rate is smaller than or equal to the deterministic input production rate (Table 7.43).	285
7-13	Alignment A-a, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. The scattergrams make evident the uncertainty in the construction cost and time caused by variability of input parameters. The deterministic total cost and total time (black dot) of the tunnels (a) is located in the center of the scattergram since the deterministic input cost and time are equal to the mean input cost and time. The deterministic total cost and total time of the viaducts (b) and of the cuts and embankments (c) lie below the respective scattergrams since the deterministic input cost and time are equal to the mode input cost and time. The deterministic total cost of all structures (d) is smaller than the mean total cost since it is equal to the sum of the deterministic total costs of tunnels, viaducts, cuts and embankments, while the deterministic total time of all structures is equal to the mean total time of the tunnels since these determine the critical path.	289
7-14	Alignment A-a, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. Due to cost correlations, the range of the total cost increases dramatically.	292
7-15	Disruptive events in tunnel construction. Cave-in: probability of occurrence and probability distributions of a) the cost impact and b) the time impact. Water inflow: probability of occurrence and probability distributions of c) the cost impact and d) the time impact.	294
7-16	Disruptive events in viaduct construction. Differing ground conditions: probability of occurrence and probability distributions of a) the cost impact and b) the time impact. Construction accident/problem: probability of occurrence and probability distributions of c) the cost impact and d) the time impact.	295

7-17 Disruptive events in earthwork (cuts and embankments) construction. Flooding: probability of occurrence and probability distributions of a) the cost impact and b) the time impact. Differing site conditions: probability of occurrence and probability distributions of c) the cost impact and d) the time impact. 296

7-18 Alignment A-a, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. Due to the disruptive events, the ranges of total cost and total time increase significantly. 299

7-19 Alignment A-a: comparison of total cost and total time for the construction in sequence (left) and in parallel (right) of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. While the total costs remain the same, the total times decrease significantly due to the shortened critical paths. 303

7-20 Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 306

7-21	Increases (from deterministic total cost/time to the 90 th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.	307
B-1	Uncorrelated but not independent variables X and Y . Independent variables are uncorrelated, whereas uncorrelated variables are not necessarily independent.	337
C-1	Activity network of viaduct Moinhos in alignment A-a.	352
C-2	Activity network of viaduct Aldeia in alignment A-a.	353
C-3	Activity network of viaduct Ave in alignment A-a. It is the only viaduct constructed from both ends due to its extended length.	354
C-4	Activity network of viaduct Este in alignment A-a.	355
C-5	Activity network of viaduct Cambeses in alignment A-a.	356
C-6	Activity network of viaduct Laje in alignment A-c.	357
C-7	Activity network of viaduct Ave in alignment A-c.	358
C-8	Activity network of viaduct Este in alignment A-c.	359
C-9	Activity network of viaduct Cambeses in alignment A-c.	360
C-10	Activity network of viaduct Mau in alignment B-a.	361
C-11	Activity network of viaduct Cavado in alignment B-a.	362
C-12	Activity network of viaduct Neiva in alignment B-a.	363
C-13	Activity network of viaduct Outeiro in alignment B-a.	364
C-14	Activity network of viaduct Lima in alignment B-a.	365
C-15	Activity network of viaduct Labruja in alignment B-a.	366
C-16	Activity network of viaduct Coura in alignment B-a.	367

C-17 Activity network of viaduct Reguengo in alignment B-a.	368
C-18 Activity network of viaduct Boriz in alignment B-a.	369
C-19 Activity network of viaduct Mau in alignment B-b.	370
C-20 Activity network of viaduct Cavado in alignment B-b.	371
C-21 Activity network of viaduct Leiras in alignment B-b.	372
C-22 Activity network of viaduct Neiva in alignment B-b.	373
C-23 Activity network of viaduct Pombarinhos in alignment B-b.	374
C-24 Activity network of viaduct Lima in alignment B-b.	375
C-25 Activity network of viaduct Moinho Velho in alignment B-b.	376
C-26 Activity network of viaduct Cabracao in alignment B-b.	377
C-27 Activity network of viaduct Bouca in alignment B-b.	378
C-28 Activity network of viaduct Formigoso in alignment B-b.	379
C-29 Activity network of viaduct Coura in alignment B-b.	380
C-30 Activity network of viaduct Bogalheiro in alignment B-b.	381
C-31 Activity network of viaduct Boriz in alignment B-b.	382
C-32 Construction of cuts and embankments. a) cuts: first the top layer is cleared, then the cut is excavated, finally the capping is placed followed by the sub-ballast. b) embankments: the top layer is cleared, if needed the ground under the embankment is improved, then the embankment is filled, finally the capping is placed followed by the sub-ballast. The cost and time of these processes depend on the volumes to be removed (clearing and excavating), to be replaced (improving), and to be added (filling, capping, and sub-ballast).	391
C-33 a) the cut depth, D , and b) the embankment height, H , are weighted averages of the depths, respectively the heights, at the sides and in the center (equations C.1 and C.2).	392
C-34 Six geometries of cuts. They correspond to a slope steepness α varying between 0 for the vertical wall (CS6) and $\frac{15}{8} \simeq 1.9$ (CS2).	395
C-35 Two geometries of embankments. They correspond to a slope steepness α varying between 2.0 (ES1) and $\frac{11}{5} = 2.2$ (ES2).	396

C-36	Construction network of a sequence of cuts and embankments. The total construction time is equal to the sum of clearing, excavating, improving, filling, capping, and sub-ballast times on the critical path of the construction network.	408
C-37	Construction network of cuts and embankments in alignment A-a.	409
C-38	Construction network of cuts and embankments in alignment A-c.	410
C-39	Construction network of cuts and embankments in alignment B-a.	411
C-40	Construction network of cuts and embankments in alignment B-b.	412
C-41	Material handling. The virtual resource (Figure C-41), $startc$, is produced by the activity clearing, only in embankments (clearing of cuts does not produce virtual material), in the quantity equivalent to the improving volume (V_{im}), and by the activity improving in the quantity equivalent to the filling volume (V_{fi}). It is used by the activity excavating in the quantity equivalent to the excavation volume (V_{ex}). By delivering the virtual resource $startc$, the activities clearing and improving trigger the activity excavating. This uses the virtual resource $startc$ in the quantity equivalent to the excavating volume (V_{ex}) and produces the resource $soilrock$ in the quantity also equivalent to the excavating volume (V_{ex}). This is used by the activities improving and filling in the quantity equivalent to the improving volume (V_{im}) and filling volume (V_{fi}), respectively.	414
D-1	Alignment A-c, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.	417
D-2	Alignment B-a, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.	418
D-3	Alignment B-b, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.	419
D-4	Alignment A-c, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.	420

D-5 Alignment B-a, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. 421

D-6 Alignment B-b, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. 422

D-7 Alignment A-c, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. 424

D-8 Alignment B-a, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. 425

D-9 Alignment B-b, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. 426

D-10 Alignment A-c. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 429

D-11 Alignment A-c. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 430

D-12 Alignment B-a. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 432

D-13 Alignment B-a. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 433

D-14 Alignment B-b. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 435

D-15 Alignment B-b. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events. 436

List of Tables

- 2.1 Structure of the chapter. Causes and cures of cost underestimation according to Flyvbjerg are presented in section 2.3, while risk definitions, causes and cures according to the transportation community are discussed in section 2.4. 55
- 2.2 Cost escalation means and standard deviations for all projects as well as for the subgroups of rail, fixed link, and road projects, respectively. Rail projects had the largest cost escalation, fixed link also lay above average, while the road projects seemed to be the best contained in terms of cost escalation (after Flyvbjerg et al. 2002). 60
- 2.3 Causes and cures for the cost underestimation in infrastructure construction projects according to Flyvbjerg. There are three types of causes: technical, economic-political, and psychological. Cost underestimation can be limited with a policy change, improved forecasting techniques, such as reference class forecasting, and debiasing techniques. 63
- 2.4 Causes and cures for cost underestimation in infrastructure construction projects according to the transportation community. Causes are subdivided into 11 external causes and 7 internal causes. Cures are eight strategies. One of these is the risk strategy, which consists of five phases: identification, analysis, mitigation and planning, allocation, and monitoring and controlling. 70
- 2.5 Tools used in the identification phase and the analysis phase of the risk strategy. Expert interviews, risk management plan, risk workshops, and risk breakdown structure are used in both phases. 78

3.1	The construction and the material handling of cuts and embankments are modeled with four cut activities, five embankment activities, one depositing or importing activity, three resources and two repositories.	115
4.1	Structures (tunnels, viaducts, cuts, and embankments) and uncertainties (variability, cost correlations, and disruptive events) in the construction of rail lines. In this research work, the time variability of viaduct-, cut- and embankment construction as well as the cost correlations and the disruptive events of all structures were not available from historical sources and they had therefore to be estimated by experts. .	126
5.1	Comparison of sample and simulation standard deviations of construction costs (Touran, 1993). Assuming independence in the simulation caused the underestimation of the construction cost standard deviation.	141
5.2	Comparison of sample and simulation standard deviations of construction costs (Newton, 1992). Assuming independence in the simulation caused the underestimation of the construction cost standard deviation.	142
5.3	The correlation models are compared against the requirements. Although sampling from a joint distribution satisfies all the requirements, it is not selected since it is either inefficient or does not converge. The NORTA method, which is based on the Gaussian copula is the selected correlation model.	172
5.4	Summary of types of correlation and structures considered in this chapter.	180
5.5	Viaduct case study. Estimates of the cost and time for the viaduct activities (after RAVE (2006b)). The estimates are the input for the cost and time of activities in the activity network.	183
5.6	Viaduct case study, scenarios 1, 2 and 3. Means, standard deviations of total cost and total time, and correlations between total cost and total time in the scenarios 1, 2 and 3. The standard deviation increases by 8.3% in scenario 2, while the total cost - total time correlation is equal to 0.38.	188

5.7	Viaduct case study, scenario 2, sensitivity analysis. Means and standard deviations of the total cost assuming independence, means and standard deviations of total cost and total time assuming correlation and normalized differences in total cost standard deviation for different numbers of units. With increasing number of units, the normalized difference in total cost standard deviation increases to a maximum of 15.5%.	191
5.8	Viaduct case study, scenario 3, sensitivity analysis. Means and standard deviations of total cost and total time, and correlations between total cost and total time for different numbers of units. With increasing number of units, the correlation between total cost and total time increases to a maximum of 0.56.	195
5.9	Viaduct case study, scenarios 1 and 3. Means, 80 th and 95 th percentiles of total cost and total time, and correlations between total cost and total time for scenario 1 and scenario 3. The correlation between total cost and total time does not impact the marginal distributions of the total cost and the total time.	197
5.10	Cost and advance rate of tunnel excavation depending on the geological conditions. They are used to calculate the deterministic total cost and total time of the tunnel construction.	199
5.11	Deterministic total cost and total time of tunnel construction. They will be compared with the the probabilistic analyses of total cost and total time.	200
5.12	Correlation type 1, scenarios A, B1, B2, and B3. Means and standard deviations of the total cost and total time, and normalized differences in total cost standard deviation for scenario A and scenarios B1 to B3. When modeling correlation type 1, the standard deviation increases dramatically: 757% in scenario B2 and 1,299% in scenario B3.	202
5.13	Correlation type 1, scenarios B1, B2, and B3, sensitivity analysis. Total cost standard deviations and normalized differences in standard deviations for different number of units. With increasing number of units, the standard deviation increases dramatically: up to 3,243% in scenario B3.	206

6.1	Available simulation tools compared to the requirements set by the construction and the uncertainty models. DAT, @Risk for Project and Pertmaster satisfy four out of five requirements: in their current form, the DAT cannot model cost correlations, while @Risk for Project and Pertmaster cannot model the uncertainties derived from geology. The DAT are expanded to model cost correlations as part of the research work presented in this thesis.	215
6.2	Example of ground parameters, ground parameter states, and ground classes of tunnels, viaducts, cuts and embankments. While for tunnels ground parameters, ground parameter states, and ground classes are geologic descriptors, for viaducts, cuts and embankments they are descriptors of a specific characteristic of the structure. In viaducts, they uniquely identify the particular viaduct; in cuts, they describe the percentage of blasting excavation; in embankments, they describe the need of improving prior to filling.	221
6.3	Ground parameters, ground parameter states, and ground classes describe a specific characteristic of a structure. In tunnels, they describe the geology along a tunnel. In viaducts, they uniquely identify the particular viaduct. In cuts, they indicate the percentage of ground excavated with blasting. In embankments, they indicate whether the ground under an embankment needs to be improved prior to filling.	223
6.4	Construction network and activity network of a tunnel. Tunnels are modeled with the simplest construction network (one arc) and the simplest activity network (one activity).	228
6.5	Construction network and activity network of a viaduct. Viaducts are modeled with the simplest construction network (one arc) but a complex activity network.	229
6.6	Construction network and activity network of a sequence of cuts and embankments. This is modeled with a complex construction network but simple activity networks (one activity).	230
6.7	Inputs required in form of probability distributions for the research presented in this thesis. The cost of activities is modeled with lognormal distributions, while the time of activities as well as the cost and time impacts of disruptive events are modeled with triangular distributions.	231

6.8	The transition matrix of the Markov process determines the occurrence (state "yes") or no occurrence (state "no") of the disruptive event. The probability of transitioning to the state "yes" is equal to p , while the probability of transitioning to the state "no" is equal to $(1 - p)$	237
6.9	The transition matrix of the Markov process determines the occurrence (states "yes1" or "yes2") or no occurrence (states "no1" or "no2") of the disruptive event. The probability of transitioning to the states "yes1" and "yes2" is equal to p , while the probability of transitioning to the states "no1" and "no2" is equal to $(1 - p)$	238
6.10	Disruptive events modeled in the construction of tunnels, viaducts, cuts and embankments. Their probability of occurrence is modeled with the modified Markov process.	239
6.11	The disruptive events water inflow and cave-in may occur in every construction cycle of a tunnel. The disruptive events construction accidents (in viaducts) and differing site conditions (in viaducts, and in cuts and embankments) may occur once for the construction of a structure. The disruptive event flooding (in cuts and embankments) may occur in each construction day of the cuts and embankments. Depending on the type of occurrence, the disruptive event is modeled differently in the DAT.	239
6.12	A disruptive event that may occur once is modeled with a fictitious structure that is one-unit long. If the disruptive event occurs, the fictitious structure is built causing additional cost and time equal to the cost and time of the disruptive event.	240
6.13	A disruptive event that may occur daily for N days is modeled with a fictitious structure that is N -unit long. If the disruptive event occurs on a day, the corresponding unit of the fictitious structure is built causing additional cost and time equal to the cost and time of the disruptive event.	241
7.1	The four alignments of the Porto-Vigo axis are assigned unique names, used throughout the chapter.	244
7.2	The input data consists of historical data and expert estimations.	245
7.3	Costs per unit length, advance rates and times per unit length in tunnel construction depending on the construction conditions.	248

7.4 Mean costs per unit length, mean advance rates and mean times per unit length for tunnels in good/medium/poor geology, calculated based on the percentages of tunnel constructed in good/medium/poor construction conditions. The cost per unit length, the advance rate and the time per unit length for tunnels built by cut & cover are considered to be independent of geology. 249

7.5 Deterministic input costs and times per unit length of mined tunnels in good/medium/poor geology and of tunnels built in cut & cover. 249

7.6 Distributions of the costs per unit length for tunnels in good/medium/poor geology. The lognormal distributions are fitted to the discrete distributions by matching the means and the standard deviations. 250

7.7 1-percentile of the cost per unit length distributions in good/medium/poor geology. Assuming the minimum of the lognormal distribution is equal to zero is reasonable since the 1-percentile is a relatively large cost per unit length. 251

7.8 Distributions of the times per unit length for tunnels in good/medium/poor geology. The triangular distributions are fitted to the discrete distributions by matching the means, the standard deviations, and one point of the distributions. 253

7.9 Minimum, mode, and 98th percentile of the lognormal distributions of the cost per unit length and minimum, mode, and maximum of the triangular distributions of the time per unit length for a tunnel in good/medium/poor geology. They are the input for the uncertainty model. 255

7.10 Correlations between the costs of the repeated tunnel excavation activity. They are the input for the uncertainty model. 256

7.11 Probability of occurrence and triangular distributions of the cost and time impacts of disruptive events in tunnel construction. They are the input for the uncertainty model. 257

7.12 Geometric parameters of the viaducts. The construction costs of viaducts are calculated based on these geometric parameters. 258

7.13	The deterministic input costs are calculated with the expressions in the table. The sources (Table 7.2) assume that activities pre-work (construction site setup) and pre-deck (deck construction setup) require time but do not cause additional cost. The activity construction method models the cost of employing staging or a gantry in the construction of the deck sections.	259
7.14	Minimum-to-mode and maximum-to-mode ratios for different span lengths and pier heights. The average minimum-to-mode and maximum-to-mode ratios are used to obtain the lognormal distribution of the viaduct cost factor.	260
7.15	The probability distribution of the cost factor is the input for the uncertainty model.	260
7.16	Deterministic input times of viaduct activities. The sources (Table 7.2) assume a pier, a footing and a pile set are constructed in parallel with a deck section, and they assume their construction times equal to zero, since the deck section requires more time to be constructed. The activity construction method models the cost of employing staging or a gantry in the construction of the deck sections and does not require additional time.	261
7.17	Minimum, mode, and maximum viaduct times estimated by the expert, and deterministic input times. The input time distributions are determined by matching the mode of the estimated time distribution with the deterministic input time (Figure 7-7).	262
7.18	Minimum, mode and maximum of the triangular distributions of the viaduct times. They are the input for the uncertainty model.	263
7.19	Probability of occurrence and triangular distributions of the cost and time impacts of disruptive events in viaduct construction. They are the input for the uncertainty model.	264
7.20	Deterministic input costs per unit volume of activities in the construction of cuts and embankments.	265
7.21	Minimum, mode, and maximum costs per unit volume of earthwork (cuts and embankments) construction, and deterministic input costs. The input cost distributions are determined by matching the mode of the cost distribution with the deterministic input cost (page 261).	266

7.22	Minimum, mode and 98 th percentile of the lognormal distributions of the cost per unit volume for the construction of cuts and embankments. They are the input for the uncertainty model.	266
7.23	Minimum, mode, and maximum production rates in earthwork (cuts and embankments) construction, as estimated by the earthwork expert.	267
7.24	Minimum, mode, and maximum production rates in earthwork (cuts and embankments) construction. They are the input for the uncertainty model.	268
7.25	Deterministic input production rates of activities in the construction of cuts and embankments.	268
7.26	Correlations between the costs of excavating and filling activities in earthwork (cuts and embankments) construction. They are the input for the uncertainty model. . . .	269
7.27	Probability of occurrence and probability distributions of the cost and time impacts of disruptive events in earthwork (cuts and embankments) construction. They are the input for the uncertainty model.	270
7.28	Tunnel ground parameters GP1 to GP8 with their states. Ground classes are determined based on the combination of ground parameter states.	271
7.29	Definition of the four tunnel ground classes: good, medium, poor and reduced overburden. Different combinations of ground parameters states can correspond to the same ground class. For instance, the ground class medium is defined with three combinations of ground parameter states, which are listed in the second, third and fourth row of the table. An asterisk means that all states of the ground parameter fit in the ground class definition.	272
7.30	Relationship between tunnel ground classes and tunnel construction methods. . . .	272
7.31	Cost per unit length and time per unit length of the four tunnel construction methods. They are used to calculate the construction cost and time of the tunnels along the four alignments.	273
7.32	Deterministic construction cost and time of the tunnels in the four alignments. . . .	273
7.33	Deterministic total cost and total time for the construction of all the tunnels in the four alignments.	274

7.34	Zones, ground parameter states, ground classes, and construction methods of the five viaducts in alignment A-a. A different construction method is assigned to each viaduct in the four alignments.	274
7.35	Deterministic construction cost and time of the viaducts in the four alignments. They are calculated with the cost and time expressions in Table 7.13 and 7.16. Cambeses is the same viaduct in alignments A-a and A-c, as are Mau and Coura in alignments B-a and B-b. On the other hand, Ave and Este in alignments A-a and A-c and Cavado, Neiva, Lima and Boriz in alignments B-a and B-b are different viaducts. . .	275
7.36	Deterministic total cost and total time for the construction of all the viaducts in the four alignments.	275
7.37	Ground parameters and their states, and ground classes modeling the construction of cuts and embankments. The ground parameter blasting determines the percentage of cut excavated with blasting, while the ground parameter improving determines the need of improving the ground under the embankment. Six ground parameter states correspond to six ground classes.	276
7.38	Deterministic input costs per unit volume and production rates to calculate the deterministic total cost and total time of cuts and embankments construction. . . .	277
7.39	Total volumes (clearing, mechanical excavating, blasting, improving, filling, capping, sub-ballast) of the four alignments. With the costs per unit volume and the production rates (Table 7.38), they determine the construction cost and time of the cuts and embankments in the alignments.	277
7.40	Deterministic total cost and total time for the construction of all cuts and embankments in the four alignments.	278
7.41	Deterministic input cost and time per unit length for the construction of tunnels in good/medium/poor geology and of tunnels built by cut & cover. Note that the deterministic input costs per unit length and the deterministic input times per unit length are equal to the mean input costs per unit length and the mean input times per unit length in Figure 7-8.	280

7.42	Expressions to calculate the deterministic input cost and time of viaducts. Note that the deterministic input times are smaller than the mean input times in Figure 7-10. Since the deterministic input cost factor is equal to 1, this is smaller than the mean cost factor (Figure 7-9).	280
7.43	Deterministic input costs per unit volume and deterministic input production rates for a sequence of cuts and embankments. Note that the deterministic input costs per unit volume are smaller than the mean input costs per unit volume in Figure 7-11, while the deterministic input production rates are larger than or equal to the mean input production rates in Figure 7-12.	282
7.44	Alignment A-a. Deterministic total cost and total time and means, 90 th percentiles, and standard deviations of the total cost and total time distributions for the construction of the tunnels, the viaducts, the cuts and embankments, and of all the structures. The variability in the construction process is responsible for the range of possible total cost and total time.	288
7.45	Correlations between the costs of repeated activities.	290
7.46	Alignment A-a, uncertainty sources: variability versus variability and cost correlations. Means, 90 th percentiles, and standard deviations of the total cost and total time for the construction of the tunnels, the viaducts, the cuts and embankments, and of all the structures. Due to cost correlations, the standard deviations of the total cost increase by an order of magnitude (in viaduct construction, no correlation is modeled).	293
7.47	Alignment A-a, uncertainty sources: variability and cost correlations versus variability, cost correlations and disruptive events. Means, 90 th percentiles, and standard deviations of the total cost and total time for the construction of the tunnels, the viaducts, the cuts and embankments, and of all the structures. Due to the disruptive events, the means, the 90 th percentiles and the standard deviations of total cost and total time increase significantly.	300

7.48	Deterministic total cost and total time, 90 th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) in alignment A-a depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 58% in tunnel construction, while the largest total time increase is 94% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90 th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations.	305
A.1	Structures (tunnels, viaducts, cuts, and embankments) and uncertainties (variability, cost correlations, and disruptive events) in the construction of rail lines. In this research work, the time variability of viaduct-, cut- and embankment construction as well as the cost correlations and the disruptive events of all structures were not available from historical data and they had therefore to be estimated by experts. . .	316
C.1	List of structures in alignment A-a. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure at the beginning of the alignment.	344
C.2	List of structures in alignment A-c. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure at the beginning of the alignments.	345
C.3	List of structures in alignment B-a. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure at the beginning of the alignment.	346
C.4	List of structures in alignment B-b. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure.	347

C.5	List of tunnels in the four alignments with the ground parameters states and the ground classes. GP1 is ground parameter overburden, GP2 granit, GP3 meta-sediment, GP4 perpendicular faults, GP5 oblique faults, GP6 acute-angled faults, GP7 parallel faults, and GP8 transversal faults.	349
C.6	Construction method, tunnel length, construction cost and time of all tunnels in the four alignments.	350
C.7	Notation of the activities in the viaduct activity networks.	351
C.8	Construction cost and time of the viaducts in the four alignments. They are calculated with cost expressions (Table 7.13) and time expressions (Table 7.16) based on geometric parameters. <i>Delta</i> is the cost adjustment (Table C.10). Cambeses is the same viaduct in alignments A-a and A-c, similarly to Mau and Coura in alignments B-a and B-b. On the other hand, Ave and Este in alignments A-a and A-c, and Cavado, Neiva, Lima and Boriz in alignments B-a and B-b are different viaducts. . .	385
C.9	The calculated construction costs of the viaducts and the construction costs in the sources are compared.	386
C.10	Viaduct cost adjustments. Due to special structures in the viaducts Ave, Lima, Neiva, and Formigoso, cost adjustments are introduced to better match the calculated costs to the costs in the sources.	387
C.11	The geometric parameters of cuts and embankments determine the volumes of these.	387
C.12	Expressions to calculated the cut volume based on geometric parameters (Table C.11). a) clearing volume, b) excavating volume, c) capping volume, and d) sub-ballast volume.	388
C.13	Expressions to calculate the embankment volume based on geometric parameters (Table C.11). a) clearing volume, b) improving volume, c) filling volume, d) capping volume, and e) sub-ballast volume.	389
C.14	For consistency, the percentage of cut volume excavated with blasting, $e[\%]$ from RAVE (2008b) are used for all cuts and embankments.	394

C.15	The volumes calculated with the expressions in Tables C.12 and C.13 are compared with the volumes from the sources: for alignments A-a and A-c (RAVE 2006b) and for alignments B-a and B-b (RAVE 2008b). The ratios between calculated volumes and volumes from the sources show a good agreement in the case of clearing, excavating, and filling. Conversely, improving, capping and sub-ballast volumes require more scrutiny.	397
C.16	Given, the volumes, the widths and the thicknesses of the capping and sub-ballast layers (RAVE 2006b, RAVE 2008b), the total length of cuts and embankments is back calculated. In three alignments, it is larger than the alignment length. Clearly, the data in the sources are contradictory.	399
C.17	Alignment A-a: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.	400
C.18	Alignment A-c: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.	401
C.19	Alignment B-a: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.	402
C.20	Alignment B-b: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.	403

C.21 Alignment A-a: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment. 404

C.22 Alignment A-c: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment. 405

C.23 Alignment B-a: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment. 406

C.24 Alignment B-b: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment. 407

C.25 Material excavated in cuts that is not reused in the construction of embankments for the four alignments. The material is transported from the virtual repository to the final destination. 415

- D.1 Alignment A-c. Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 56% in tunnel construction, while the largest total time increase is 90% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations. 428
- D.2 Alignment B-a. Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 48% in tunnel construction, while the largest total time increase is 90% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations. 431

D.3 Alignment B-b. Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 52% in tunnel construction, while the largest total time increase is 90% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations. 434

Acknowledgements

Lieber Doktorvater, ich möchte mich herzlich bei Ihnen bedanken für Ihre Unterstützung von Anfang bis Ende meiner Dissertation, für die Förderung meines freien Denkens und für die Gelegenheit, in den letzten vier Jahren in Ihrer Forschungsgruppe mitzuwirken. Es war eine akademische sowie eine persönliche Reise, die ich immer in guter Erinnerung behalten werde. Danke, Professor Einstein!

My doctoral studies at MIT were possible thanks to the support of many people. I would like to express my gratitude to Prof. Roy Welsch, Prof. Sarah Slaughter and Prof. Richard de Neufville, members of the Doctoral Committee, for the invaluable insights, advice and support throughout my dissertation. I am very thankful to the providers of the financial support that made the research presented in this thesis possible: the Schoettler Fellowship, the Louis Berger Fellowship, the Janggen-Pohn Stiftung Fellowship, and in particular the MIT Portugal Program. I am sincerely thankful to the many people who contributed to my doctorate with their expertise and the project data: Dr. Claude Indermitte, Massimo Marino, Lisa O'Donnel, Prof. Luis Sousa, Ricardo Leite, Pedro Pereira, and Prof. Campos e Matos.

At MIT I have met numerous special people. I would like to thank Kris for stirring with me through and out of difficult times, Ruth for teaching me the nuts and bolts of negotiation, and Mary for her support and her warmth. In the Department of Civil and Environmental Engineering, I would like to thank headquarters for their constant cheerfulness and my research colleagues, in particular Rita and Sherif, whom I will always remember fondly. I would like to remember three special friends: Alessia for dragging me out of the office when it was time for a break; Leila for being my first friend and for the reciprocal self-irony; Stephanie for the glorious weekends in New York and for my first Thanksgiving. To the three of you: thank you for making my time at MIT unforgettable.

A special thank goes to the Swiss crowd in Boston, especially to swissnex and SwissLinkBoston. I would like to thank the swissnex team for initiating me into science diplomacy, for the fruitful interactions and for connecting so well my dots. I would like to dedicate a special thought to Pascal, Emil, Andreas, Thomas, Jacqueline, Nina, Kati and Pierre: it was a pleasure working and interacting with you. I would like to thank the initiators of SwissLinkBoston Daniela, Steffi, and Karl for making me feel at home, for the Swissness made in Boston and for the fun events. It has been a gorgeous journey together!

Provo la più profonda gratitudine per la mia famiglia. Non potrò mai ringraziare abbastanza mio padre per avermi trasmesso la sua ambizione, mia madre per il suo sostegno incondizionato e mio fratello per capirmi meglio di quanto io stessa sia mai stata capace.

Terminar esta tesis de doctorado con mens sana in corpore sano no habría sido posible nunca sin Alberto: ¡mi amor, gracias!

Chapter 1

Introduction

1.1 Problem Statement and Motivation

Infrastructure construction projects are often plagued by cost overruns and time delays. Two notorious examples are the Big Dig in Boston and the Transalpine tunnels in Switzerland. The Big Dig in Boston was estimated to cost 6 billion US dollars and to be completed in 2001. Unfortunately, the project cost almost tripled to a total of nearly 15 billion US dollars and was completed in 2007, six years after initially forecast (Salvucci 2003). The Transalpine tunnels discussed here, namely the Lötschberg and the Gotthard base tunnels, connect Italy to Central Europe. The Lötschberg base tunnel cost 4,365 million Swiss francs instead of the estimated 3,214 million. It was completed according to schedule. The Gotthard base tunnel is still under construction. However, it is known that the initially estimated cost was 7,716 million Swiss francs and the current additional cost amounts to 2,833 Swiss francs (Teuscher 2007). Initially, the Gotthard base tunnel was planned to be completed by 2012/2013; however, currently it is expected to be completed by 2017. The Big Dig and the Transalpine tunnels are only two examples of a long list of infrastructure projects plagued by cost overruns and delays.

Flyvbjerg et al. (2002) conducted a statistical study of the construction costs of 258 infrastructure projects. First, they found that cost underestimation (final construction cost is larger than initially estimated cost) was far more common than cost overestimation. More specifically: On average a project cost was underestimated by 27.6%. The probability of cost underestimation was 86%, while the probability of overestimation was 14%. The cost underestimation was substantially

larger (maximum of +280%) than the cost overestimation (maximum –80%). Second, they found that rail projects (high speed, urban, and conventional) were the project type most plagued by cost overruns: the mean cost underestimation of rail projects was significantly larger than for the other two types of projects, fixed links (bridges and tunnels) and roads (highways and freeways).

These examples and the study by Flyvbjerg et al. (2002) clearly show that the construction costs of transportation infrastructure projects are often underestimated. Technical, economic-political, psychological and legal causes can explain the frequent underestimations (Molenaar et al. 2010, Flyvbjerg 2007). One of the technical causes is poorly performing estimation methods used to calculate the cost and duration of project construction. This indicates that there is the need for modeling tools that can capture the uncertainties in the construction process and their impact on construction cost and time of transportation infrastructure projects.

1.2 Models of Uncertainty for Infrastructure Construction Projects

Molenaar et al. (2010) reviewed risk analysis tools and management practices to control transportation project costs. They found that for the majority of infrastructure projects, the most probable construction cost and time (best guess) are calculated, and a contingency for unidentified risks is added to the cost estimate. In some cases, risks are identified with "risk checklists" including risks observed in past projects. In a few cases, the identified risks are analyzed, usually qualitatively, using so called *PxI* (Probability x Impact) matrices. In *PxI* matrices, first the probability of occurrence of the risk is ranked, usually from 1 to 5; second, the impact of the risk is ranked, also from 1 to 5; then, according to the probability rank and the impact rank, the risk is classified as high, moderate or low risk. At the Washington State Department of Transportation (WSDOT), a quantitative analysis of the identified risks is performed for large projects. In the quantitative analysis, using Monte Carlo simulation the occurrence of the risk is modeled and the impact of the risk is estimated. The quantitative analysis developed at WSDOT is the joint effort of the transportation agency, academic researchers and the industry (WSDOT 2005, Molenaar 2005, Reilly et al. 2004). To date, it is the most advanced tool to model uncertainty in transportation projects.

The current approaches to modeling uncertainty show two limitations: first, the reliance on modeling risks without an in-depth understanding of the construction process does not ensure that all uncertainties in the construction process are captured; second, modeling only risks rather than

e.g. the variability in the construction process, the correlations between costs and the occurrence of disruptive events does not capture the cumulative effect of different sources of uncertainty.

1.3 Objectives, Approach and Contributions

1.3.1 Objectives

The research objectives are:

1. Deepen the understanding of the construction process of rail lines and develop a construction model to represent the construction of the four main structures (tunnels, viaducts, cuts and embankments) of rail lines.
2. Deepen the understanding of uncertainty in the construction process of rail lines, and based on this develop an uncertainty model to capture the sources of uncertainty and model their cumulative impact on construction cost and time of rail lines.
3. Implement the construction model and the uncertainty model in a simulation tool that can be used to model the uncertainties in the construction of rail lines and other linear/networked transportation projects.
4. Apply the construction model and the uncertainty model by simulating the uncertainties in the construction of a rail line project in order to show the feasibility of the proposed models and their effectiveness in capturing the impact of uncertainty on the construction cost and time.

1.3.2 Approach

The research objectives have been pursued in four steps. First, the construction process of the four main types of structures in rail lines (tunnels, viaducts, cuts, and embankments) is analyzed at the level of single activities and represented in networks. The activity networks of the tunnels, the viaducts, and the cuts and embankments are interconnected in a construction network that models the construction of a rail line. The activity networks and their interconnection in the construction network are the construction model.

Second, sources of uncertainty in the construction process are identified and quantitative models of the sources of uncertainty are developed. The three sources of uncertainty identified in the construction process of rail lines and developed in this research work are the variability in activity cost and time, the correlations and the disruptive events. The variability is modeled with probability distributions. Five types of correlations are identified and their impact compared. The correlation with the largest impact is further pursued. It is modeled with the NORTA method (Cario & Nelson 1997). The occurrence of disruptive events is modeled with Markov processes and the cost and time impacts are modeled with probability distributions. These three sources of uncertainty and their quantitative models are the uncertainty model.

Third, the construction model and the uncertainty model are implemented in the simulation tool DAT (Decision Aids for Tunneling). The networks modeling the construction process of the four main structures (tunnels, viaducts, cuts and embankments) are implemented in the DAT. The quantitative models of the three sources of uncertainty are introduced in the DAT: the required probability distributions, the NORTA method, and the Markov processes.

Fourth, the construction model and the uncertainty model are brought together in the application to the construction of a section of the new Portuguese high speed rail line. The construction of all the structures of four alignments are modeled with networks, and the sources of uncertainty are modeled with probability distributions, the NORTA method, and Markov processes. The impact of the different sources of uncertainty are compared and the cumulative impact of the sources of uncertainty is analyzed. Through the application to the rail line project, the effectiveness of the proposed construction and uncertainty models is brought to light and the contributions of the research work become evident.

1.3.3 Contributions

The research work presented in this thesis makes two original contributions. First, it deepens the understanding of the construction process in the four main structures of rail lines: tunnels, viaduct, cuts and embankments. Second, it contributes to the understanding of the sources of uncertainty by quantitatively modeling variability, cost correlations, and disruptive events, and by capturing their cumulative impact.

Understanding of the construction process. The construction process of rail lines and its

four main structures (tunnels, viaducts, cuts and embankments) is analyzed down to the activity level and the interconnections between activities. The construction of rail lines is modeled with activity networks in a bottom-up approach: single activities are connected into sub-networks modeling the repetitive processes, sub-networks are connected into structure networks modeling the construction of each structure, and structure networks are connected into the construction network modeling the construction of the entire rail line. The representation of the construction process of rail lines with activity networks deepens the understanding of the construction process and provides insight when identifying the sources of uncertainty (second contribution).

Understanding, modeling and quantitatively capturing the sources of uncertainty.

In a novel approach, uncertainty is analyzed at the activity level, from the perspective of its sources and considering both positive and negative outcomes. Three sources of uncertainty have been identified. The first identified source of uncertainty is the variability of the activity cost and time modeling the construction process: when an activity is repeated, the cost and time of the activity change from repetition to repetition. The second source of uncertainty is the correlations between the costs of the activities: positive correlations between the activity costs cause the standard deviation of the total cost to increase. The third source of uncertainty is the disruptive events. Due to their large cost and time impact, they can cause large increases in the total cost and total time. For the first time, these three sources of uncertainty are modeled jointly at the activity level: the cost and the time of an activity are variable; the cost of the activity is correlated with the costs of other activities; and during the activity, one or more disruptive events can occur. Analyzing and modeling the sources of uncertainty shows the impact of the single source of uncertainty and, most importantly, it captures the cumulative impact of different sources of uncertainty on project cost and time.

Simulation tool to model the uncertainty in rail line construction. The simulation tool DAT in its extended form integrates the construction and uncertainty models to simulate the uncertainty in the construction of rail lines. The activity networks modeling the construction of not only tunnels but also viaducts, cuts and embankments are implemented in the DAT. Also implemented in the DAT are: 1) the probability distributions, which model the variability of the activity time and cost and the cost and time impacts of disruptive events, 2) the correlation model NORTA, which can represent the cost correlations, and 3) the Markov processes, which model the

occurrence of disruptive events. In their extended version, the DAT are a simulation tool that can be used to model the uncertainty in the construction of rail lines and other linear/networked transportation projects.

Demonstration of the feasibility and the effectiveness of the construction and uncertainty models. The practical application of the construction and uncertainty models to the construction of a section of the new Portuguese high speed rail line clearly shows the feasibility of the models and their effectiveness in capturing the uncertainty in the construction process. The 90th percentiles of the simulated construction cost and time are significantly larger than the deterministic (single number estimate) construction cost and time: in tunnel construction the 90th percentile of the construction cost is 58% larger than the deterministic cost estimate, while in earthwork (cuts and embankments) the 90th percentile of the construction time is 94% larger than the deterministic time estimate. Thus, the construction and uncertainty models provide invaluable insight on the magnitude and the impact of the sources of uncertainty on the construction cost and time of transportation infrastructure projects, such as rail lines.

1.4 Thesis Structure

This thesis is organized in eight chapters:

Chapter 1 contains the introductory material and the presentation of the research topic.

Chapter 2 reviews causes of cost underestimation and presents a detailed literature review of the available models of uncertainty in transportation.

Chapter 3 addresses the first research objective, i.e. the construction model is developed: the construction processes of tunnels, viaducts, cuts and embankments are studied at the activity level and are represented in corresponding activity networks.

Chapters 4 and 5 fulfill the second research objective. Chapter 4 proposes the uncertainty model: this includes the three sources of uncertainty variability, cost correlations and disruptive events. These are modeled quantitatively with probability distributions, the NORTA method, and Markov processes.

Chapter 5 focuses on the second source of uncertainty: correlations. It reviews available correlation measures and correlation models, and identifies as best suited the Spearman correlation coefficient and the NORTA method. These are then used to compare the impact of five types of

correlations identified in the construction of rail lines on project cost and time. The correlation between the costs of a repeated activity emerges as the correlation type with the by far largest impact on the total cost.

Chapter 6 presents the simulation tool to implement the construction model and the uncertainty model. The simulation tool, DAT (Decision Aids for Tunneling), is expanded in its capabilities to model not only tunnels but also viaducts, cuts and embankments. The construction model is implemented with the introduction of the activity networks. The uncertainty model is implemented with the introduction of a new probability distribution (lognormal), the correlation model NORTA, and by modeling the occurrence of disruptive events with Markov processes.

Chapter 7 brings together the construction model and the uncertainty model in an application to the construction of four alignments of the new Portuguese high speed rail line. In the DAT, first the construction of the four alignments is simulated based on activity networks. Second, the construction is simulated including the sources of uncertainty (variability, cost correlations, disruptive events). Then, the individual and the cumulative impacts of the sources of uncertainty on project cost and time are analyzed. Through the application to the rail line construction project, Chapter 7 makes the contributions of the construction and uncertainty models evident.

Finally, Chapter 8 summarizes the findings of the research work and provides suggestions for future work.

Chapter 2

Literature Review: Risk in the Construction of Transportation Infrastructure Projects

Cost and time underestimations are widespread in the construction of transportation infrastructure projects. This shows the need for modeling tools that can capture the uncertainty and its impact on construction cost and time. The author reviewed the academic and professional literature on models of uncertainty for transportation infrastructure projects. After presenting examples and a statistical study of cost underestimation in transportation infrastructure projects, this chapter presents the literature review on identified causes and proposed cures for cost underestimation. Cost underestimation is widespread and more common in rail line projects than in other transportation infrastructure projects. A group of literature sources identified technical, psychological and economic-political causes for cost underestimation and suggested the use of improved estimation tools, debiasing techniques and a change in policy to counteract the identified causes. The transportation community identified risk factors and proposed eight strategies to counteract the impact of the risk factors on the construction cost and time of transportation infrastructure projects. The proposed cures will show two limitations: the lack of an in-depth understanding of the construction process and its uncertainties, and the limitation of modeling risks only without capturing the cumulative effect of different sources of uncertainty. To overcome these limitations, a construction model and an uncertainty model will be introduced.

The detailed structure of the chapter is as follows: first, two notorious examples of cost and time underestimations in transportation are described (section 2.1). Second, a statistical study of cost underestimation in transportation infrastructure projects is presented (section 2.2). Then, the definition of risk, the causes of cost underestimation and the tools to counteract cost underestimation are discussed in the following sections (Table 2.1):

- Three explanations for cost underestimation and two cures recommended by Flyvbjerg et al. (2002) and Flyvbjerg (2007) are presented in section 2.3.
- Definitions of risk, 18 risk factors causing cost underestimation and eight strategies suggested by Anderson et al. (2007) and Molenaar et al. (2010) are discussed in section 2.4, with a focus on one of the eight strategies: the risk strategy.

Since the core of the research work presented in this thesis are a construction model and an uncertainty model to identify and analyze uncertainties in the construction of rail lines, all sections in this chapter focus on uncertainties, on their causes and on tools to identify and analyze such uncertainties, as indicated in Table 2.1.

Table 2.1: Structure of the chapter. Causes and cures of cost underestimation according to Flyvbjerg are presented in section 2.3, while risk definitions, causes and cures according to the transportation community are discussed in section 2.4.

	Flyvbjerg	Transportation community
references	Flyvbjerg et al. (2002) Flyvbjerg (2007)	Anderson et al. (2007) Molenaar et al. (2010)
section	2.3	2.4
definition	–	risk
causes	three explanations	18 risk factors
cures	policy change and reference class forecasting	eight strategies focus: risk strategy

2.1 Big Dig and Swiss Transalpine Base Tunnels: Examples of Cost Underestimation

Two notorious examples of cost underestimation in transportation infrastructure projects are the Big Dig in Boston and the Swiss transalpine base tunnels. Construction cost and time underestimation in these projects and their causes are discussed.

In 1990, it was estimated that the Big Dig in Boston would have cost \$6 billions and would have been completed by 2001. Unfortunately, the project cost almost triplicated to a total of nearly \$15 billions and was completed only in 2007, six years after initially forecast (Salvucci 2003). Several reasons caused the cost increase: among others scope changes, new safety requirements, technology advancements and a lacking cost control mechanism. The scope of the project was changed by several additions of tunnels and viaducts to the project: for example, the Ted Williams tunnel was expanded with an additional tunnel connection to the airport, which was not originally planned. Considering safety requirements, the federal government requested the replacement of an existing 35 m.p.h. multiple lane underground curve with a 45 m.p.h. curve which required extensive reconstruction work. A large increase in cost was caused by the introduction of the Intelligent Highway Vehicle System: this technology cost approximately three quarters of a billion dollar. Last and probably most importantly, a comprehensive cost control from project inception to start of operation was missing (Salvucci 2007). In 1990 the project included a contingency for unidentified risks of \$100 millions: interestingly, these funds are equal to just 1.7% of the initial estimate of \$6 billions, an extremely low percentage for contingency in the design phase.

The transalpine base tunnels in Switzerland will connect Italy to Central and Norther Europe with two new tunnels: the Lötschberg base tunnel and the Gotthard base tunnel, which will be the longest rail tunnel in the world. While the first is complete and operational, the second has been excavated and will be operational in 2017. Both projects were plagued by the same problem: major scope changes were added to the project by political entities, although no further funding was allocated so that the scope changes are being paid for with project contingencies (Pfisterer 2005, Teuscher 2007, Zbinden 2007). Scope changes occurred mostly in three areas: more stringent safety requirements, new technology in rail infrastructure, and project improvements in favor of the environment and of people impacted by the project (Pfisterer 2005, Zbinden 2007). The new technology in rail infrastructure increased the construction cost of the two projects by 290

million Swiss francs, while the cost of the Gotthard base tunnel increased further by half a billion Swiss francs to construct a tunnel rather than an open track section at the Northern portal. In the case of the Gotthard base tunnel, scope changes were so extensive that the contingency had been exhausted in the early stages of the project (Pfisterer 2005).

In the case of the Lötschberg base tunnel, 3,214 million Swiss francs were initially budgeted with a 15% contingency of 482 millions. Eventually, in order to cover the rising costs due to scope changes, an additional 582 millions were allocated with additional contingencies of 15%, corresponding to 87 millions (all 1998 prices). This amount sums to a total of 4,365 million Swiss francs, over a billion Swiss francs more than initially estimated. At the start of the operational life, the Lötschberg tunnel had cost approximately 4,300 millions, i.e. 65 millions less than estimated (Teuscher 2007). In the case of the Gotthard base tunnel, a final assessment of the construction cost is not possible since the project has not been completed yet. Nevertheless, it is known that the initially estimated cost was 7,716 million Swiss francs and the current additional cost amounts to 2,833 millions. Some insight is offered by the breakdown structure of the current additional cost of the Gotthard base tunnel: over three fourth are due to measures in favor of the environment, new technology and higher safety standards, and delays in the construction process; technology and safety are responsible of 50% of the overall cost increase. On the other hand, only 22% of the cost increase is caused by geology and additional costs to award contracts (Figure 2-1).

The Big Dig and the Swiss transalpine base tunnels are just two among many examples of transportation infrastructure projects completed with large cost- and time overruns. A statistical study of cost underestimation in transportation infrastructure projects in the last 70 years by Flyvbjerg et al. (2002) is presented in the next section.

2.2 A Statistical Study of Cost Underestimation in the Construction of Transportation Infrastructure Projects

This section presents the results from a study (Flyvbjerg et al. 2002) that showed cost underestimation was widespread in the construction of transportation infrastructure projects, with interesting trends about underestimation magnitude and depending on type of project.

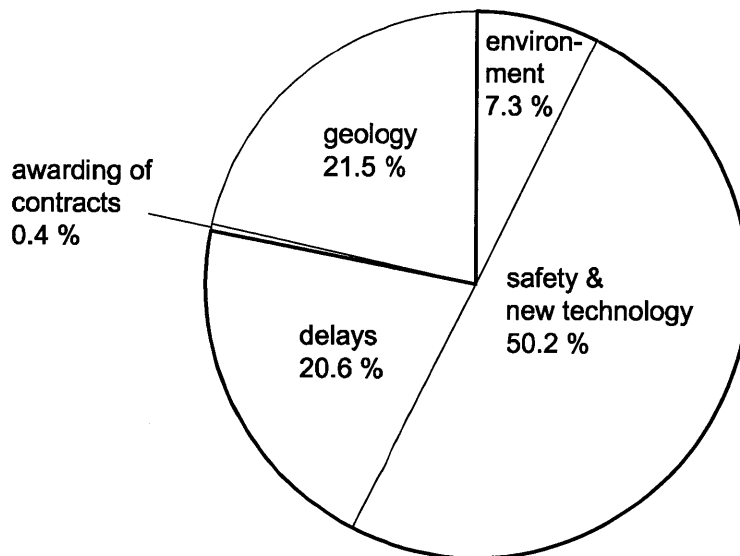


Figure 2-1: Cost breakdown structure of the current additional costs for the Gotthard base tunnel: while 22% of the additional costs are caused by geology and the contract awarding process, 78% of the costs are caused by additional investments in the project in the form of measures to protect the environment, new technology and higher safety standards, and delays in the construction and decision process (Teuscher 2007).

Flyvbjerg et al. (2002) conducted a study on cost underestimation in transportation infrastructure projects involving 258 projects. In the study, they attempted to answer, among other things, two questions: 1) do overestimated costs occur as often as underestimated costs?, and 2) does the type of project have an impact on the magnitude of the cost underestimation?

Before presenting the results from that study, it is important to elucidate the study framework, in particular how cost underestimation was calculated, general characteristics of the projects analyzed, and the sources of data. Cost escalation was defined as the difference between actual construction cost and estimated construction cost, in percent of the estimated construction cost. If cost escalation was positive the construction cost was underestimated, while if cost escalation was negative the construction cost was overestimated. The study analyzed the costs of 258 projects ranging from bridges, tunnels, highways, freeways to high-speed rail, urban rail and conventional rail. These were constructed between 1927 and 1998 in 20 different countries on 5 continents. Their construction costs varied between \$1.5 million to \$8.5 billion summing up to a project portfolio of approximately \$90 billion. Construction costs were discounted to 1995 level prices, corrected

using geographical and historical indices and converted from their original currency using the applicable (at the date) exchange rates. The study was based on information from several sources, mainly from their own data but also from other sources (Fouracre et al. 1990, Hall 1980, Leavitt et al. 1993, Lewis 1986, Merewitz 1973, National Audit Office 1985, National Audit Office & Scottish Development Department 1988, National Audit Office 1992, Pickrell 1990, Riksrevisionsverket 1994, Vejdirektoratet 1995, Walmsley & Pickett 1992).

The first relevant finding of the study (Flyvbjerg et al. 2002) concerned the distribution of the construction cost: underestimation errors were more common than overestimation errors (Figure 2-2). Several findings supported this statement: first, the average cost escalation was 27.6% (rather than 0%, as it were the case for a symmetric distribution). Second, the probability of cost underestimation (positive cost escalation) was 86%, while the probability of cost overestimation (negative cost escalation) was 14%. Assuming a normal distribution for the cost estimation error, the mean is equal to the mode and the median and should be equal to 0% if there is the same probability of cost underestimation and overestimation. Third, the cost underestimation (positive cost escalation) was substantially larger (maximum of +280%) than the cost overestimation (negative cost escalation) (maximum of -80%) since the distribution is skewed to the right. These observations were corroborated by two statistical tests: a statistically significant two-sided test (using the binomial distribution) rejected the null hypothesis that a cost underestimation error was as common as a cost overestimation error, and a nonparametric Mann-Whitney test proved statistically significant and rejected the null hypothesis that the magnitude of cost underestimation errors (positive cost escalation) was the same as the magnitude of cost overestimation errors (negative cost escalation).

The second finding related to the type of project: subdividing the projects into rail (high speed, urban, and conventional), fixed link (bridges and tunnels), and road (highways and freeways) revealed that their cost escalation distributions differed in mean and standard deviation (Table 2.2 and Figure 2-3). This finding was confirmed by a statistically significant F-test rejecting the null hypothesis that the type of project had no impact on the cost escalation. Most importantly for this doctoral thesis, the rail projects were characterized by the largest mean cost escalation, when compared to the other two types of projects, the fixed link projects and the road projects. Finally, since the sample size of fixed link projects was relatively small (33, Table 2.2), statistical tests based on the fixed link sample should be considered with caution.

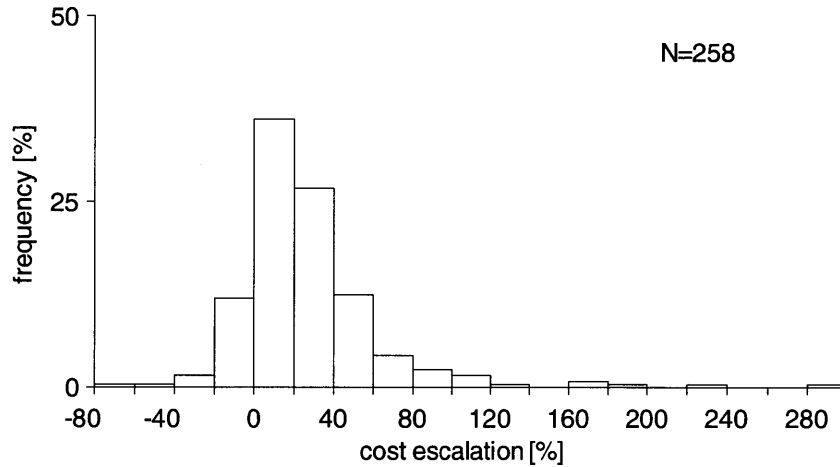


Figure 2-2: Distribution of the cost escalation in transportation infrastructure projects. Underestimation (positive cost escalation) was far more common (86%) than overestimation (negative cost escalation) (14%), and the underestimation was also substantially larger (up to 280% compared with -80%). After Flyvbjerg et al. (2002).

Table 2.2: Cost escalation means and standard deviations for all projects as well as for the subgroups of rail, fixed link, and road projects, respectively. Rail projects had the largest cost escalation, fixed link also lay above average, while the road projects seemed to be the best contained in terms of cost escalation (after Flyvbjerg et al. 2002).

Project type	Sample size	Cost escalation [%]	
		mean	st. dev.
all projects	258	27.6	38.7
rail	58	44.7	38.4
fixed link	33	33.8	62.4
road	167	20.4	29.9

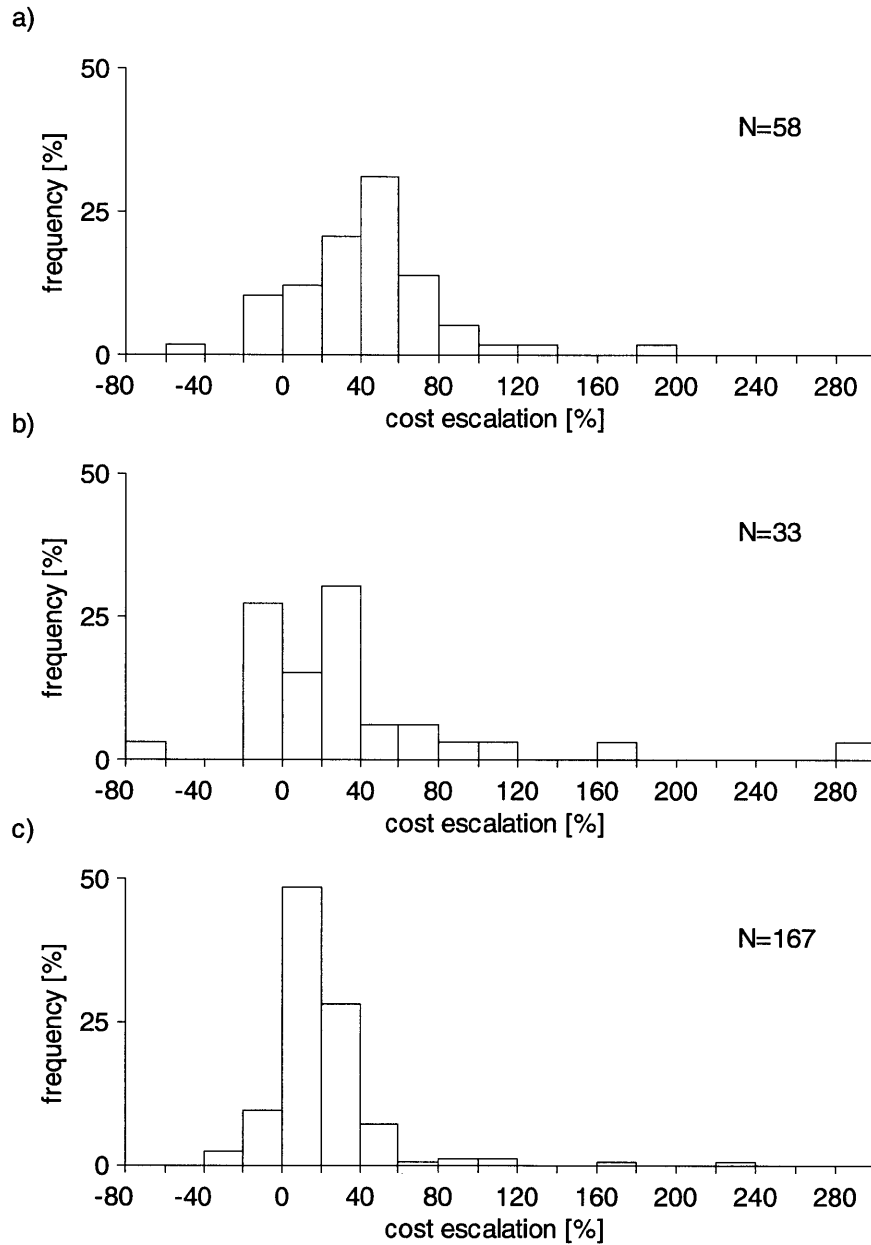


Figure 2-3: Distribution of cost escalation for a) rail projects, b) fixed-link projects (tunnels and bridges), and c) road projects. The type of project influences the magnitude of cost escalation, with rail projects suffering from the largest percentages. After Flyvbjerg et al. (2002).

Flyvbjerg et al. (2002) offered a groundbreaking insight in the realm of cost underestimation in transportation infrastructure projects: it is clear that project costs are consistently underestimated. Before discussing causes and potential remedies to the recurring cost escalations, I would like to add two open-ended comments to Flyvbjerg et al.'s (2002) study. First, treating cost escalations above 160% as outliers (5 data points out of 258) and excluding them in the statistical analyses might change findings and conclusions, particularly about fixed link projects and/or road projects. Second, although aggregating bridge and tunnel projects in one type of projects, named fixed link projects, increases the sample size to 33, engineering judgment indicates that structure type and most importantly construction processes are radically different between bridges and tunnels, thus questioning the aggregation choice. Nevertheless, the pattern of cost escalation is clearly identifiable in the statistical analyses presented by Flyvbjerg et al. (2002).

Flyvbjerg et al. (2002) thus showed that in transportation infrastructure projects cost underestimation occurred more often than cost overestimation, and rail projects were characterized by larger cost underestimations than fixed link (tunnels and bridges) and road projects. Following these observations, the next sections present an in-depth literature review of the causes of cost underestimation in transportation infrastructure projects and the cures currently available to counteract cost underestimation.

2.3 Causes and Cures of Cost Underestimation according to Flyvbjerg

Flyvbjerg et al. (2002) and Flyvbjerg (2007) identified causes for the widespread occurrence of cost underestimation in transportation infrastructure projects, and suggested potential cures (Table 2.3). This section discusses the causes (section 2.3.1) and the cures (section 2.3.2) of cost underestimation proposed by Flyvbjerg. A specific cure, called reference class forecasting, is presented in section 2.3.3. The transportation community has identified other causes for cost underestimation and suggested other cures. These are discussed in section 2.4.

2.3.1 Causes

Flyvbjerg et al. (2002) and Flyvbjerg (2007) proposed three explanations for the widespread oc-

Table 2.3: Causes and cures for the cost underestimation in infrastructure construction projects according to Flyvbjerg. There are three types of causes: technical, economic-political, and psychological. Cost underestimation can be limited with a policy change, improved forecasting techniques, such as reference class forecasting, and debiasing techniques.

Flyvbjerg's	
causes	cures
technical	policy change
economic-political	improved estimation tools
psychological	debiasing techniques

currence of cost underestimation in transportation infrastructure projects: technical explanations, economic-political explanations, and psychological explanations.

Technical explanations interpreted forecasting errors in technical terms: estimation errors caused by imperfect forecasting techniques, inadequate data, honest mistakes, the intrinsic problem of forecasting the future, inexperienced forecasters and/or other technical difficulties. According to Flyvbjerg et al. (2002), such explanations could not explain the overwhelming evidence of recurrent cost underestimation for several reasons. First, if cost underestimation had been caused by technical errors, its distribution would be symmetric around zero, i.e. there would be equal probability of cost underestimation and cost overestimation; however, this was not the case (section 2.2). Second, error sources and forecasting techniques are expected to improve over time, while cost and time overruns did not decrease in the last years and were still recurrent. Third, although geological, environmental, safety and general construction problems for a specific project could not be always forecast, the fact that such problems occur was known and it should have been accounted for in the cost estimation. The question thus arises if and how the technical causes of recurrent cost underestimation could be contained. Flyvbjerg (2007) suggested that cost estimation errors due to technical reasons could be limited or eliminated with better forecasting techniques, improved data and experienced forecasters.

Flyvbjerg et al. (2002) mentioned two types of economic-political explanations for cost underestimation: economic self-interest and economic public interest. In the former case, planners may have been interested in increasing the chances of a project obtaining funding with the help of favorable cost forecast, since a funded project created further work for engineers and construction firms. For them, cost underestimation would have been economically rational because it potentially increased revenues. Flyvbjerg et al. (2002) defined such deliberate cost underestimation as a lie

and, since it was economically-driven, they note that the cost underestimation lie paid off. In the case of economic public interest, planners and project promoters underestimated costs to provide an incentive to cut costs and save public money, according to the principle that larger cost estimates give incentives for wasteful expenditures. For reasons of economic efficiency, this argument by the planners and the project promoters was clearly flawed: cost underestimation led to falsely favorable cost-benefit ratio, meaning that a not viable project may have been granted funding and/or a project with larger returns may have been denied financial support. Both self-interest and public interest driven cost underestimations could be reined in by measures of accountability (Flyvbjerg 2007) (see below).

The last type of explanations for the occurrence of cost underestimation was a psychological one: optimism bias. This was the tendency by planners and project promoters to be excessively optimistic by focusing on success scenarios and overlooking the chance of mistakes and failure. Due to optimism, decision makers did not weigh gains, losses and their probabilities objectively but rather optimistically. It has been argued that optimism bias is common among decision makers in companies and industries different from transportation infrastructure (Lovallo & Kahneman 2003). Nevertheless, optimism bias could be reined in with simple reality checks and by using debiasing techniques (Flyvbjerg 2007).

Flyvbjerg et al. (2002) do not consider two important aspects in their analysis of the technical causes of cost underestimation: the increase in cost variance with larger project scope, and the increase over time in technical complexity. Scope creep causes increases in the construction cost, as well as in the construction cost variance. Thus, the statistical study (Flyvbjerg et al. 2002) should have checked the projects in the study sample for scope creep. The second aspect is the new technologies developed over time that make it possible to construct projects inconceivable in the past: new technologies often add to the technical complexity of projects, which is a main driver of cost escalation.

As was just shown Flyvbjerg et al. (2002) and Flyvbjerg (2007) identified three causes of cost underestimation: technical causes, economic-political causes, and psychological causes. In order to counteract these causes, Flyvbjerg and his research group proposed cures to cost underestimation (next section).

2.3.2 Cures

Technical causes and psychological causes (biases) could be limited or eliminated with improved estimation tools and debiasing techniques, while economic-political causes would require a change in policy. A change in policy that creates incentives to reward honesty and punish economic-political misrepresentations is briefly introduced here. In section 2.3.3, a specific cure for technical and psychological causes of cost underestimation (reference class forecasting) is presented.

Flyvbjerg et al. (2003) proposed a policy change based on accountability. This could be implemented and guaranteed through four basic instruments: transparency, performance specifications, specifying the regulatory regime and risk capital. These four cornerstones are briefly highlighted in the following:

- Transparency is the central instrument to implement accountability in the public sector: it means a public discussion of projects involving stakeholders and civil society.
- Performance specifications derived from policy objectives and public interest requirements, such as economic performance, sustainability, environmental impact and safety. Performance specifications shift the perspective from a technical solution-driven approach to a goal-driven approach. It also means that all project requirements must be decided before exploring technical alternatives (note that, differently from Flyvbjerg et al. (2003), in the construction industry performance specifications refer to the structure performance in terms of e.g. deflections).
- The regulatory regime is the compound of economic rules defining project construction and operations. For instance, it describes the project financial sources, such as public funding, use of tolls, etc. It is central to the implementation of accountability since it forces the planner to identify all costs before reaching decisions and it determines the level of risk in the project.
- Flyvbjerg et al. (2003) also proposed to condition the construction of a project on the participation of private investors without a sovereign guarantee (i.e. the government does not guarantee the investment in case of default). The idea behind this proposal is that private investors would finance a project, thus bear the consequences of a wrong decision, only if the project is indeed viable.

More details about policy-driven solutions to cost estimation can be found in Flyvbjerg et al. (2003) and Flyvbjerg (2007).

Flyvbjerg et al. (2003) proposed a policy change to counteract economic-political causes of cost underestimation. In order to counteract technical and psychological causes, Flyvbjerg (2006) suggested reference class forecasting, an estimation tool that include debiasing techniques. This is presented in the next section.

2.3.3 A Specific Cure: Reference Class Forecasting

Flyvbjerg (2006) proposed to apply to transportation construction projects reference class forecasting, a methodology developed in the fields of psychology and management (Lovallo & Kahneman 2003), in order to counteract technical and psychological causes of cost underestimation and improve the accuracy of cost and duration estimations in transportation.

Reference class forecasting is a tool that estimates the performance, e.g. cost or time, of a project based on statistical analyses of past projects ("outside view") rather than on the specifics of the project itself ("inside view"). Differently from the inside view, the outside view relied on statistics of performance in comparable projects.

Developed to increase the accuracy of forecasts, reference class forecasting was effective against the negative impact of optimism bias, the psychological explanation to cost underestimation according to Flyvbjerg et al. (2002) and Flyvbjerg (2007). The majority of people were very optimistic: they tended to exaggerate their talents, they took credit for successes while they attributed failures to external factors, they overestimated the level of their control over events dismissing the specifics of the case (Lovallo & Kahneman 2003). Being problem solvers, engineers in particular were biased when considering risk since they naturally focused on solutions (Molenaar et al. 2010). Interest-driven cost underestimation occurred when planners and politicians are not interested in improving forecasts but rather in getting a project approved (Flyvbjerg et al. 2003): given that planners and politicians were interested in better forecasts, reference class forecasting was a tool that sidestepped optimism bias to deliver more realistic estimates (Flyvbjerg 2008).

The difference between the outside and inside view and the impact either view could have on a forecast was given with an anecdote (Lovallo & Kahneman 2003). A team of academics and teachers developing a new school curriculum was asked how long it would take to complete the

project: their individual estimates ranged between 18 and 30 months. A team member, expert in curriculum development, was asked to recall from his experience similar projects with comparable teams and at a comparable stage, and how long it took to complete such projects. Surprisingly, around 40% of the projects were abandoned and the rest required between 7 and 10 years. In the light of this second estimate not surprisingly, it took the team of academics and teachers 8 years to develop the new school curriculum. In this example, the first estimate, between 18 and 30 months, is the inside view: the team members focused on the specific project, its objectives, resources, obstacles, possible scenarios, and current trends. In contrast, in the outside view the expert ignored the details of the project but rather recalled his experience with a class of similar projects and offered a distribution of the outcome. The divergence of estimations between inside and outside view has been confirmed in research studies (Gilovich et al. 2002).

Besides being the traditional approach, the inside view was intuitively adopted by the majority of individuals and organizations when planning major initiatives. It was the natural approach to focus on the project, to dissect it and analyze all details especially focussing on its specific characteristics, while it might have been somehow counter-intuitive to gather statistics on similar projects. Unfortunately, the inside view is less likely to yield realistic estimates than the outside view. Outside-view thinking could be contributed by outsiders, such as consultants, with data on other companies and their projects (Lovallo & Kahneman 2003).

Reference class forecasting can be applied to transportation projects in three steps that allow one to systematically take an outside view (Flyvbjerg 2006). First, a reference class of past, similar projects is identified. A class should strike a balance between two conflicting goals: it should be broad enough to be statistically meaningful but narrow enough to include only truly comparable projects. For example, in Flyvbjerg et al. (2002) a class included all fixed links, tunnels and viaducts: the class size was 33 projects. Despite being debatable, the class was considered statistically meaningful with its 33 data points, and tunnels and viaducts were considered comparable projects. Second, the probability distribution of the relevant performance measure (e.g. cost, time) is created based on empirical data from the reference class projects. In the fixed links class above, the cost and time distributions would be generated based on the empirical cost and time of the 33 projects. Third, the project considered is compared with the reference class distribution in order to estimate the most likely outcome. For instance, planners designing a new subway would establish a

reference class of comparable projects. If the performance measure of interest were cost escalation, they would collect information and generate the cost escalation distribution of the reference class, with engineering judgment position the new subway project in the distribution, and read out the probability of cost escalation for the new subway project (after Flyvbjerg 2006).

Flyvbjerg & Cowi (2004) applied reference class forecasting to three categories of projects in transportation: roads, rail and fixed links. Roads included projects such as highways, bicycle and pedestrian facilities; the rail category consisted of metro, conventional and high speed rail, and others; bridges and tunnels were in the fixed link category. The three categories corresponded to three reference classes, for which data were gathered to establish probability distributions of cost overruns. From the cumulative distribution of the cost overrun, one could read the cost overrun for a given probability. For example, in the rail reference class, there was a probability of 50% to incur into a cost overrun of 40%.

As an estimation method, reference class forecasting has two caveats: it may be impossible to predict extreme outcomes and it may be difficult to find comparable projects (Lovallo & Kahneman 2003, Flyvbjerg 2006). Based on historical, empirical data, reference class forecasting may be ill-suited to forecast extreme outcomes that due to their nature could not be captured by statistical analysis. On the other hand, creating a reference class for a project involving the introduction of a new technology may be difficult: one could compare projects that introduced other new technologies, or projects by other companies that introduced the same technology, or adopt other approaches. In transportation construction, a project usually involves the construction of new infrastructure (e.g. the first subway in the city) with well-known technologies. In such cases, the reference class forecasting could be well suited.

In the construction industry, the use of reference class forecasting suffered from one major disadvantage: it may lead to too large construction cost estimates with diverging consequences. If the construction cost is deemed too high, the project could be abandoned. This aspect is connected to the interest-driven cost underestimation (section 2.3). On the other hand, if the project is approved, such large estimates could lead to inefficiency and overspending: since the money had been allocated, it would be perceived as available to the project and its contractors. Thus, the cost estimate established with the reference class forecasting may create an incentive working against cost control, if the money allocated is more than the money actually needed for

construction (Flyvbjerg 2006).

Although Flyvbjerg (2006) attempts to make a strong case in favor of reference class forecasting and the outside view in transportation, he also acknowledged the usefulness of other approaches based on the inside view, such as the Cost Estimate Validation Process (CEVP) developed by the Washington State Department of Transportation (WSDOT 2005) and the use of Monte Carlo simulations. While these make use of subjective data, the reference class forecasting is exclusively used in the framework of empirical, statistically-analyzed data.

Flyvbjerg (2006) proposed reference class forecasting to counteract technical and psychological causes of cost underestimation. Flyvbjerg and his research group identified three causes of cost underestimation (technical, psychological, economic-political), and proposed cures (policy change, reference class forecasting) to counteract these causes. Differently, the transportation community has identified other causes and proposed other cures for cost underestimation in infrastructure construction projects (next section).

2.4 Causes and Cures according to the Transportation Community

The transportation community has identified causes of cost underestimation in infrastructure construction projects different from the ones identified by Flyvbjerg and his research group, and has suggested other cures. In this section factors causing cost escalation are listed and strategies to counteract cost escalation are introduced. Before discussing causes and cures, first the concept of risk needs to be defined (section 2.4.1). It will become clear that there is no agreement in the transportation community on such a central aspect as the definition of risk. Second, causes of cost underestimation are presented (section 2.4.2). They are subdivided into external and internal causes (Table 2.4). Third, eight cures/strategies to eliminate and/or limit cost underestimation are discussed (section 2.4.3). Among them, the risk strategy is discussed in detail. It consists of five phases: risk identification, risk analysis, risk mitigating and planning, risk allocation, and risk monitoring and controlling (Table 2.4). Specifically, risk identification tools (section 2.4.4) and risk analysis tools (section 2.4.5) are described in detail. The remaining phases of the risk strategy (mitigating and planning, allocation, and monitoring and controlling) are presented in section 2.4.6.

Finally, the most advanced tool in risk identification and analysis, the Cost Estimate Validation Process, is discussed in detail (section 2.4.7).

The focus on the tools of the first two phases of the risk strategy, identification and analysis, is needed since the uncertainty model proposed by the author in chapter 4 allows one to identify and analyze uncertainties. Thus, there is an interest in giving the complete overview of the identification and analysis techniques used in the transportation community before introducing the new tool.

Table 2.4: Causes and cures for cost underestimation in infrastructure construction projects according to the transportation community. Causes are subdivided into 11 external causes and 7 internal causes. Cures are eight strategies. One of these is the risk strategy, which consists of five phases: identification, analysis, mitigation and planning, allocation, and monitoring and controlling.

Transportation community	
causes	cures
external	management strategy
bias	scope and schedule strategy
delivery and procurement approach	off-prism strategy
project schedule changes	risk strategy
engineering and construction complexities	– risk identification
scope changes (controllable by the agency)	– risk analysis
scope creep (controllable by the agency)	– risk mitigation and planning
poor estimations	– risk allocation
inconsistent application of contingencies	– risk monitoring and controlling
faulty execution	delivery and procurement strategy
ambiguous contract provisions	document quality strategy
contract document conflicts	estimate quality strategy
internal	integrity strategy
local concerns and requirements	
effects of inflation	
scope change (not controllable by the agency)	
scope creep (not controllable by the agency)	
market conditions	
unforeseen events	
unforeseen conditions	

2.4.1 Risk Definition

The fact that the concept of risk is at the very core of any risk strategy makes it even more remarkable that there is no agreement on the definition of risk in the scientific and professional communities. Definitions of risk in the construction engineering literature and in the more general project management literature are presented here and compared.

Even publications by a same institution may not agree on the definition of risk. The National Cooperative Highway Research Program publishes a dozen reports a year regarding highways in the US and in the past four years it has published one report about cost estimation management (Anderson et al. 2007) and one report about risk analysis and management (Molenaar et al. 2010). The first publication defines risk as "the combination of the probability of an adverse event and its consequences", while the second publication defines it as "an uncertain event or condition that, if it occurs, has a negative or positive effect on a project's objectives". Although both definitions rotate around an event, the first identifies the risk with the combinations of its probability and its consequences, while the second identifies the risk with the uncertain/adverse event itself. As we will see, these two diverging definitions of risk are recurrent in the literature.

Definitions of risk reported in the construction engineering-, in the natural hazard modeling- and in the project management literature are quoted in the following:

- Anderson et al. (2007): risk is "the combination of the probability of an adverse event and its consequences".
- Molenaar et al. (2010): risk is "an uncertain event or condition that, if it occurs, has a negative or positive effect on a project's objectives".
- Caltrans (2007): a project risk is "an uncertain event or condition that, if it occurs, has a positive or negative impact on at least one project objective".
- WSDOT (2005): a risk is "an uncertain event or condition that, if it occurs, has a positive or negative impact on a project".
- WSDOT (2008): a risk is "a combination of the probability of an uncertain event and its consequences".
- FTA (2004): a risk is "a potentially adverse circumstance, expressed mainly in terms of causing undesired cost growth or time delays".
- Guglielmetti et al. (2008): "the risk, R , associated with an identified hazard is defines as the product $R = PxI$, and is called "initial risk". P is the probability (or likelihood) of occurrence of a hazard, and I is the impact (or consequence, or severity) of the hazard, in terms of safety, time, and cost".

- ISSMGE (2004): a risk is the "measure of the probability and severity of an adverse effect to life, health, property, or the environment". It can also be expressed as "probability of an adverse event times the consequences if the event occurs".
- Einstein (2002): "Risk can be described as $R = P[U] \times \text{Consequence}$ where $R = \text{Risk}$, $P[U] = \text{Probability of unsatisfactory performance}$, and where the consequence can be expressed in financial or other terms".
- PMI (2008): a risk is "an uncertain event or condition that, if it occurs, has a positive or negative effect on a project's objectives".
- Institution of Civil Engineers and Faculty of Actuaries (2005): a risk is "a threat (or opportunity) which could affect adversely (or favorably) achievement of the objectives of an investment".

As anticipated, the definitions of risk in the literature identify risk with either an event with some consequences on the project objectives (Molenaar et al. 2010, Caltrans 2007, WSDOT 2005, FTA 2004, PMI 2008, Institution of Civil Engineers and Faculty of Actuaries 2005) or the combination of the probability and the consequences of such an event (Anderson et al. 2007, WSDOT 2008, Guglielmetti et al. 2008, ISSMGE 2004, Einstein 2002).

Another important divergence in the definition of risk is its character: some definitions opt for a neutral definition where the event has positive or negative consequences, others speak of an adverse event or an hazard, while others do not specify the negative and/or positive connotation of risk. On the other hand, it should be noted that the word risk has inherently a negative connotation implying the presence of a threat.

Although the concept of risk is not clearly defined, the word risk is profusely used in the next sections since risk is the word mainly used in the literature reported here.

In the transportation community there is no agreement on the definition of risk. Despite the lack of a unifying definition, identified causes (risks) and proposed cures by the transportation community are presented in the next sections.

2.4.2 Causes: Risk Factors

The risk factors causing project cost underestimation have been described in detail in the literature (Anderson et al. 2007, Molenaar 2005, WSDOT 2005, Caltrans 2007, FTA 2004). Among these research studies, Anderson et al. (2007) describe 18 main factors that lead to cost escalation in construction projects.

Risk factors are subdivided into factors from internal sources and factors from external sources although there is some overlap since factors can be a combination of external and internal sources.

The 11 factors from internal sources are (type of factor in brackets):

1. Bias: purposeful underestimation of project costs to ensure that a project remains in the construction program (economic-political).
2. Delivery and procurement approach: cost escalation is caused by shifting risk to a party unable to control it, and by lack of experience with a delivery method or procurement approach (economic).
3. Project schedule changes: these are due to project extensions, budget constraints and similar. Schedule changes cause additional costs depending on two primary components: the inflation, and the time of expenditure (technical).
4. Engineering and construction complexities: internal coordination errors between project components and constructibility problems (technical).
5. Scope changes (controllable by agency): it involves additions to the project (economic-political).
6. Scope creep (controllable by agency): accumulation of minor changes (economic-political).
7. Poor estimations: general errors, omissions, inadequacies and poor performance in planning and estimation procedures and techniques (technical).
8. Inconsistent application of contingencies: misuse and failure to define what costs contingency amounts cover (technical).
9. Faulty execution: faulty execution by the agency in managing a project (technical).

10. Ambiguous contract provisions: misunderstanding between the agency and other contractual parties (legal).
11. Contract document conflicts: they cause confusion and lead to changes during project construction (legal).

The 7 factors from external sources are (type of factor in brackets):

1. Local concerns and requirements: perceived negative impacts of construction on the local societal environment, as well as on the natural environment (economic-political).
2. Effects of inflation: relevant when 1) the project estimates are not communicated in year-of-construction costs, 2) the project completions is delayed, 3) the rate of inflation is greater than anticipated (economic).
3. Scope changes (not controllable by agency): additions to the project requested by external participants (economic-political).
4. Scope creep (not controllable by agency): accumulation of minor scope changes from external participants (economic-political).
5. Market conditions: lack of competition for a project (economic).
6. Unforeseen events: floods, hurricanes, tornadoes, other weather related incidents (technical).
7. Unforeseen conditions: unknown ground conditions, contaminated ground, utilities, others (technical).

As indicated there is some overlap between internal and external factors, namely between "project schedule changes" (internal factor 3) and "effects of inflation" (external factor 2), and between "scope changes" and "scope creep" that are controllable versus not controllable by the agency (internal factors 5 and 6 versus external factors 3 and 4). The first overlap is due to the fact that although inflation is an external factor, its impact is influenced by the time of expenditure, thus by schedule changes and delays. The second and third overlaps, scope changes and scope creep, can be decided by the agency or they can be imposed on the agency through e.g. public opinion pressure or political decisions.

A central difference in the search of sources for cost escalations is observed: while Flyvbjerg et al. (2003) presented technical, economic-political, and psychological explanations (section 2.3.1), Anderson et al. (2007) presents 18 factors of technical, economic, economic-political and legal nature, and none considering the psychological aspect of optimism bias. However, optimism bias is addressed in some of the estimation techniques presented below.

The transportation community has identified risk factors causing cost underestimation. It was shown that Anderson et al. (2007) subdivide risk factors into external and internal factors. In the next section, the cures proposed by the transportation community to counteract these risk factors are discussed.

2.4.3 Cures: Risk Strategy and Other Strategies

In order to counteract the risk factors causing cost underestimation, Anderson et al. (2007) propose eight strategies. In this section, these strategies are introduced with special attention on the risk strategy. The risk strategy consists of five phases: risk identification, risk analysis, risk mitigation and planning, risk allocation, risk monitoring and control. The first and second phases (identification and analysis) are the focus of the next two sections (sections 2.4.4 and 2.4.5). As mentioned earlier, the focus on the risk strategy and the first two phases is needed to give the background to the uncertainty model proposed by the author in chapter 4. In fact, the proposed uncertainty model is an improved tool to identify and analyze uncertainties.

To counteract the factors causing cost underestimation in projects, Anderson et al. (2007) propose eight strategies:

1. Management strategy: manage the costs and the cost estimation process during the entire project development. It is based on accuracy, consistency and transparency, and it involves training for the agency personnel, established estimation processes and the critical review of the estimates.
2. Scope and schedule strategy: create processes to control changes in project scope and project scheduling.
3. Off-prism strategy: proactive engagement of external stakeholders and assessment of macro-environmental conditions possibly influencing project costs. External stakeholders and macro-

environmental conditions are called off-prism cost drivers since they are not within the road/rail prism. The off-prism strategy means reaching out to the community to hear interests and concerns and analyzing the market and macroeconomic conditions.

4. Risk strategy: risk identification, quantification and mitigation.
5. Delivery and procurement strategy: apply appropriate delivery methods to better manage cost since project delivery influences both project risk and cost.
6. Document quality strategy: improve project documents aiming at accuracy and consistency of cost estimates.
7. Estimate quality strategy: achieve improved accuracy and consistency through uniform approaches and by employing qualified personnel.
8. Integrity strategy: minimize the influence of outside pressures that can cause biases in the estimates.

In this thesis, the focus is on the fourth strategy to counteract cost escalation: risk strategy.

The risk strategy involves a comprehensive consideration of risks by identifying the risks, quantifying their impact and mitigating their impact throughout the project. More specifically, it consists of five phases, which are repeated iteratively throughout the project development (Molenaar et al. 2010):

1. Risk identification: determine the risks that may affect the project, and document their characteristics (section 2.4.4).
2. Risk assessment/analysis: qualitative or quantitative analysis to analyze the probability of occurrence of risks and their consequences if they occur (section 2.4.5).
3. Risk mitigation and planning: decide how to approach (acceptance, avoidance, mitigation, or transfer) risks and plan risk management activities (section 2.4.6).
4. Risk allocation: allocate responsibility for a risk to one or several parties, typically through a contract (section 2.4.6).

5. Risk monitoring and control: report project performance compared to the risk management plan (section 2.4.6).

Risk management is iterative since the different phases can and should be repeated during the development of the project to keep track of new situations and changed risks. Through the development of a project, new risks arise and are identified, meaning that their impact and possible mitigation measures need to be considered while such risks are identified. On the other hand, other risks disappear in more advanced stages of the project and/or are averted with mitigation measures leaving new resources available to counteract the remaining risks.

Another important aspect of a risk strategy is its scalability. This means that the risk strategy can be applied successfully to small and relatively simple projects as well as to large and complex ones. It is valid for each phase of the risk strategy: for small projects, risks can be identified with the help of risk lists and analyzed with qualitative risk analysis, mitigation and monitoring may consist of few measures for the most important risks. For large projects, risk identification may require brainstorming with construction experts, risk analysis may include quantitative risk analysis based on Monte Carlo simulation, risk mitigation may involve a series of measures, risk allocation may be covered with insurance contracts, and risk monitoring and control may consist of possibly alarm systems for incumbent risks.

The terminology risk assessment/analysis requires some clarification: although Molenaar et al. (2010) use the terms analysis and assessment interchangeably, here the term analysis is preferred according to the definition by Fell et al. (2005) from the field of landslide risk. Risk assessment is defined as "the process of making a decision recommendation on whether existing risks are tolerable and present risk control measures are adequate, and if not, whether alternative risk control measures are justified or will be implemented", while risk analysis is defined as "the use of available information to estimate the risk to individuals or populations, property or the environment, from hazards". Also Fell et al. (2005) define qualitative risk analysis as "an analysis which uses word form, descriptive or numeric rating scales to describe the magnitude of potential consequences and the likelihood that those consequences will occur", and quantitative risk analysis as "an analysis based on numerical values of the probability, [...] and consequences, and resulting in a numerical value of the risk". Thus, the definition of (qualitative/quantitative) risk analysis by Fell et al. (2005) corresponds to the description of risk analysis/assessment according to Molenaar et al. (2010), i.e.

qualitative or quantitative analysis to analyze the probability and impact of risks.

In this section, eight strategies to counteract risk factors causing cost underestimation have been introduced. The focus has been on one of these strategies, the risk strategy, and its phases. In the next two sections, tools to identify (first phase of the risk strategy) and tools to analyze (second phase) risks are discussed in detail. To offer a better overview of the available tools to identify and analyze risks, the tools are summarized in Table 2.5.

Table 2.5: Tools used in the identification phase and the analysis phase of the risk strategy. Expert interviews, risk management plan, risk workshops, and risk breakdown structure are used in both phases.

Identification tools	analysis tools
red flag items risk checklists assumption analysis Crawford slip method SWOT analysis risk register	percentage contingency identified contingency Monte Carlo analysis three-point estimates risk priority ranking P <i>x</i> I matrix risk comparison matrix risk map risk management information system self modeling worksheet
combined identification & analysis tools	
expert interviews risk management plan risk workshops risk breakdown structure	

2.4.4 Cures: Risk Identification

Risk identification is the first phase of the risk strategy, one of the eight strategies to counteract risk factors. The tools available to identify risks are presented here.

Risk identification, the first phase of the risk management strategy (Anderson et al. 2007, Molenaar et al. 2010), consists in identifying and documenting risks that could affect the project. It includes the examination of the project description, the design, cost estimates, construction schedules, the work breakdown structure, and others. It typically delivers a list of risks that is used in subsequent phases in the risk management strategy (analysis, mitigation, allocation, and monitoring).

Several tools have been developed by transportation agencies and project management entities to identify potential risks in projects (Molenaar et al. 2010). Some of them can be used not only during risk identification but also during risk analysis (Table 2.5). The risk identification tools are:

- Red flag items: red flag items are items critical in terms of cost and schedule impacts. They are listed, or red flagged, at the earliest stage of project development and updated at major milestones and anytime new items are identified. The two main goals of red flag items are setting contingencies and controlling cost escalation, and facilitate the communication between designers and cost/schedule estimators. Typically developed by interdisciplinary work groups in brainstorming sessions, red flag items do not involve any formal qualitative or quantitative risk analysis of the items but remind the project team of their existence.
- Risk checklists: risk checklists are lists of risks that have occurred or have been identified in past projects. They pursue two goals: first, the transfer of risk knowledge from past projects and past team members; second, they ensure that common risks are not overlooked. Risk checklists should be used after some form of risk brainstorming (e.g. red flag items) for two reasons: they might prevent from thinking out of the box, and they may not contain crucial risks specific to the project. During the development of the project, they should be reviewed. The California Transportation Department (Caltrans) has published a quite comprehensive risk checklist, subdivided into technical, external, environmental, organizational, project management, right-of-way, construction, and regulatory risks (Caltrans 2007).
- Assumption analysis: assumption analysis documents assumptions as potential risks, whose impact on the project cost and time is analyzed. It has the advantage of bringing the attention of designers and estimators to the assumptions and of being an alternative way to brainstorm risks. Examples of potential risks are assuming that environmental regulations do not change, calculating project costs in current dollars disregarding inflation and energy cost, or assuming the presence of good ground on the construction site.
- Expert interviews: experts are individual qualified in their field that can make reasonable subjective assessments on costs and schedule of the project. Their experience is used to identify risks that are not initially apparent. Besides identifying risks, they can indicate qualitatively and quantitatively the probability of occurrence and the impact of risks. It is

sensible to conduct expert interviews in risk identification, risk analysis and generally when additional input is needed.

- Crawford slip method: in the Crawford slip method, each participants of a group writes down one risk each minute for ten minutes. The advantage of such a rapid brainstorming session is obtaining the participant's opinion independently in a short period of time, while the disadvantages are the large number of risks generated (for a group of ten participants, 100 risks are generated including duplicates) and the fact that such a procedure cannot deliver a comprehensive list of risks. The Crawford slip method can make sense at the beginning of the risk identification step.
- SWOT analysis: a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis is used in the most different fields to develop strategic plans. In construction projects, it is a useful tool to identify uncertainties related to threats and opportunities. Used early in the risk identification process, it can be the starting point for the brainstorming as well as an additional tool to provide a comprehensive picture of the identified risks.
- Risk management plan: a risk management plan is a detailed plan of action for the management of risks. It consists of the risk strategy, including the methods to be used and the resources to be allocated, and the five steps of the risk management strategy (compare above): identification, analysis, mitigation, allocation and monitoring. It pursues three goals: formalizing the risk management process, providing guidance during the project, and serving as a communication tool between project team members. Differently from other risk identification tools, it is appropriate for small projects and for large, complex projects, the latter requiring a higher level of details. As mentioned above, it is used during identification, as well as during analysis, mitigation, allocation and monitoring of risks.
- Risk workshops: risk workshops are formal meetings between designers, estimators, experts, facilitators, and possibly stakeholders who are convened to identify and analyze risks. They aim at aligning team members in the understanding of the risks by identifying, ranking, and quantifying risks. Their output can include the description of project risks, the quantification of probability of occurrence and impact, ranges of project cost and time to calculate contingencies (section 2.4.5), preliminary risk registers (see next risk identification tool), and/or a

risk management plan (see previous risk identification tool). They are beneficial in the risk identification phase as well as in the other four steps of the risk management strategy. The Washington State Department of Transportation (WSDOT) has developed risk workshops to one of the central pieces of their cost estimation process: the CEVP (Cost Estimate Validation Process) workshop (more details in section 2.4.7). Applied for the first time to 12 projects in 2003, the CEVP workshop allows the WSDOT engineers to identify and analyze project risks with the help of team engineers, risk managers and experts. A more synthetic version (CRA: Cost Risk Assessment) of it has been developed to address risk in smaller, simpler projects (WSDOT 2005).

- Risk register: a risk register is a document that describes comprehensively all identified risks including risk triggers, probability of occurrence and severity, overall risk rating, responses, resources allocated, and current status. It should be understood and maintained as a living document recording the evolution of risks throughout the project. Thus, not only it is a risk identification tool but also a tool to analyze, mitigate and monitor risks. Similarly to the risk management plan, it is scalable by varying the level of detail from minimal for small, simple projects to comprehensive for large, complex projects. Its main goal is communicating project risks and their status and development during the entire project.
- Risk breakdown structure: a risk breakdown structure visually illustrates the interrelations between different risks belonging to different aspects of a project. It is particularly appropriate for complex projects (Figure 2-4).

Some of the tools presented are also used in the risk analysis phase (section 2.4.5) of the risk management strategy, specifically: expert interviews, risk management plan, risk workshops, and risk register.

The tools used in the first phase (risk identification) of the risk strategy have been presented. The tools used in the second phase (risk analysis) of the risk strategy are discussed in the next section.

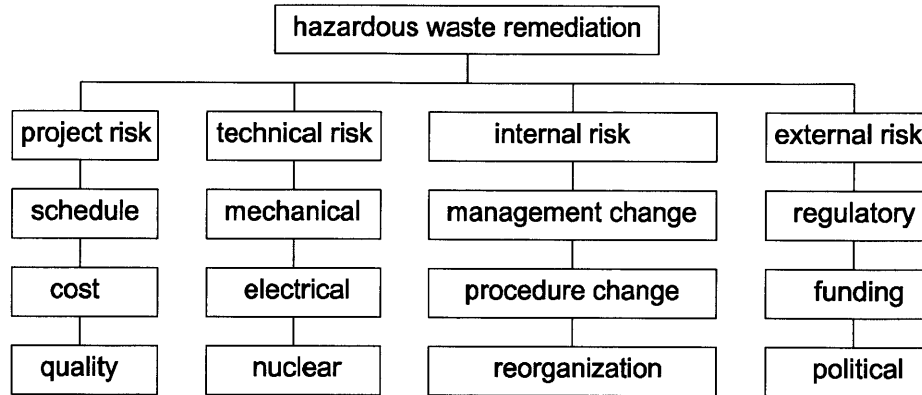


Figure 2-4: The risk breakdown structure is a tool to identify risks. In the figure four types of risk are identified in the case of a hazardous waste remediation project: project, technical, internal and external risks. Examples of technical risk can be mechanical, electrical and nuclear (after Molenaar et al., 2010).

2.4.5 Cures: Risk Analysis

Risk analysis is the second phase of the risk strategy, one of the eight strategies to counteract risk factors. The objectives, the two types of risk analysis (qualitative/quantitative) and the tools available to analyze risks are presented here.

Risk analysis, the second phase of the risk management strategy (Molenaar et al. 2010), is the process of evaluating the project risks documented in the risk identification phase of the risk management strategy. Its objective is to systematically consider risks, their probability of occurrence (frequency) and the consequences of their occurrence (severity). Risk analysis tools are scalable in that they can be used to prioritize red flag items of a relatively simple project as well as estimate probabilistically the cost of complex projects.

There are two types of risk analysis, qualitative and quantitative (Fell et al. 2005), that ideally are combined into a comprehensive risk analysis. While, qualitative tools are used to prioritize risks and to decide mitigation strategies and risk allocation, quantitative tools allow one to calculate the risk exposure of a project.

Several tools are available to analyze qualitatively and/or quantitatively the risk of a project. Expert interviews, risk management plan, risk workshops and risk register are used in both the risk identification phase (section 2.4.4) and the risk analysis phase, while others are specific to the risk

analysis phase of the risk management strategy (Table 2.5). The risk analysis tools are (Molenaar et al. 2010):

- **Percentage contingency:** a contingency is a percentage of the cost estimate to account for unforeseen, unidentified costs. Assuming that the total cost of a project is given by the sum of 1) known and quantifiable costs, 2) known but not quantified costs, and 3) unknown costs, the contingency is set to cover known but not quantified costs and unknown costs. The contingency is set according to historic data from past projects and decreases as the project advances. It should always be accompanied by a description of which risks it is intended to cover, e.g. project scope changes, underestimation of project costs, or an optimistic projection of the rate of inflation. Although it is the most prevalent risk analysis tool used by transportation agencies, it should be utilized only when more sophisticated risk analysis tools are unavailable.
- **Identified contingency:** in this risk analysis tool, the contingency is calculated as the sum of the percentage contingency (see previous tool) plus a contingency based on risks identified through a qualitative risk analysis. The risk estimator must provide justification why the identified risks are not captured by the percentage contingency. An alternative implementation of the identified contingency consists of setting the range of the percentage contingency from historical data and then estimating the expected costs of the top 20% of the prioritized risks. If the range of the percentage contingency is exceeded, additional contingency is warranted.
- **Monte Carlo analysis:** Monte Carlo analysis is a simulation tool that generates samples from probability distributions (Figure 2-5a). By simulating a project repetitively, it generates probability distributions for total cost and total time (Figure 2-5b), and tornado diagrams (graphic depictions of the impact of risks on the project total cost and total time). It has several important advantages: it visualizes the project uncertainties, it gives insight into which risks have the greatest impact, and it can be used to generate range estimates and to calculate contingencies. On the other hand, it has several requirements: it requires trained professional to be performed and all risks must be described quantitatively with probability distributions. These must be defined in terms of distribution type (triangular, lognormal, etc.) and distribution parameters (e.g. mean, standard deviation). The Monte Carlo analysis has

been used successfully in the CEVP framework for the risk management of complex projects (WSDOT 2005).

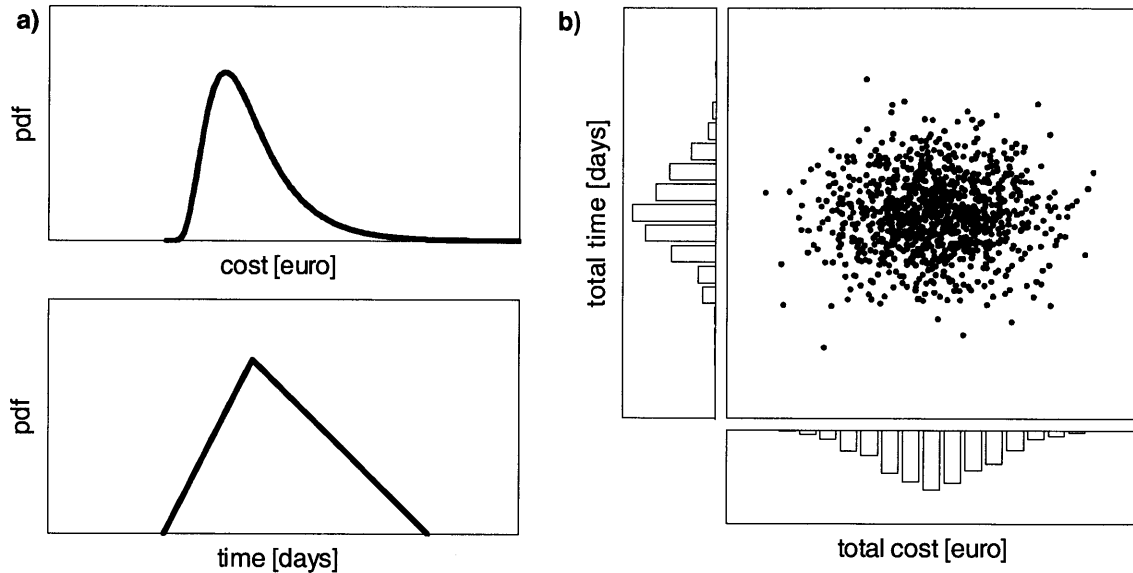


Figure 2-5: Monte Carlo simulation: a) input cost and time distributions, and b) output cost-time scattergram and cost and time histograms. The Monte Carlo simulation generates samples from the input distributions of the cost and time parameters, and sums the generated costs and times to calculate the total cost and the total time for each simulation run. The scattergram shows 1,000 simulation runs.

- Three-point estimates: Three-point estimates is a type of Monte Carlo simulation where each risk is evaluated in terms of a three-point estimate: the optimistic, the most-likely and the pessimistic estimate (Figure 2-6). The optimistic estimate is the lowest possible cost if everything works well, the pessimistic estimate is the largest possible cost if everything goes wrong, and the most-likely estimate is the expert's best guess. It is assumed that the mean is larger than the most-likely estimate, i.e. the distribution is skewed to the right. Three-point estimates are widely used to quantitatively estimate risks in transportation projects (Caltrans 2007).
- Risk priority ranking: risk priority ranking allows one to rank risks based on qualitative or quantitative risk analysis or experts' judgement. It aims at using the resources available

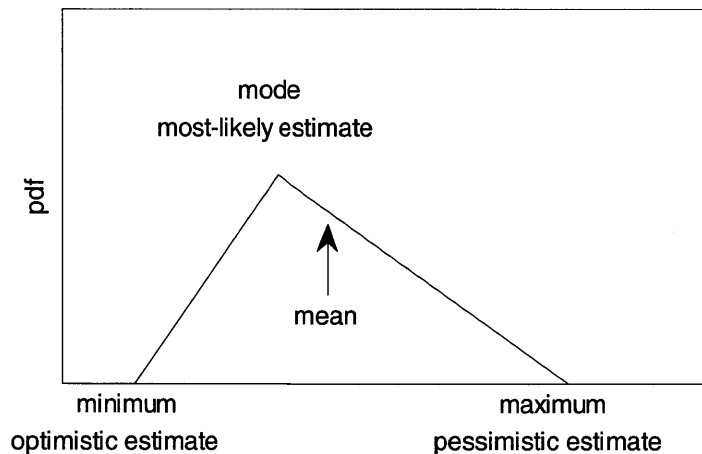


Figure 2-6: Three-point estimates: the probability distribution is assumed triangular, where the minimum is equal to the optimistic estimate, the maximum is equal to the pessimistic estimate, and the mode is equal to the most-likely estimate.

efficiently for planning and mitigation and it facilitates the communication between designers and risk managers. The choice of analysis, qualitative or quantitative, depends on the project complexity: while a qualitative approach (PxI matrix or risk map, see next risk analysis tools) is sufficient to allocate scarce resources, a Monte Carlo analysis (see previous risk analysis tool) may be required to rank risks to determine the contingency.

- PxI matrix (Probability \times Impact (severity) matrix): a PxI combines the qualitative evaluation of the probability of occurrence and of the impacts of a risk in three steps. First, a probability ranking table is set up to match a probability of occurrence to a ranking number. Second, impact tables for every risk are formulated in order to relate this to the objectives of the project, such as cost and time. Last, from the combination of the probability ranking table and the impact table, the PxI matrix is created. If more than one project objective is taken into consideration (e.g. cost and time), there is a PxI matrix for each objective. Depending on the combination of probability and impact, risks fall into the red, yellow or green areas of the PxI matrix (Figure 2-7). Risks in the red area, large probability/moderate impact or moderate probability/large impact, cause unacceptable major disruptions, thus warranting risk management attention. Risks in the yellow area cause some disruptions and are distinguished into two groups: 1) large probability/low impact risks should be pooled

together to model their combined effect, whilst 2) low probability/large impact require individualized scrutiny. Risks in the green area are low probability/low impact. Probability of occurrence and impact of each risk are analyzed by an engineer or an expert. The goal of a P*x*I matrix is twofold: prioritize risks to efficiently allocate resources and identify in a initial analysis the risks that require further quantitative analysis.

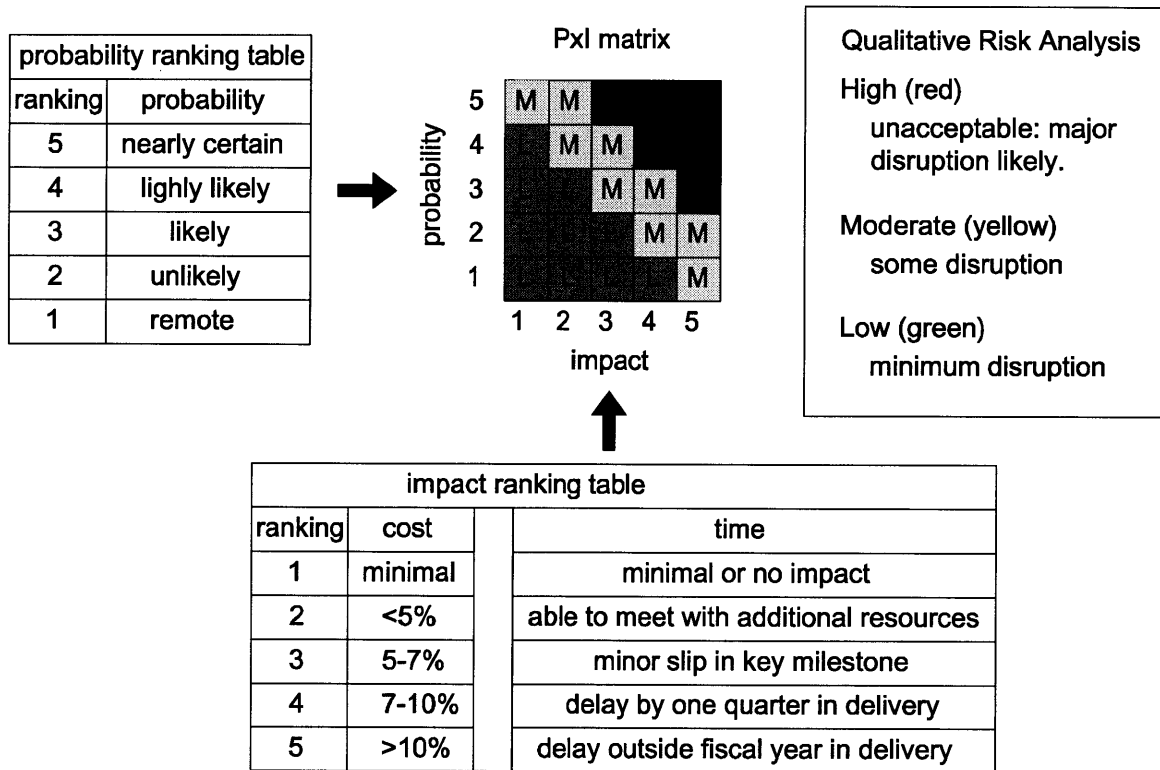


Figure 2-7: A P*x*I matrix (Probability *x* Impact matrix) is a tool for qualitative risk analysis: first, the probability of occurrence of a risk is matched to a ranking number (probability ranking table); second, the impact of a risk on the project objective is also matched to a ranking number (impact ranking table); finally, according to the probability ranking number and the impact ranking number, a risk falls into the red (H-high), the yellow (M-moderate), or the green (L-low) area. Depending on this qualitative risk analysis result, the risk is warranted risk management attention.

- Risk comparison table: a risk comparison table is a tool to compare risks to one another. The project team discusses two risks at a time and votes on their relative importance. Two at the time, all risks are compared. Being a prioritization tool, this specifically aims at an efficient resources allocation.

- Risk map: a risk map is a probability/impact diagram on which the risks are positioned. It is a powerful communication tool that visualizes probability and consequences of a risk relative to other risks (Figure 2-8). It is also used to track changes over time: for example, the comparison of the probability/impact diagrams before and after mitigation shows the positive effect of mitigation measures.

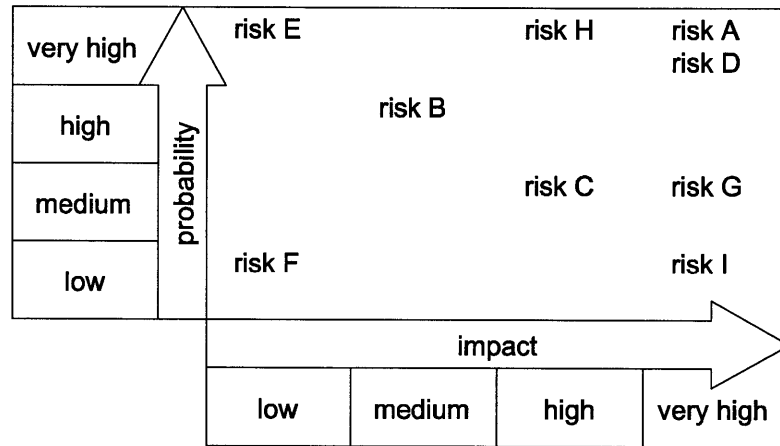


Figure 2-8: A risk map is a visualization of the probability and impact of the risks in a project. It allows one to visually compare more than one risk.

- Risk management information system: This tool is an interactive data management system to document and monitor risks through the project development. It can also be used to generate standard reports such as the 20 top risks by cost, delay or other objectives. Since it requires significant information technology, a less sophisticated option is the use of the risk register (section 2.4.4).
- Self modeling worksheet: a self modeling worksheet is an Excel spreadsheet to quantitatively analyze risks. It works as a Monte Carlo simulation and three-point estimates to calculate the impact of risks on a project. Upon inputting the three point estimates for each variable in the spreadsheet cells, it runs a simulation and provides the distributions of project cost and time based on the input. It is less complex than risk analysis software and thus more accessible for project teams. It is being successfully used for \$10 – 15 million projects at the Washington State Department of Transportation.

The objectives, the types (qualitative/quantitative) and the tools of the risk analysis (second phase of the risk strategy) have been presented. The last three phases of the risk strategy are discussed in the next section.

2.4.6 Cures: Other Phases of the Risk Strategy

Recall that a risk strategy includes five phases: identification, analysis, mitigation and planning, allocation, and monitoring and controlling. In this section, the third, the fourth and the fifth phase of the risk strategy are briefly presented.

The third phase of the risk strategy, risk mitigation and planning, explores response strategies for the risks that have been prioritized in the qualitative or quantitative risk analysis. Risk mitigation identifies which risk response strategy is best suited for each risk and designs specific actions to implement the selected risk response strategy. There are four main risk response strategies (Caltrans 2007):

1. Avoidance: the project plan is changed to eliminate a risk. This can be achieved by changing the scope, adding time, or adding resources.
2. Transference: the financial impact of the risk is transferred by subcontracting part of the work. This is effective only if the subcontractor is more capable of reducing the risk and does so.
3. Mitigation: the probability of occurrence or the consequences of a risk are reduced to an acceptable threshold.
4. Acceptance: certain risks are accepted.

Risk planning is the development, implementation and monitoring of the risk response strategies. It involves the detailed formulation of a plan of action, in the form of a risk register for simple projects and a formal risk management plan for a complex project.

The fourth phase of the risk strategy, risk allocation, is the process of identifying and assigning the responsibility of a risk and a risk response to a party. This could be an agency planner, an engineer, a construction manager, or a private sector contractor. Risk allocation should follow four principles (Molenaar et al. 2010):

1. Allocate a risk to the party best able to manage it. For instance, the risk of a breakdown in equipment is best borne by the contractor, while the risk of securing project funds is best borne by the agency.
2. Allocate a risk in alignment with project goals (it requires that project objectives are clearly defined). For example, if the public needs a project completed sooner, the agency can ask the contractor to assume more risk for expedited completions and it must compensate the contractor for assuming this risk.
3. Share a risk when appropriate to accomplish project goals. Despite the misleading terminology, risk sharing is the definition of the point where risk is transferred from one party to another. For instance, a contract may grant the contractor a time extension for severe weather but may not provide additional compensation of costs.
4. Allocate a risk to promote alignment with customer-oriented performance goals. Examples of customer-oriented performance goals are the client satisfaction with the service, the predictability of time, or safety. Aligning to such goals can be achieved with alternative contracting techniques. For example, the contracting technique "(cost + time) procurement" is used to pass the risk for accurately setting the fastest construction completion date from the agency to the contractor.

The fifth phase of the risk strategy, risk monitoring and controlling, pursues four objectives (Molenaar et al. 2010): first, systematically track the identified risks; second, identify new risks; third, effectively manage the contingency reserve; and fourth, capture the lessons learnt for future risk management. Risk monitoring and controlling should continue throughout the project: in periodic risk reviews the identification, the analysis, the mitigation and the allocation phases are repeated. Specifically, risk monitoring and controlling have three tasks (Molenaar et al. 2010):

1. Develop comprehensive reporting procedures.
2. Monitor risk and contingency reserves.
3. Provide feedback for future risk management.

The last three phases (risk mitigation and planning, risk allocation, and risk monitoring and controlling) of the risk strategy have been presented. In the next section, the most advanced tool pursuing the risk strategy is discussed.

2.4.7 A Specific Cure: Cost Estimate Validation Process

The Cost Estimate Validation Process (CEVP) is the most advanced tool pursuing the risk strategy. It is a tool especially suited for risk identification and risk analysis (first and second phases of the risk strategy). In this section, the goals and the two central processes, the validation process and the risk identification process, of the CEVP are discussed.

WSDOT (Washington State Department of Transportation) has developed the Cost Estimate Validation Process (CEVP) and the CRA (Cost Risk Assessment) to identify, quantitatively analyze and evaluate risk that could impact cost and/or time during project development. CEVP and CRA are similar processes but they differ in the degree of detail of the risk analysis: if the estimated cost of a project is more than \$100 million, a CEVP is required while for costs between \$5 million and \$100 million a CRA is carried out (WSDOT 2005).

CEVP pursues several goals (WSDOT 2008). The main goal is estimating ranges of project cost and time by modeling uncertainty in the positive and negative forms of opportunities and threats, according to the concept that it is better to be approximately right rather than precisely wrong (WSDOT 2008), thus favoring ranges over single number estimates. A second goal is to establish consistency in the practice of risk-based estimations. Since 2002, CEVP has been successfully implemented at WSDOT and it has been adopted by federal agencies and state agencies nationwide (Reilly et al. 2004). Another goal is providing a flexible and scalable tool that can be adapted depending on size, location and complexity of the project. The existence of two versions of the same estimation tool, CEVP and CRA, offers some scalability while the steps of the estimation (see below) guarantee great flexibility.

A brief description of CEVP will highlight the two central elements, base cost and risk cost, and the two central processes, the validation process and the risk identification process. In CEVP, cost estimates are given by the base cost and the risk cost. The base cost is the most probable cost that can be expected in the case the project develops as planned. It is estimated during the validation process by eliminating contingencies from the initial cost estimate. The risk cost is given

by risk events, defined by their probability of occurrence and their impact. It is determined and quantified during the risk identification process (WSDOT 2008).

Project team and CEVP specialists conduct the cost validation process: starting from an initial cost estimate they determine the base cost. First, the project team briefs the CEVP specialist on the project scope and the risks included in the project estimate. Then, successively the project scope, cost and time are reviewed. Unit prices and production rates are also reviewed; in particular, the contingencies in the unit prices are removed. Finally, an agreed upon base cost of the project is calculated, which is the base to which the risk cost is added (Reilly 2005).

In the risk identification, project team members and experts identify risks led by an experienced risk analyst who is acquainted with uncertainty theory, debiasing techniques, and the cost and risk models. The composition of the team participating in risk identification should strike a balance between project knowledge, risk analysis expertise, cost estimating experience, and objectivity. The calibration process requires three pieces of input: a graphical display of the project plan and strategy (e.g. in a flowchart), a preliminary list of risks and opportunities, and the base cost from the validation process. Ideally before starting the formal risk identification process, the participants should be calibrated, i.e. trained albeit briefly, on relevant risk concepts and biases. The goal of the risk identification process is to 1) identify, 2) quantify and 3) model the uncertainty of project cost and time. Starting from the preliminary list, potential risks and opportunities are identified in a brainstorming session, and if needed they are prioritized with a screening process, such as a qualitative risk analysis. Then, they are quantified in terms of probability of occurrence and cost/time impact through a combination of empirical data and subjective judgment. Independence and correlation, positive or negative, between risks is also defined; however, the participants should aim at defining risks to be independent as far as possible, since independence greatly simplifies the modeling (Reilly 2005). The risk identification process is a type of risk workshop, as explained in section 2.4.4.

Following the validation and the risk identification processes, the base cost and risk cost are integrated in a probabilistic model and quantitatively analyzed with a Monte Carlo procedure, which delivers probability distributions of project cost and time. Such distributions are used to communicate ranges of probable cost and time to the public (Reilly 2005). It should be noted that the quantitative analysis as well as the risk identification processes are iterative, since some risks

are mitigated or eliminated and others arise. The quantitative analysis of CEVP lies within the class of Monte Carlo analyses described in section 2.4.5.

A central advancement of CEVP is that, since one cannot predict the future accurately, a range of project cost and time is a more sensible representation of uncertain outcomes. However, a practical downside of ranges is that for budget planning purposes a single number estimate is usually needed in order to gain legislature endorsement. The practice at WSDOT has been to indicate the 90th percentile of the probability distribution for approval and budgeting purposes. Additionally, results presented as ranges can cause communication difficulties: the project can appear too expensive and unrealistic if the decision makers focus on the extreme values of the ranges (Reilly et al. 2004). However, this is not always the case: in fact, when the budget for the Lötschberg tunnel (Swiss transalpine tunnels) was allocated, it included a range for the construction costs of $\pm 15\%$ (Teuscher 2007). When CEVP was introduced at WSDOT in 2002, an initial negative reaction of the public turned into acceptance and gratitude for more realistic estimates: within a short period of time both the larger cost estimates and the concept that ranges are a more sensible representation of cost estimates than single numbers were accepted by the public (Reilly et al. 2004).

The successful use of CEVP depends on one central conditions: the owner is interested in knowing the objective potential cost (Reilly et al. 2004). This observation is tightly connected to the body of literature that identified the causes of cost escalation in inadequate estimation tools, optimism bias and economic-political interests (section 2.3). CEVP is an advanced tool to estimate uncertainty in cost and time; it can address biases in the risk identification process with the help of the risk analyst but it cannot address politically motivated cost underestimations that aim at obtaining approval for a project by accentuating benefits and reducing costs. In order to guarantee the professionalism in cost and time estimates within CEVP, WSDOT binds the participants to a code of ethics in ten points that cover issues such as the highest standards of practice in the industry, honest and effective communication, accountability, broad participation in the process without exercising pressure when developing estimates, consideration for public funds, strengthening the understanding of risk and of cost/time estimation (WSDOT 2008).

Transportation agencies abroad have developed risk identification and analysis tools similar to CEVP: it is the case for HARM (Highways Agency Risk Management) used at the Highways Agency in England and the Public Sector Comparator developed by the PPP (public-private partnership)

Knowledge Centre in the Netherlands. In both tools, risks are first identified and prioritized, and then they are analyzed with quantitative risk analysis (Monte Carlo simulations). In particular, HARM aims at creating a full-fledged risk management including, besides risk identification and analysis, also risk mitigation and risk allocation. Differently, the special feature of the Dutch Public Sector Comparator is the capability of comparing total project cost including the entire project life cycle (FHWA 2005).

The CEVP, with its validation process and risk identification process, is the most advanced tool pursuing the risk strategy. This is one of eight strategies proposed by the transportation community to counteract the risk factors causing cost underestimation in infrastructure construction projects.

2.5 Conclusions

Cost and time underestimations are widespread in infrastructure construction projects. Flyvbjerg and the transportation community have identified different causes and proposed different cures for these underestimations. Flyvbjerg identified technical, psychological and economic-political causes for which he encouraged the use of improved estimation techniques and a change in policy. Differently, the transportation community identified risk factors of different nature and proposed eight strategies to counteract these. One of the proposed strategies is the risk strategy, whose tools aim at identifying and qualitatively/quantitatively analyzing the risks.

The proposed cures to counteract cost and time underestimation show two limitations: First, estimation techniques and risk modeling tools without an in-depth understanding of the construction process do not ensure that all uncertainties in the construction process are captured. Second, modeling risks only does not capture the cumulative effect of different sources of uncertainty.

The research work presented in this thesis provides a construction model and an uncertainty model. The construction model deepens the understanding of the construction process of linear/networked infrastructure projects and gives insight when identifying sources of uncertainty. The uncertainty model deepens the understanding of uncertainty in the construction process, models the sources of uncertainty and captures their cumulative impact on the construction cost and time of linear/networked infrastructure projects. The construction model is presented in chapter 3 and the uncertainty model in chapters 4 and 5. The effectiveness of the construction model and the uncertainty model are brought to light with their application to a rail line project in chapters

6 and 7.

Chapter 3

Construction Model for Rail Lines

Rail line construction projects are often plagued by cost and time underestimations. The available estimation tools do not attempt to understand the construction process of rail lines before quantifying the risk. The author proposes an estimation method consisting of two components: the construction model and the uncertainty model. This chapter presents the construction model, which is later implemented in the simulation tool DAT (Decision Aids for Tunneling) in chapter 6.

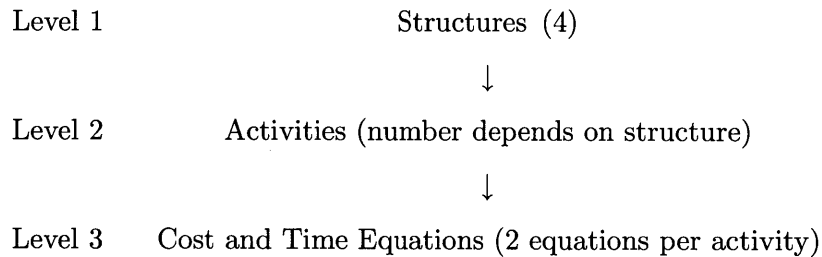
The construction model analyzes at the level of the single activities the construction processes of the four main types of structures in rail lines: tunnels, viaducts, cuts and embankments. It develops activity networks representing the specific construction processes and calculates the construction cost and time of rail lines. This chapter presents the activity networks of the structures, the cost and time equations of the activities, and the material handling model for cuts and embankments. In the construction model, the activity networks consist of sub-networks that are repeated numerous times during the construction of a structure. Thus, the interconnection of sub-networks models one structure and the interconnection of the different structure networks model the construction of the rail line. The representation of the construction process of rail lines with activity networks deepens the understanding of the construction process and provides the starting point to identify the sources of uncertainty (chapter 4).

The detailed structure of the chapter is as follows: first, the expressions to calculate the construction cost and time of a rail line are presented (section 3.1). Second, activity networks are introduced (section 3.2), followed by the activity networks specifically developed to represent the construction of tunnels (section 3.2.1), viaducts (section 3.2.2), and cuts and embankments (sec-

tion 3.2.3). Last, material handling schemes in cuts and embankments construction are discussed (section 3.3).

3.1 Construction Cost and Time

The construction of a rail line can be modeled as the construction of a sequence of four main structures: tunnels, viaducts, cuts and embankments. Every structure is modeled as a sequence of activities (see next sections), which are characterized by cost and time equations. A hierarchy can be identified:



Assuming the contract structure in rail line construction projects is unit cost based (note that some costs, such as overhead costs, are not unit cost based), the total construction cost is given by the sum of the costs of tunnel activities, $t = 1, \dots, n_t$, the costs of viaduct activities, $v = 1, \dots, n_v$, the costs of cut activities, $c = 1, \dots, n_c$, the costs of embankment activities, $e = 1, \dots, n_e$:

$$TotalCost = \sum_{t=1}^{n_t} Cost_t + \sum_{v=1}^{n_v} Cost_v + \sum_{c=1}^{n_c} Cost_c + \sum_{e=1}^{n_e} Cost_e. \quad (3.1)$$

Differently from the total cost, the total time is not the sum of all activities time because some activities are performed in parallel, so that the total construction time of a rail line is given by the sum of the activities, $j = 1, \dots, m$, on the critical path:

$$TotalTime = \sum_{j=1}^m Time_j. \quad (3.2)$$

Note that the word "time" is used to describe the concept of construction schedule duration throughout this thesis, since time is the word used in the simulation tool DAT (chapter 6).

In this section, the expressions to calculate the construction cost and time of rail lines have

been introduced. They are based on the activity costs and times, which are presented in the next sections.

3.2 Activity Networks

Activity networks are the instrument to model the construction of the structures (tunnels, viaducts, cuts and embankments) of the rail line. In this section, the building blocks of activity networks (arcs, nodes, dummies) and their use are explained.

Activity networks are used to model complex processes comprising many activities. These are modeled with arcs, the connection between nodes. They are connected in series and in parallel. In the series connection, an activity must wait for the preceding activity to be completed before being activated, while in the parallel connection activities are activated simultaneously (Figure 3-1a).

Nodes differentiate into OR and AND nodes depending on the activation of the activity following the node. The OR node activates the following activity as soon as any of the preceding activities is completed, while the AND node must wait for all preceding activities to be completed before activating the following activity (Figure 3-1b). While the OR node suffices to model activities in sequence (Figure 3-1a top), the AND node is useful to model parallel activities (Figure 3-1b bottom).

Dummy activities are used when OR and AND nodes are not sufficient to correctly model the logical sequence of activities in the process. They are activities that do not strictly contribute to the process but help modeling the logical sequence of activities. For example, in Figure 3-1c two activities in the bottom sequence depend on two activities in top sequence, which is modeled with two dummy activities connecting the end of the activities in the top sequence to the start of the activities in the bottom sequence. The two AND nodes indicate that the dummy activity and the preceding activity must be completed before the node activates the following activity.

Arcs, OR and AND nodes are the building blocks of activity networks that can model complex processes such as rail line construction. In different ways, they are used and combined into activity networks to model the construction of tunnels, viaducts, cuts and embankments, as detailed in the next sections.

In the upcoming sections, the activity networks modeling the construction of tunnels, viaducts, cuts and embankments are presented.

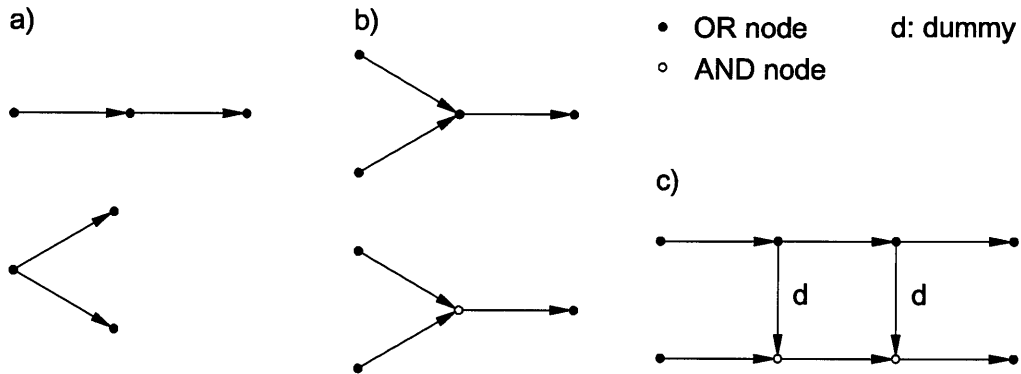


Figure 3-1: a) two activities in series (top) and in parallel (bottom). b) the OR node activates the following activity as soon as any of the preceding activities is completed (top), whereas the AND node activates the following activity after both preceding activities are completed (bottom). c) in the bottom sequence, the second and the third activities depend on the first and the second activities of the top sequence, respectively. This is modeled with two dummy activities connecting the end of the activities in the top sequence to the start of the activities in the bottom sequence.

3.2.1 Tunnels

The construction of a tunnel is modeled with a one-activity activity network. Cost and time equations of the activity are presented.

The construction of a tunnel includes the excavating and the supporting process, where these processes include several operations. For example, the excavating process with blasting includes operations such as drilling, loading, detonating, ventilating and mucking, while the supporting process includes operations such as shotcrete lining, concrete lining and tunnel finishing. The construction of a tunnel can be modeled at three different levels: the overall construction level, the processes level, or the operations level.

The construction model presented here reproduces the construction of a tunnel at the overall construction level, i.e. the tunnel construction is modeled with one activity that includes all operations of excavating and supporting a unit length of tunnel (Figure 3-2). The tunnel construction activity is characterized by cost and time equations, which quantify the cost and the time of constructing the tunnel as a function of cost per unit length, time per unit length, and tunnel length:

$$\text{cost} = (\text{cost per unit length}) \times \text{length}, \quad (3.3)$$

$$\text{time} = (\text{time per unit length}) \times \text{length}. \quad (3.4)$$

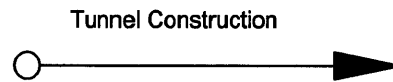


Figure 3-2: The construction of a tunnel is reproduced with an activity modeling the construction of one unit length of the tunnel. The cost and the time of the activity are calculated with equations 3.3 and 3.4.

Hence, the construction of tunnels is modelled with the simplest activity network: one activity. Differently, viaducts, cuts and embankments are modelled with more complex activity networks (next section).

3.2.2 Viaducts

The construction of a viaduct is modelled with an activity network, which depends on the construction method of the viaduct. Three construction methods, their activity networks, and the cost and time equations of the activities are presented.

The construction of a viaduct includes the construction of its elements, such as abutments, deck sections, piers, footings, pile sets, technical blocks (soil structure preceding and following a viaduct aiming at reducing settlements), and finishing (Figure 3-3). The construction model presented here models the construction of a viaduct at the elements level, where the construction of each element is modeled with one activity (Figure 3-4). This is characterized by cost and time equations, which quantify the cost and the time of constructing an element:

$$\text{cost} = (\text{element cost}), \quad (3.5)$$

$$\text{time} = (\text{element time}). \quad (3.6)$$

The activities modeling the construction of the elements are combined in an activity network,

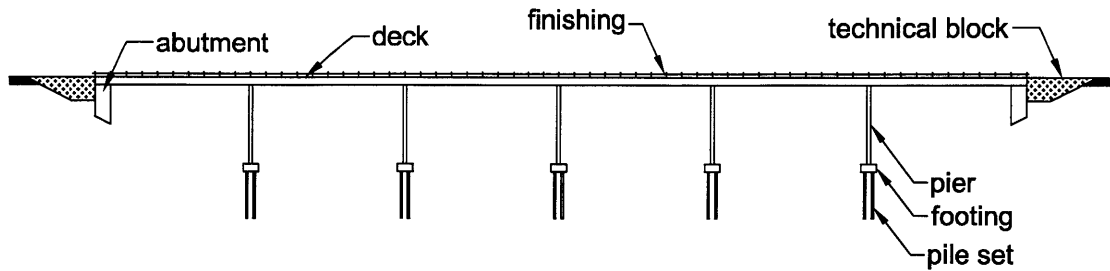


Figure 3-3: Elements of a viaduct. Each element is modeled with one activity characterized by its cost and time.

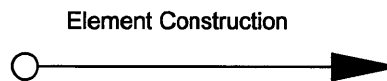


Figure 3-4: Activity modeling the construction of a viaduct element. The cost and the time of the activity are calculated with equations 3.5 and 3.6.

where the sequence of the activities depends on the construction method. There are three main methods to construct viaducts: balanced cantilever, span-by-span, and launching (O'Donnell 2008). These methods are explained in the following assuming construction starts from an end of the viaduct; however, the activity networks can easily be adapted to other construction starting points. In the balanced cantilever construction method (Figure 3-5a), the foundation pile set and footing are constructed, followed by the pier and the deck section. The deck section is constructed by attaching and post-tensioning precast concrete elements, or by casting-in-place and post-tensioning concrete elements on both sides of the pier, thus exploiting the cantilever action of the deck elements balancing each other on the two sides of the pier. In the span-by-span construction method (Figure 3-5b), a deck section between two piers is built after both pile sets, footing and piers are constructed. The deck section between piers is built with the aid of a gantry or formwork. The materials used are precast concrete (deck elements and girders), cast-in-place concrete (deck elements and girders), and steel girders plus deck elements (precast or cast-in-place concrete) placed on girders. In the launching construction method (Figure 3-5c), similarly to the span-by-span construction method a deck section between two piers is built after both pile sets, footings and piers, are constructed.

After the first pile set, footing and pier are built, a first deck section is pushed forward (launched) from the abutment towards the first pier. A second deck sections is launched by pushing forward the first deck section towards the second pier, and so on. Precast concrete, cast-in-place concrete and steel are used in launching.

In the balanced cantilever method, the construction of the deck section follows the construction of the pier, whereas in the span-by-span and the launching methods the construction of the deck follows the construction of the preceding and the following pier. Based on this observation, two activity networks are developed: an activity network modeling the balanced cantilever construction method (Figure 3-6), and an activity network modeling the span-by-span and the launching construction methods (Figure 3-7).

In the balanced cantilever activity network (Figure 3-6), the abutments and the piers' foundations, followed by the piers, can be started simultaneously. Starting from the first abutment, the deck sections are constructed sequentially. In the activity network, a repeating sub-network is identified, which models the construction of a unit including the foundation (pile set and footing), the pier, and the deck section, and a dummy activity connecting it to the next sub-network. For longer viaducts, the activity sub-network is repeated as many times as the number of viaduct units.

In the span-by-span and the launching activity network (Figure 3-7), the first abutment and the first pier's foundation (pile set and footing), followed by the pier, are constructed simultaneously. After the abutment and the first pier are completed, the deck sections are constructed sequentially in parallel with the sequential construction of foundations and piers. A repeating sub-network is identified also in this activity network. It models the construction of a unit including the foundation, the pier, the deck section preceding the pier, and a dummy activity connecting it to the next sub-network. For long viaducts, the activity sub-network is repeated as many times as the number of viaduct units, and the activities modeling the construction of abutments, technical blocks and finishing (Figure 3-3) are added at the beginning and at the end of the sequence of repeated sub-networks (arcs "A", "T" and "N", respectively, in Figures 3-6 and 3-7).

In the preceding activity networks representing different construction methods of viaducts have been illustrated. After presenting the activity networks of tunnels and viaducts, the activity network of a sequence of cuts and embankments is discussed in the next section.

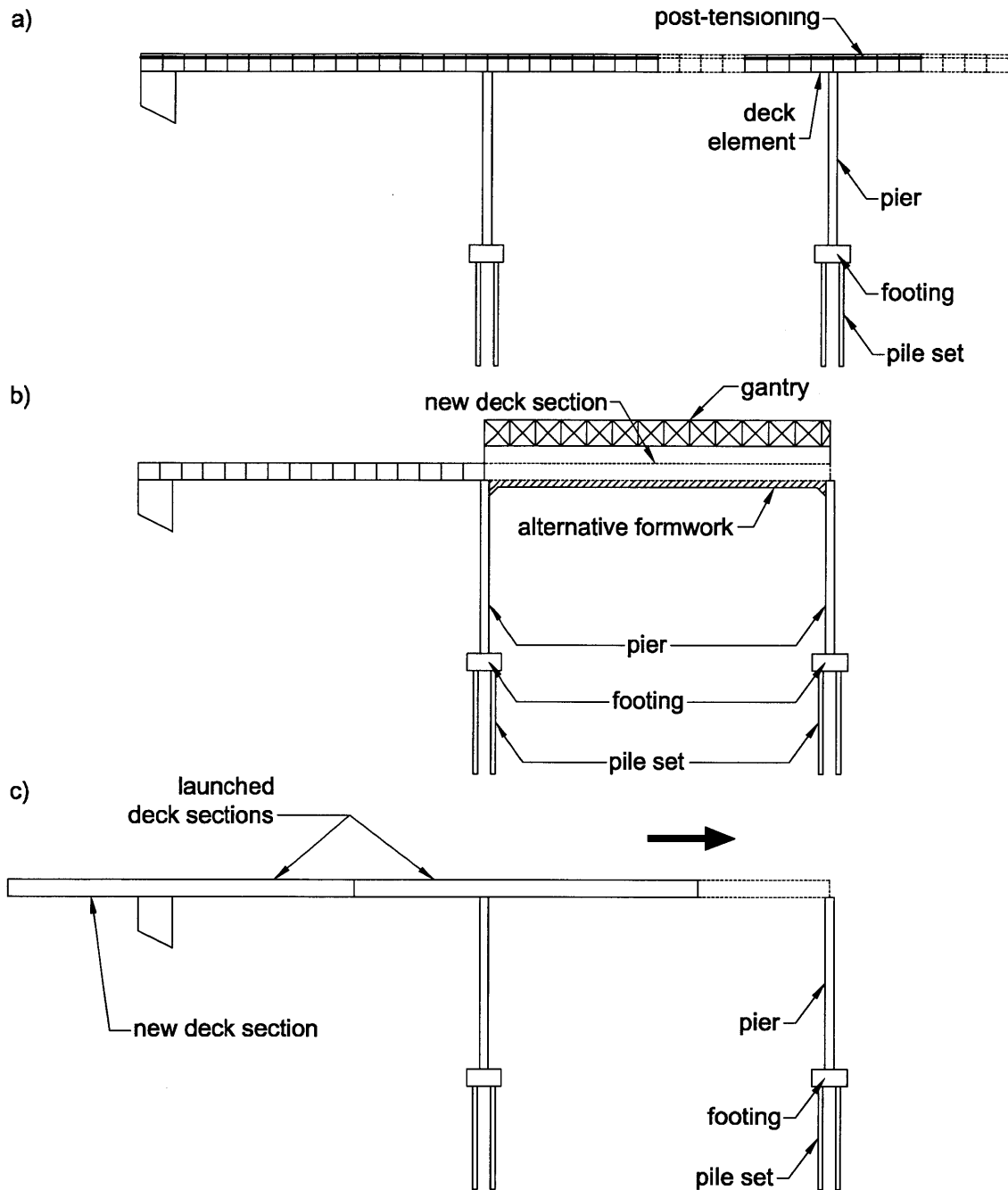


Figure 3-5: a) construction of a viaduct with the balanced cantilever method: the deck section is built after the pile set, the footings and the pier are constructed. b) construction of a viaduct with the span-by-span method: the deck section is built after the preceding and following pile sets, footing and piers are constructed. c) construction of a viaduct with the launching method: similarly to the span-by-span construction method, the deck section is built (launched) after the preceding and following pile sets, footings and piers are constructed.

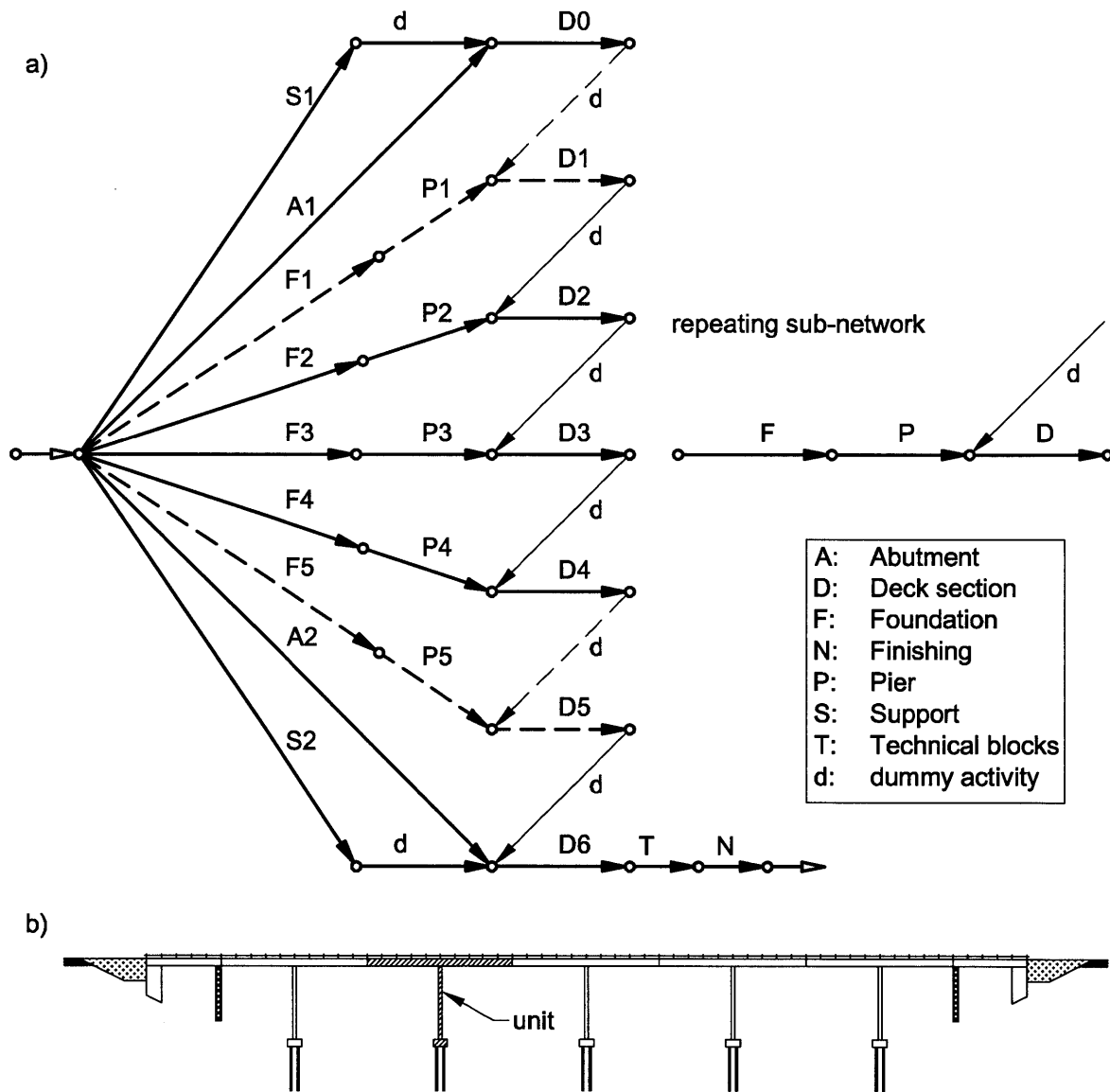


Figure 3-6: Balanced cantilever construction method: a) activity network and repeating sub-network of b) a six-span viaduct. The repeating sub-network models the construction of a unit (dashed in b) including the foundation (pile set and footing), the pier and the deck section. In the activity network the sub-network is repeated five times, the number of the units in the viaduct (the first and the fifth repetitions of the sub-network are dashed in a).

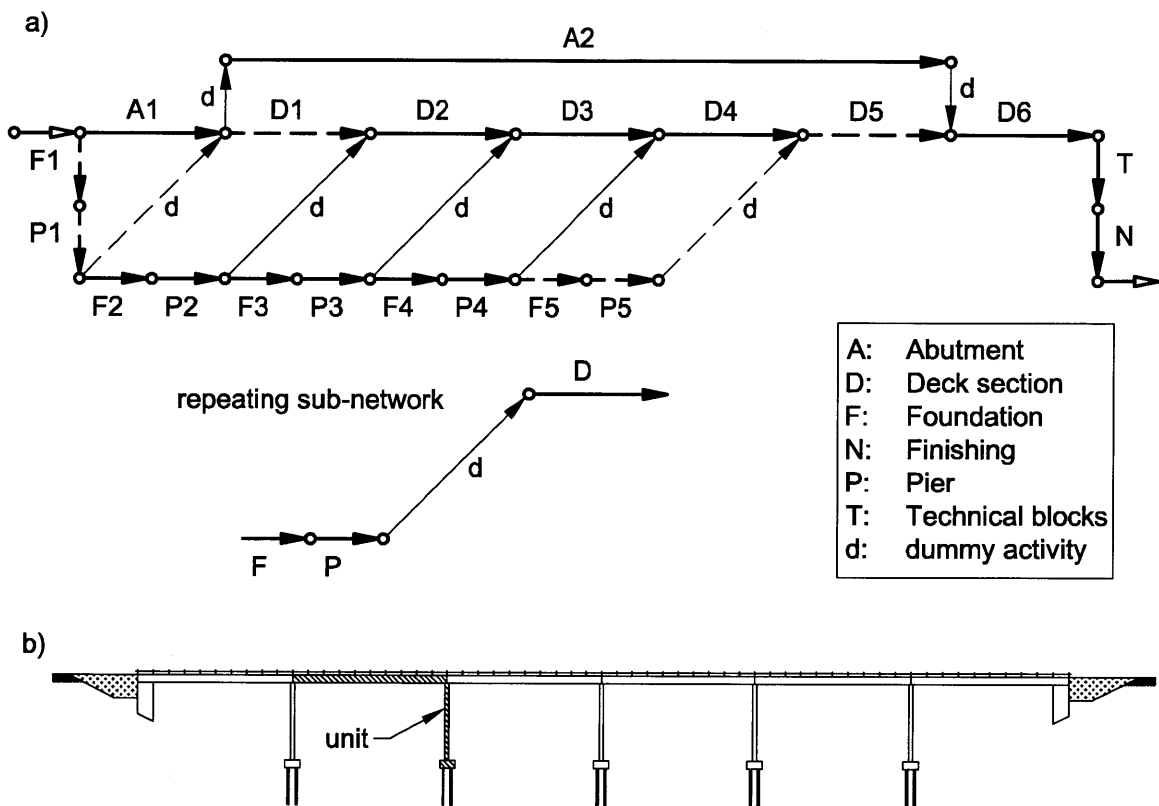


Figure 3-7: Span-by-span and launching construction methods: a) activity network and repeating sub-network of b) a six-span viaduct. The repeating sub-network models the construction of a unit (dashed in b) including the foundation (pile set and footing), the pier and the deck section. The sub-network is repeated five times, the number of the units in the viaduct (the first and the fifth repetitions of the sub-network are dashed in a).

3.2.3 Cuts and Embankments

The construction of cuts and embankments is modeled jointly with a single activity network. The cost and time equations of the activities are presented, and the construction processes determining the sequence of activities in the activity network are explained.

Cuts and embankments are modeled jointly as a sequence of the two types of structures since cuts and embankments share common processes, e.g. clearing the construction area and placing the sub-ballast, and a mass balance between excavated material in the cuts and filled material in the embankments is often sought. The construction of a cut includes processes (Figure 3-8a) such as

1. clearing the soil,
2. excavating the cut,
3. capping the cut, and
4. placing the sub-ballast,

while the construction of an embankment includes processes (Figure 3-8b) such as

1. clearing the soil
2. possibly improving the in situ material,
3. filling the embankment,
4. capping the embankment, and
5. placing the sub-ballast.

Note that improving the in situ material might be also required in cuts, although it does not occur often.

In rail line construction, these processes are followed by placing the ballast, placing the railroad ties, planing the rail and checking the track elevation (Figures 3-8), while in road construction, they are followed by placing the base course and then the concrete or asphalt road surface.

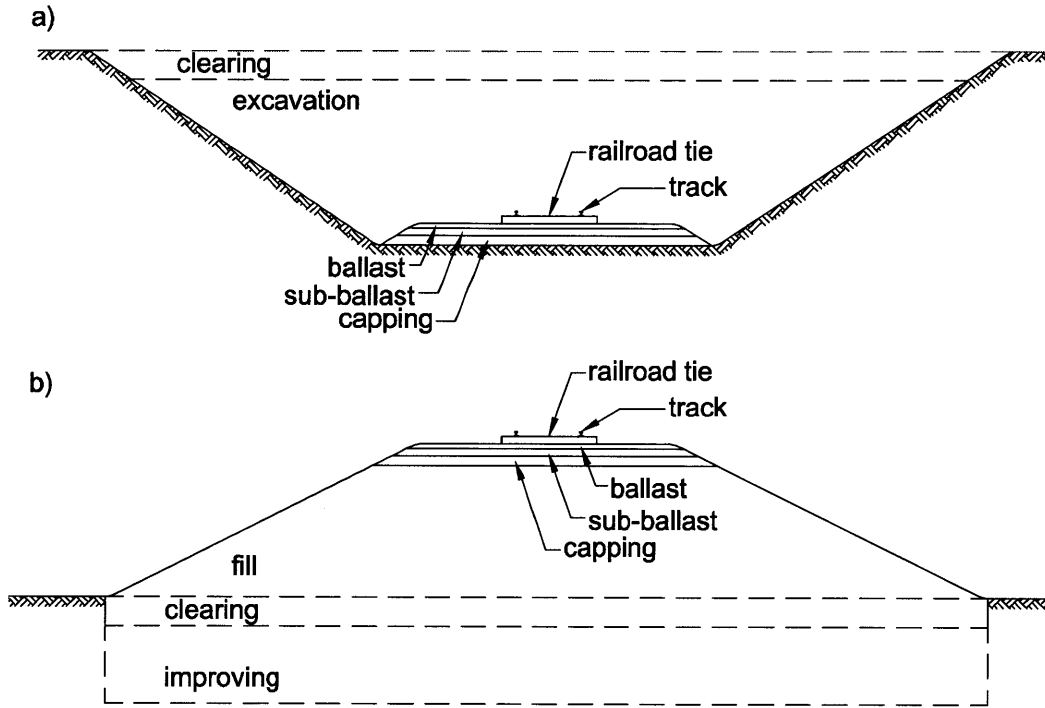


Figure 3-8: Cross-sections of a) a railroad cut and b) a railroad embankment. They show the construction processes: clearing the soil, placing the capping layer, placing the sub-ballast, the ballast, the railroad ties and the tracks are common processes, while excavating is a process in cut construction, and improving the in situ soil and filling the embankment are processes in embankment construction.

Cuts and embankments are constructed with parallel and sequential processes. The construction of a cut can be started from either end or from both ends, is advanced sequentially, possibly in parallel layers at different depths (Figure 3-9a). The construction of an embankment can be started from any point or any combination of points, also advances sequentially, possibly in parallel layers at different heights (Figure 3-9b). Depending on site access and equipment used, embankments are constructed sequentially not only longitudinally but also laterally (Figure 3-9c).

The construction model presented here includes cuts and embankments processes starting from the clearing until placing the sub-ballast (Figure 3-8). It assumes construction starts from one end but it can easily be adapted to other starting points. Each process is modeled with one activity (Figure 3-10), characterized by cost and time equations quantifying its cost and time as a function

of the cost per unit volume, the production rate, and the volume:

$$\text{cost} = (\text{cost per unit volume}) \times \text{volume}, \quad (3.7)$$

$$\text{time} = \frac{\text{volume}}{\text{production rate}}. \quad (3.8)$$

The activities modeling the construction of cuts and embankments are combined in an activity network, which is governed by two relationships. First, the (i) structure (cut or embankment) is constructed before the ($i + 1$) structure: in a sequence of two embankments and two cuts (Figure 3-11a), the first embankment is filled before the second embankment, and the first cut is excavated before the second cut (activity F1 precedes activity F2, E1 precedes E2 in Figure 3-11b). Second, the sequence of activities within a cut is clearing, excavating, capping, sub-ballast, and within an embankment is clearing, improving, filling, capping, sub-ballast. These sequences are modeled with dummy activities connecting the end of the preceding activity with the start of the following activities (thin arrows e.g. from I1 to E1 in Figure 3-11b). In the activity network, two sub-networks are identified: the cut sub-network consists of the clearing activity, the capping activity, and the excavating activity connected by two dummy activities, while the embankment sub-network consists of the clearing activity, the improving activity, the capping activity, and the filling activity connected by three dummy activities (Figure 3-11c). The cut and embankment sub-networks are repeated as many times as the number of cuts and embankments in the sequence (two times in Figure 3-11a and 3-11b). The sub-ballast activities are added at the end of the activity network (Figure 3-11b) based on the assumption that the sub-ballast is placed after all cuts are excavated and capped, and all embankments are filled and capped.

As has been shown, the construction of cuts and embankments are modelled jointly in one activity network. The construction time of cuts and embankments is determined by the activity network but also by the construction material availability. This is discussed in the next section.

3.3 Material Handling in Cuts and Embankments

Material is used and produced in the construction of tunnels, viaducts, cuts and embankments. Depending on the quality of the material excavated in tunnels, this can be e.g. reused as aggregate in concrete liner of the tunnel. In the case of viaducts, material is mainly used in form of concrete

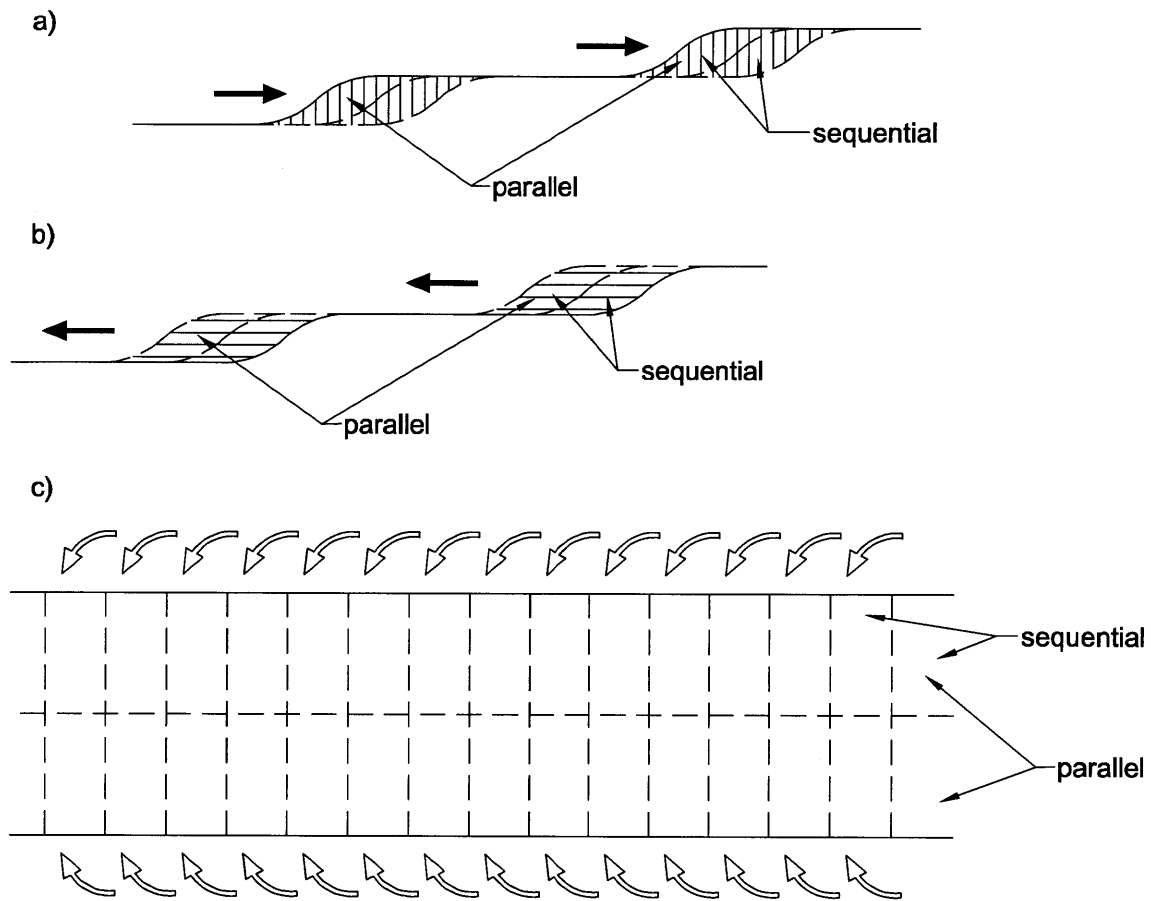


Figure 3-9: a) parallel and sequential excavation of a cut. b) parallel and sequential fill of an embankment. c) lateral parallel and sequential fill of an embankment. Cuts and embankments can be constructed in several ways, of which all include parallel and sequential processes.

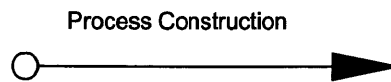


Figure 3-10: Activity modeling a process of the cut or embankment construction. The cost and the time of the activity are calculated with equations 3.7 and 3.8.

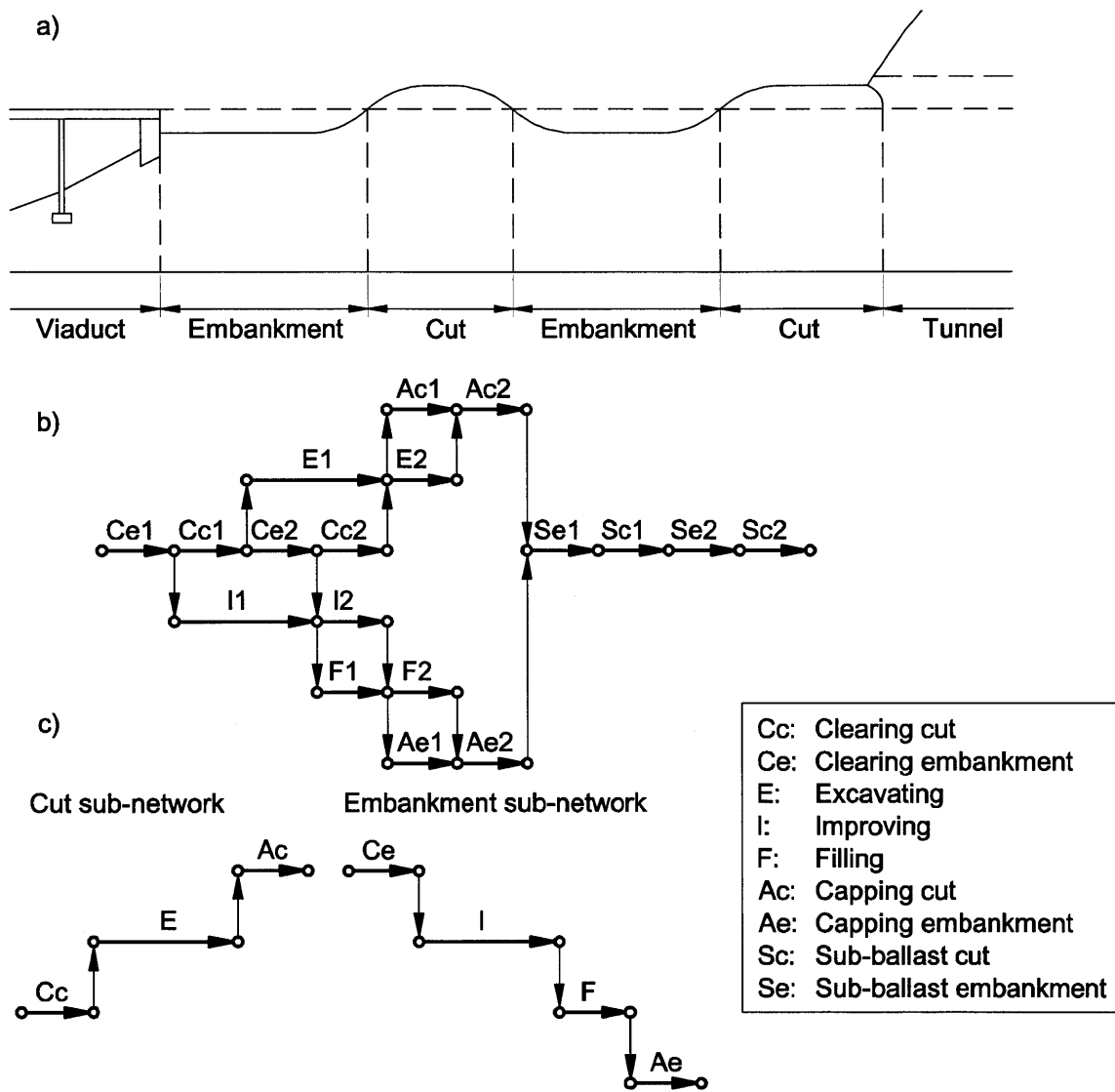


Figure 3-11: a) sequence of cuts and embankments between a viaduct and a tunnel, modeled with b) an activity network consisting of c) repeated sub-networks. The cut sub-network includes the clearing, the excavating and the capping activities, while the embankment sub-network includes the clearing, the improving, the filling and the capping activities. Each sub-network is repeated two times in the activity network, the number of cuts and embankments in the particular sequence (Figure a). In b) and c), thinner arrows are dummy activities.

and steel. Differently, the construction of cuts and embankments is characterized by the recycling of material excavated in the cuts, which is reused in the embankments. This process is further developed in this section.

The reuse of material in cuts and embankments is modeled seeking a mass balance between excavated material and filled material. The activity "excavating" produces material, while the activities "filling", "capping" and "sub-ballast" use material; however, the last two necessitate material of higher quality that is usually imported from outside the project. The activity "clearing" usually produces organics rich material, which is considered waste, and the activity "improving" produces and/or uses material depending on the improving method employed (e.g. replacing, injecting). Generally, the material excavated in the cuts is reused in the embankments fill, since it is not rich in organics and is of sufficient quality for the embankment body.

Peterson & Einstein (1993) and in Sinfield & Einstein (1996) modeled the reuse of material in a tunnel. In their model, one activity reproduces the transport of the excavated material from one origin point (the excavation site inside the tunnel) to one destination point (the material handling location outside the tunnel), where the material is transformed into aggregate for the tunnel liner concrete. Kollarou (2002) and Marzer (2002) modeled the construction of tunnels with activities and the transport of material with flows. These can have several origin and destination points, such as the aggregate plant, the concrete plant, the final repositories, and others. For each of them, origin point and destination point are known. In the case of the construction of cuts and embankments, there are many points of origin (cuts) and many destination points (embankments) and it is not known a priori from which cut and to which embankment the material flows.

The transport of material is modeled with activities that produce and use the excavated material. The material can leave the cuts and reach the embankments obeying different schemes (Marino 2009):

1. Excavate the cut, transport the material to a repository (location where the excavated material is temporarily stored), stockpile the material in the repository, excavate stockpiled material, transport the material to the embankment, fill the embankment.
2. Excavate the cut, transport the material to the embankment, fill the embankment.
3. Excavate the cut, push the material into the adjacent embankment(s), fill the embankment(s).

Schemes 1 and 3 are not further developed since the volumes of adjacent cuts and embankments are often unbalanced (scheme 3), and stockpiling material in the repository causes additional, undesirable costs (scheme 1). Scheme 2, which is further developed, comprises three processes: excavating the cut, transporting the material to the embankment, filling the embankment, of which excavating the cut produces the excavated material, and filling the embankment uses it. The three processes, the production and the use of excavated material are modeled with two activities:

- Activity: excavating.

Included processes: excavating the cut, transporting the material to the embankment.

Resource: excavated_material (produced).

- Activity: filling.

Included processes: receiving the excavated material from the cut, filling the embankment.

Resource: excavated_material (used).

These activities are changed to adapt the model to the constraints set by the simulation environment DAT (chapter 6), which require produced resources to be delivered to a repository and used resources to be retrieved from a repository (this can be a "virtual" repository):

- Activity: excavating.

Included processes: excavating the cut, transporting the material to a virtual repository.

Resource: excavated_material (produced).

- Activity: filling.

Included processes: transporting the material from the virtual repository to the embankment, filling the embankment.

Resource: excavated_material (used).

The excavation of the material and the filling of the material are modeled seeking a mass balance between excavated and filled material. This is modeled as follows: (1) the activity filling cannot start unless excavated material is available in the virtual repository, and (2) the activity excavating cannot produce excavated material unless the embankment can receive it, i.e. after the

clearing of the embankment is completed. The first condition is complied with through the resource `excavated_material`, while the second condition requires the introduction of a virtual resource, a resource that does not physically exist but allows one to model dependencies between activities. The virtual resource (Figure 3-12), `virtual_material`, is produced by the activity clearing, only in embankments (clearing of cuts does not produce virtual material), in the quantity equivalent to the filling volume (V_{fi}). It is used by the activity excavating in the quantity equivalent to the excavation volume (V_{ex}). The produced material is delivered to the repository in steps at every unit length of embankment, totalling the volume V_{fi} at the end of clearing. The used material is retrieved from the repository also in steps at every unit length of cut, totalling the volume V_{ex} at the end of excavating. By delivering the virtual resource, the activity clearing triggers the activity excavating. This uses the virtual resource in the quantity equivalent to the excavating volume (V_{ex}) and produces the resource `excavated_material` in the quantity also equivalent to the excavating volume (V_{ex}). This is used by the activity filling in the quantity equivalent to the filling volume (V_{fi}). Similarly to the virtual resource, the produced material is delivered to the repository in steps at every unit length of cut, totalling the volume V_{ex} at the end of excavating. The used material is retrieved from the repository also in steps at every unit length of embankment, totalling the volume (V_{fi}) at the end of filling.

The two conditions, (1) the activity filling cannot start unless excavated material is available in the virtual repository and (2) the activity excavating cannot produce excavated material unless the embankment can receive it, are considered with the introduction of a virtual resource and modeled with three activities:

- Activity: clearing embankment.

Included processes: clearing the embankment, transporting the virtual material to a virtual repository.

Resource: `virtual_material` (produced).

- Activity: excavating.

Included processes: retrieving the virtual material from the virtual repository, excavating the cut, transporting the excavated material to a virtual repository.

Resource: `virtual_material` (used)

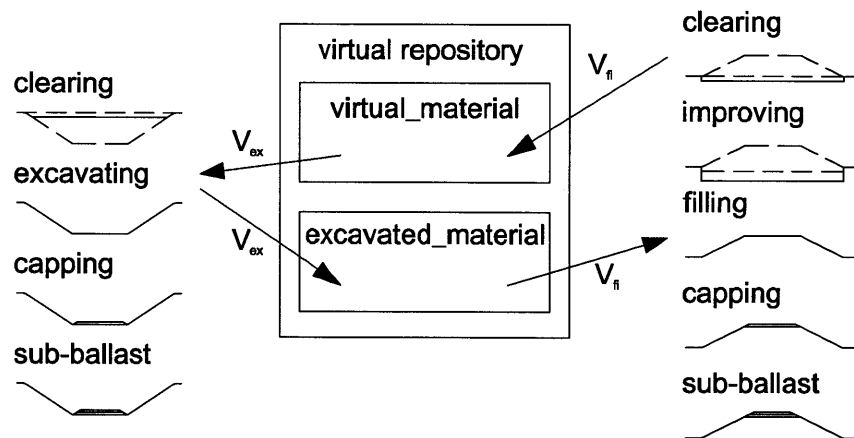


Figure 3-12: The virtual material is delivered by the clearing activity of the embankment to the repository in the quantity equivalent to the filling volume (V_{fi}). It is used by the activity excavating in the quantity equivalent to the excavation volume (V_{ex}). By delivering the virtual resource, the activity clearing triggers the activity excavating. This uses the virtual resource and produces the resource `excavated_material` in the quantity also equivalent to the excavating volume (V_{ex}). This is used by the activity filling in the quantity equivalent to the filling volume (V_{fi}). This concept, which applies to all cuts and embankments in the alignment, controls the mass balance between cuts and embankments.

Resource: `excavated_material` (produced).

- Activity: filling.

Included processes: transporting the excavated material from the virtual repository to the embankment, filling the embankment.

Resource: `excavated_material` (used).

For the mass balance we compare materials with the same volume, that is loose material. The material varies in volume in situ, loose or compacted; notably, the volume of loose material is largest, and usually the in situ volume is larger than the compacted volume (equation 3.9)*. Changes in volumes are taken into account through loosening and compacting factors (equations 3.10 and 3.11).

$$Vol_{insitu} < Vol_{compacted} < Vol_{loose} \quad (3.9)$$

*This is not always the case: for instance, the in situ volume of loose sand tends to be smaller than the compacted volume.

$$Vol_{insitu} = \rho_l \times Vol_{loose} \quad (3.10)$$

$$Vol_{compacted} = \rho_c \times Vol_{insitu} \quad (3.11)$$

If the loose volume produced is larger than the loose volume used in the project, the material remaining in the repository is not used in the project and deposited e.g. in a landfill, whereas if the loose material produced is smaller than the loose material used, additional material is imported from outside the project. These two processes are modeled with two activities: a depositing and an importing activities. The depositing activity transports the excavated material not used from the repository to the landfill. This is modeled with an activity and two resources:

- Activity: depositing.

Included processes: depositing the material remaining from the virtual repository to a final destination.

Resource: excavated_material (used)

Resource: deposited_material (produced).

The importing activity increases the volume of excavated material in the virtual repository. It is modeled with an activity producing excavated material:

- Activity: importing.

Included processes: importing material to the virtual repository.

Resource: excavated_material (produced).

The construction and the material handling of any sequence of cuts and embankments are modeled with four cut activities, five embankment activities, one depositing or importing activity, three resources and two repositories (Table 3.1).

The material handling in the construction of cuts and embankments has been analyzed at the level of the resources used and produced by the activities. With the proposed model of activities and resources, the construction and the material handling of a sequence of cuts and embankments can be represented.

Table 3.1: The construction and the material handling of cuts and embankments are modeled with four cut activities, five embankment activities, one depositing or importing activity, three resources and two repositories.

Cut activities	embankment activities	other activities	resources	repositories
clearing	clearing	depositing	excavated_material	virtual repository
excavating	improving	importing	virtual_material	final destination
capping	filling		deposited_material	
sub-ballast	capping			
	sub-ballast			

3.4 Conclusions

The construction model consists of activity networks modeling the four main structures of rail lines: tunnels, viaducts, cuts and embankments. The activity costs and times of the structures determine the construction cost and time of the entire rail line. The complexity of the activity networks vary from the simplest one-activity activity network of tunnels to the complex activity networks of viaducts. The cost and time equations of activities in the tunnel-, viaduct-, and cut- and embankment- networks have been presented.

The analysis of the activity networks showed that there is a repetition of sub-networks during the construction of a structure. The interconnection of these sub-networks models one structure, and the interconnection of the structure networks models the construction of the rail line.

The representation of the construction process of rail lines with activity networks deepens the understanding of the construction process, provides insight when identifying the sources of uncertainty and enables the development of an innovative approach to uncertainty modeling in transportation infrastructure projects: representing the uncertainty sources at the activity level (chapter 4).

Chapter 4

Uncertainty Model

Rail line construction projects are often plagued by cost and time overruns. Most available estimation tools do not capture the uncertainty in the construction projects. There is the need for estimation tools that, based on an in-depth understanding of the construction process, model its uncertainties. The author developed an uncertainty model that represents the uncertainties in the construction process. This chapter presents the different sources of uncertainty, how the uncertainty model captures them, and their impact on project cost and time. With the uncertainty model, it has been found that the sources of uncertainty have a significant impact on the total cost and total time of the project. The uncertainty model also deepens the understanding of uncertainty.

The detailed structure of the chapter is as follows: first, the concept of uncertainty is defined (section 4.1). Second, the uncertainty model based on the sources of uncertainty (variability, cost correlations, and disruptive events) is presented (section 4.2). Then, the source of the data (historical data and expert opinions) for the uncertainty model are discussed with special attention to the process of obtaining expert opinions (section 4.3). At last, the effectiveness, the advantages and the contributions of the uncertainty model are discussed (section 4.4).

4.1 Uncertainty Definition

In this section, the concept of uncertainty is defined and examples of an uncertainty with a negative outcome, a threat, and of an uncertainty with a positive outcome, an opportunity, are given.

For the uncertainty model, there is uncertainty when something is not known (besides un-

certainty due to lack of knowledge, there is uncertainty due to randomness). Typically, in the estimation of construction projects several variables are not known since construction projects are populated by uncertainties.

Compared to risk, the concept of uncertainty has a neutral connotation: an uncertainty, when it is realized, can have a positive or a negative impact on the project objectives. It is an opportunity if it has a positive impact, a threat if it has a negative impact on the project objectives. An example with a threat and an opportunity is illustrated for the case of a tunnel excavated in dry rock with a tunnel boring machine. A possible threat is a sudden water inflow: its probability of occurrence is 1%, and the potential impact (loss) is an additional cost, which is modeled with a triangular distribution with a minimum value of \$100K (100 thousand dollars), a mode value of \$200K, and a maximum value of \$500K. The probability of occurrence of the disruptive event and the cost loss if the event occurs are:

$$P(\textit{Threat}) = 1\% \quad (4.1)$$

$$f_{\textit{Loss/Threat}} = \textit{triangular}(\$100K, \$200K, \$500K) \quad (4.2)$$

where $f_{\textit{Loss/Threat}}$ is the probability distribution of the cost loss given the threat (sudden water inflow) has occurred.

Similarly, an opportunity can be modeled: a possible opportunity is the delivery of disks for the tunnel boring machine that are more resistant against rock abrasion. Its probability of occurrence is 30%, and the potential impact (gain) is uniformly distributed with a minimum value of \$1K/m (1 thousand dollars per meter excavated) and a maximum value of \$2K/m. The probability of occurrence of the disruptive event and the cost gain if the event occurs are:

$$P(\textit{Opportunity}) = 30\% \quad (4.3)$$

$$f_{\textit{Gain/Opportunity}} = \textit{uniform}(\$1K/m, \$2K/m) \quad (4.4)$$

where $f_{\textit{Gain/Opportunity}}$ is the probability distribution of the cost gain given the opportunity (delivery of tunnel boring machine disks) has occurred.

In this section, the concept of uncertainty has been defined and examples of uncertainties

with positive and negative outcome have been given. In the next section, the author proposes an uncertainty model that captures three sources of uncertainties: variability, correlations, and disruptive events. These and their impact on the distributions of the project cost and time are discussed.

4.2 Sources of Uncertainty

The uncertainty model considers three sources of uncertainty: the variability in the construction process, the correlations between construction costs and the occurrence of disruptive events. These are described, their models are presented and their impacts are discussed with a tunnel example. They can be similarly modeled in the construction of the other structures (viaducts, cuts and embankments).

In the next section, the first source of uncertainty, variability, is considered.

4.2.1 Variability

In the uncertainty model, variability is one of the three sources of uncertainty. Here, it is described, the model and generation of variability are presented, and the impact on construction cost and time is discussed.

The variability is the change in a variable under normal conditions, i.e. a regular construction process, such as the change in cost of excavating one meter of tunnel and the next meter of tunnel in the same geology and construction environment. This variability is modeled with probability distributions: the lognormal distribution for the variability in cost variables and the triangular distribution for the variability in time variables (Figure 4-1). The lognormal distribution is the selected probability distribution to model cost variables since it often underlies the distribution of construction cost variables (Touran & Wiser 1992). Time variables are modeled with the triangular distribution for four main reasons: 1) it is closed-ended in the lower tail (time variables are positive), 2) it can be either skewed to the left or skewed to the right, 3) the minimum, mode, and maximum parameters can be relatively easily estimated by an expert, and 4) it is often used in construction modeling (Chau 1995, Back et al. 2000, Haas & Einstein 2002). When modeling the variability of costs and times, the range of the possible total cost and total time is given by a cloud of points (Figure 4-4b, page 125), differently from the deterministic estimate which corresponds to one point

(Figure 4-4a, page 125). The cloud of points visually represents the variability of the resulting total cost and total time.

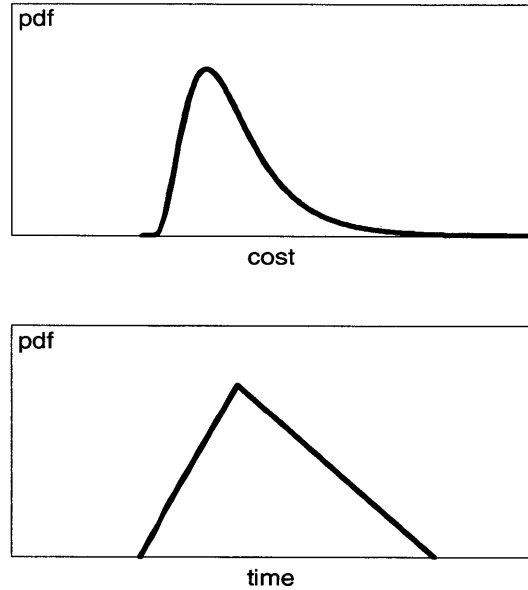


Figure 4-1: Lognormal and triangular probability distributions are used to model the variability in the cost and time variables, respectively. The lognormal distribution is the selected probability distribution to model cost variables since it often underlies the distribution of construction cost variables. Time variables are modeled with the triangular distribution for four reasons: it has a positive minimum, it can be skewed either to the left or to the right, it is common in construction modeling, and its parameters are relatively easily estimated by an expert.

In the following, a tunnel example is used to describe 1) how the variability is modeled, 2) how the costs and the times are generated, 3) how the total cost and time are calculated, and 4) to show the impact of variability on the total cost and time. The variability in the cost and the time of excavating one unit length of tunnel is modeled for each unit length of the tunnel. This is modeled with a lognormal distribution for cost and a triangular distribution for time (Figure 4-1). From the probability distribution, one number each for cost and time to excavate a unit length of tunnel is generated per simulation with a Monte Carlo process. The number generation is repeated for every unit length of tunnel until the entire tunnel has been excavated. At the end of the simulation run, the costs, respectively the times, are summed to calculate the total cost and the total time of the

tunnel for the specific simulation run. The procedure is repeated for every simulation run until the simulation is completed. The total cost and the total time are the expression of the accumulated uncertainties in the construction of all unit lengths of the tunnel. Since they change from simulation to simulation due to the construction process uncertainty, the cost-time scattergram is represented by a different point for every simulation run (Figure 4-4b, page 125).

In the preceding, the first source of uncertainty, variability, and its impact have been discussed. In the next section, the second source of uncertainty, cost correlations, is considered.

4.2.2 Cost Correlations

In the uncertainty model, cost correlations are one of the three sources of uncertainty. Here, cost correlations are described, the model and generation of correlations are presented, and the impact on construction cost and time is discussed.

The relation between two variables is expressed with a correlation: if the value of one variable is above average, the value of the second variable tends to be above average when they are positively correlated, while it tends to be below average when they are negatively correlated. The correlation varies between -1 (fully negatively correlated) and $+1$ (fully positively correlated), and for a correlation equal to 0 , the two variables are uncorrelated (Figure 4-2). The correlation is measured with the Spearman correlation coefficient and it is modelled with NORTA (Cario & Nelson 1997). The choice of the Spearman correlation amongst the available correlation measures and of NORTA amongst the available correlation models is discussed in chapter 5. In the uncertainty model, the focus is on correlations between cost variables. Several different types of correlations have been identified in the construction of rail lines, of which the most relevant is the correlation between the costs of a repeated activity in a structure, since it has the largest impact on the standard deviation of the total cost (chapter 5). Other types of correlation between variables and their impact are analyzed also in chapter 5. Compared to the total cost and the total time when modeling only the variability, the cost correlations cause the cloud of points to expand in the total cost direction (Figures 4-4b and 4-4c, page 125). In other words, the range of possible total costs increases.

In the following, a tunnel example is used to describe 1) how the correlations are modeled, 2) how the costs and the times are generated, 3) how the total cost and time are calculated, and 4) to show the impact of cost correlations on the total cost and time. The correlation between the costs

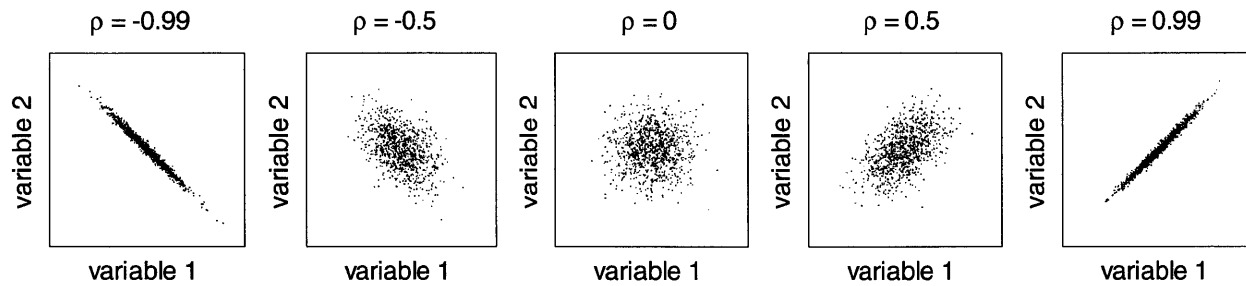


Figure 4-2: Varying degrees of correlation between variables 1 and 2: from left to right, highly negatively correlated, mildly negatively correlated, uncorrelated, mildly positively correlated, and highly positively correlated. Correlations between cost variables are modeled in the proposed methodology using Spearman correlation coefficients and the correlation model NORTA (Chapter 5).

of constructing subsequent unit lengths of tunnel is represented with the Spearman correlation coefficient and it is modeled with the NORTA correlation method. Differently from a random number generator, NORTA first generates random numbers from a uniform distribution, then through two linear transformations it obtains random numbers correlated with the desired correlations and distributed with the desired probability distribution. It generates the costs of excavating the unit lengths of the tunnel, which are correlated, then these are summed to obtain the total cost of the tunnel. The NORTA generation and the summing of the costs of all the unit lengths are repeated for each simulation run. With the NORTA generation the uncertainty in each unit length of tunnel is modeled, as well as the correlation between the costs of the unit lengths of the tunnel. A positive correlation causes the standard deviation of the sum of the correlated variables to increase (chapter 5): since cost correlations in construction are usually positive, the standard deviation, i.e. the spread of the tunnel total cost, increases (Figure 4-4c, page 125). Since the costs are positively correlated, while the times are modeled as independent, the spread in the total cost increases, while the spread in the total time remains the same (compare Figures 4-4b and 4-4c, page 125).

In this section, the second source of uncertainty, cost correlations, and their impact have been discussed. In the following, the last source of uncertainty, disruptive events, is considered.

4.2.3 Disruptive Events

In the uncertainty model, disruptive events are one of the three sources of uncertainty. They are first described below, then the model and generation of disruptive events are presented, and the impact on construction cost and time is discussed.

A disruptive event is an event with a large cost and/or time impact and usually a small probability of occurrence, such as a flooding. It can severely disrupt the construction process. Due to the rare occurrence of disruptive events, there are in general no studies on the underlying distributions of cost and time impacts of such events. The cost and the time impacts are modeled with the triangular distribution for three reasons: 1) it is closed-ended in the lower tail (time and cost variables are positive), 2) it can be either skewed to the left or skewed to the right, and 3) the minimum, mode, and maximum parameters can be relatively easily estimated by an expert. When modeling the disruptive event in addition to the variability and cost correlations, the cloud of points explodes and the range of possible total costs and total times increases dramatically (Figure 4-4d, page 125).

In the following, a tunnel example is used to describe 1) how the disruptive events are modeled, 2) how the costs and the times are generated, 3) how the total cost and time are calculated, and 4) to show the impact of disruptive events on the total cost and time. The occurrence of a disruptive event, e.g. a large water inflow, is modeled at every excavated unit length. If the disruptive event occurs, the cost and time impact of the disruptive event on the construction are also modeled. The occurrence or non-occurrence of the disruptive event is modeled with a simple random number generation: for a unit length of the tunnel a number is generated from a uniform distribution $[0, 1]$. If the generated number is larger than the probability of occurrence, the disruptive event does not occur, and the simulation proceeds to the next unit length. On the other hand, if the generated number is smaller or equal to the probability of occurrence, the disruptive event occurs in the unit length, and its cost and time impacts must also be simulated. These are modeled with triangular distributions (Figure 4-3), which are used to generate the cost and time impacts of the disruptive event. These cost and time impacts are added to the cost and time of the tunnel unit length. Since disruptive events have large cost and time impacts, the cost and time of the unit lengths, where the disruptive event occurs, can be significantly larger than the cost and time of a unit length where the disruptive event does not occur. The cost and time of all unit lengths are summed to obtain the total tunnel cost and time for a simulation run, which is one point in the cost-time scattergram.

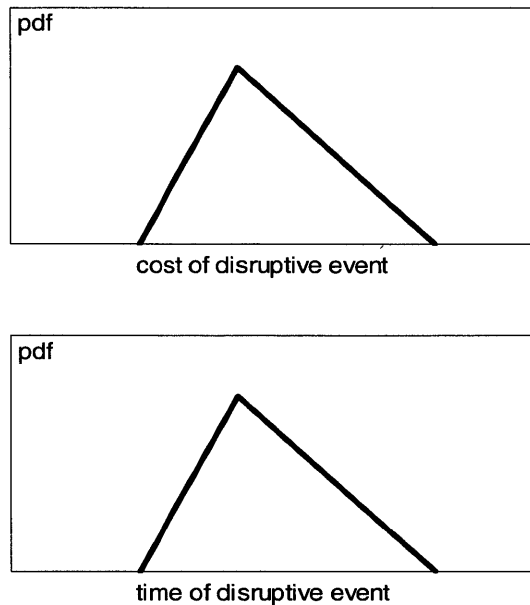


Figure 4-3: The cost and the time impacts of a disruptive event are modeled with triangular distributions. Since there are only very few studies on the underlying distributions of cost and time impacts of disruptive events, due to their rare occurrence, the triangular distribution is preferred for three reasons: it has a positive minimum, it can be skewed either to the left or to the right, and its parameters are relatively easily estimated by an expert.

Since each simulation run is different, the occurrences of the disruptive event change: it may not occur, it may occur once, or it may occur more than once. In Figure 4-4d (page 125), the majority of the simulation runs did not register the occurrence of a disruptive event (compact cloud in the same position as the cloud in Figure 4-4c), while in some cases it occurred at least once (spread cloud with larger total cost and total time in Figure 4-4d).

In the preceding, the last source of uncertainty, disruptive events, and their impact have been discussed. In the next section, the impacts of the three sources of uncertainty are compared and their cumulative impact is described.

4.2.4 Cumulative Impact of the Sources of Uncertainty

Variability, cost correlations and disruptive events have different impacts on construction cost and time. Here, these are compared and their cumulative impact is presented.

The uncertainty model generates a cost time scattergram displaying the uncertainty in the cost and time of the project construction. Depending on the uncertainties modeled, it generates strikingly different scattergrams (Figure 4-4). These vary from a single point when no uncertainty is modeled (best estimate) to a cloud when uncertainties are considered: modeling variability creates a small cloud around the best estimate, modeling correlations between costs causes the cloud to expand significantly along the cost axis, and modeling disruptive events changes the scattergram into a wide spread one. These observations are simply indicative since the magnitude of the impacts depends on the characteristics of the three uncertainties and the structure analyzed. However, from the above comparison of the impacts of the three sources of uncertainty, one can see that disruptive events have the largest impact on the range of possible total cost and total time. Most importantly, Figure 4-4d is the representation of the cumulative impact of the three sources of uncertainty: a widely scattered cloud of cost-time data points.

The impacts of the single uncertainty sources have been compared and their cumulative impact has been briefly discussed. Detailed comparisons and analyses based on an application of the uncertainty model to a part of the new Portuguese high speed rail line will follow in chapter 7.

Three sources of uncertainty (variability, cost correlations, and disruptive events) have been presented and their impact on construction cost and time illustrated. Modeling them requires a certain amount of input data. These are discussed in the next section.

4.3 Expert Estimation

An uncertainty model that considers three uncertainty sources (variability, cost correlations, and disruptive events) applied to four structures (tunnels, viaducts, cuts, and embankments) requires a significant amount of input data: for each structure probability distributions of activity cost and time are required as well as correlations between the costs of activities, the probability of occurrence of disruptive events and the distributions of their cost and time impacts.

The research work presented in this thesis is based on historical data and expert opinions (Table 4.1). Historical data were available on the costs and the times of tunnel construction (RAVE 2006d, RAVE 2008d), on the costs of viaduct construction (RAVE 2006c, RAVE 2008c), and on the costs of cuts and embankments (RAVE 2006b, RAVE 2008b). Other information, namely the times of viaduct construction, the times of cuts and embankments, the cost correlations

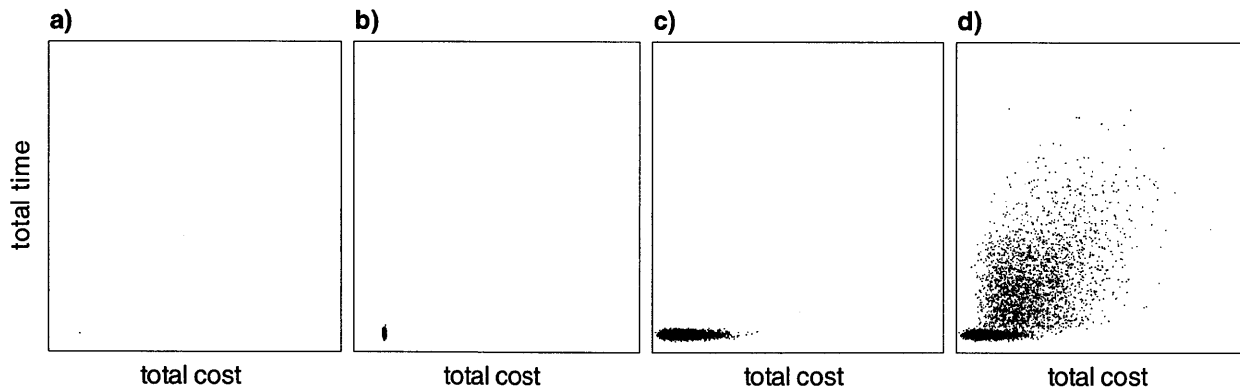


Figure 4-4: Total cost and total time distributions for the construction of a tunnel a) modeling no uncertainty, b) modeling variability, c) modeling variability and cost correlations, d) modeling variability, cost correlations and one disruptive event. Compared to the single point of the deterministic calculation in a), the clouds of points in b), c) and d) visualize the impact of uncertainty on the total cost and the total time. In b) the cloud of points visually represents the variability of the total cost and total time. In c) the cost correlations cause the cloud of points to expand in the total cost direction. In d) the cloud of points explodes due to the additional cost and time of the disruptive event. The largest impact is produced by the disruptive event for which the range of possible outcomes drastically increases in both total cost and total time. Note that the graphs are shown with same axis scales in order to guarantee comparability.

of all structures, and the disruptive events (probability of occurrence and cost and time impacts) of all structures had to be obtained from expert opinions. When using expert opinions to obtain data, particular care was needed in the data gathering process in order to guarantee the data quality.

In this section, the main aspects of expert estimation are discussed. In particular, biases are described, expert calibration and questionnaire preparation are discussed, findings and lessons learnt are reported, and recommendations on the process to gather expert opinions when estimating probabilities are given.

When gathering input data, preference was given to historical data but when these were not available, the author resorted to expert estimation. An expert can be described as a very skillful person who underwent lots of training and is very knowledgeable in her/his special field. From her/his knowledge he/she can provide expert opinion in the process of expert estimation. The expert opinion is the formal judgment of the expert on a matter in which her/his advice is sought (Ayyub 2001). In this research work, a tunnel expert, a viaduct expert and an earthwork (cuts and embankments) expert estimated the required information on variability, cost correlations and

Table 4.1: Structures (tunnels, viaducts, cuts, and embankments) and uncertainties (variability, cost correlations, and disruptive events) in the construction of rail lines. In this research work, the time variability of viaduct-, cut- and embankment construction as well as the cost correlations and the disruptive events of all structures were not available from historical sources and they had therefore to be estimated by experts.

Structures	Uncertainties	variability		cost	disruptive
		cost	time	correlations	events
tunnels		data	data	expert	expert
viaducts		data	expert	expert	expert
earthwork (cuts & embankments)		data	expert	expert	expert

disruptive events. The experts have over 20 years experience in their respective fields.

When asking an expert to estimate probabilities, one should be aware of biases that can lead to errors in the estimation. Biases are caused by the incorrect use by experts of rules of thumb, or heuristics, when estimating probabilities. Biases also occur when e.g. an axiom of probability is violated or an estimation does not represent the expert's belief (Cooke 1991). Biases in expert estimation are a well known and researched problem and are described in detail in the literature. Here, the biases relevant to the research work presented in this thesis are discussed (Cooke 1991, Ayyub 2001, Hallowell & Gambatese 2010, Flyvbjerg et al. 2002):

- availability: experts tend to base their estimates on the ease, which allows them to retrieve the information from their memory; in other words, an expert usually estimates the probability of an event by the ease with which she/he can recall instances of such an event. Typically, experts overestimate the probability of well-publicized events and underestimate the probability of less-publicized events: for instance, the probability of death by plane accident is overestimated, while the probability of death by car accident is underestimated.
- anchoring: when asked to estimate a probability, experts may fix their estimate to an initial value and adjust, but with an insufficient adjustment. This applies, in particular, to the estimation of small and large quantiles, e.g. 5% and 95%, when the expert fixes to a central value and adjusts. As a consequence, the small quantile is overestimated and the large quantile is underestimated.
- overconfidence: experts tend to provide narrower confidence intervals compared to real intervals. When estimating the confidence interval of a variable, an expert tends to estimate

a larger minimum value than the real minimum and a smaller maximum value than the real maximum so that the estimated interval is narrower than it should be.

- recency: experts are more likely to inflate the probability of an event if such an event has recently happened to them.
- contrast: an individual's perception of a variable may be influenced (enhanced/diminished) by the exposure to a (larger/lower) value of the immediately preceding variable.
- optimism bias: planners and project promoters tend to be excessively optimistic by focusing on success scenarios and overlooking the chance of failure.

Two well-known biases are worthwhile mentioning although they do not influence the estimation of variables of this thesis work: the representativeness and the base rate fallacy biases. The representativeness bias (Cooke 1991) occurs when individuals are asked to judge the conditional probability $p(A|B)$ that event A occurs given that event B has occurred, it seems that they rely on an assessment of the degree of similarity between event A and B. This heuristic leads to bias since similarity is symmetrical (A resembles B in the same way B resembles A), while conditional probability is not symmetrical, since

$$p(A|B) = p(B|A) \times \frac{p(A)}{p(B)}. \quad (4.5)$$

The base rate fallacy bias can be summarized in the inability of individuals to estimate a conditional probability. An example from the field of medicine is provided by Cooke (1991): a diagnosis for breast cancer is accomplished by means of a biopsy, a surgical operation, which is often performed on the basis of an x-ray image. When asked for the conditional probability of cancer given a positive x-ray result, physicians estimate it at 75%. This conditional probability of cancer given a positive x-ray result, $p(C|+)$, can be calculated with the Bayes' theorem

$$p(C|+) = \frac{p(+|C)p(C)}{p(+)} \quad (4.6)$$

and available data on the conditional probability $p(+|C)$ and the two probabilities $p(C)$ and $p(+)$: it is equal to 8%, almost a tenth of the value estimated by physicians (75%).

Biases can be avoided with the calibration of the experts and carefully prepared questionnaires. Calibration involves training the experts on the use of the estimated values, probability concepts, and biases. In our calibration, the expert was first informed about the use of the questionnaire's answers in the estimation of total construction cost and time of a new rail line. Second, the concepts of probability of occurrence, triangular distribution and correlation were explained and visualized to enhance the familiarization of the expert with the subject (Figure 4-5 and Figure 4-2, respectively). Third, the expert was warned about the existence of biases such as anchoring, availability, recency and their effect on the outcome of an expert estimation. Also overconfidence when estimating minimum and maximum values or small and large correlations was discussed in order to make the experts aware of this bias and to prevent them from committing the same type of error. The bias of contrast was not discussed with the expert; however, it was addressed with the randomization of the questions' sequence in the questionnaire (see below). Optimism bias did not apply to the experts in this specific case since none of them was a planner or a promoter of the project. Additionally, the questionnaire takes the outside view (section 2.3.3), i.e. it does not allow the expert to focus on the details of the project but rather it makes her/him draw the answers from her/his extensive experience.

The second instrument to avoid biases in expert estimations are well-prepared questionnaires. Ayyub (2001) gives guidelines on the construction of questions and stating issues in a questionnaire:

- Each item on the questionnaire should include only one question, since it is poor practice to include two questions in one.
- Special attention should be given to concept definition within the context of each question to make sure this is not ambiguous.
- It is good practice to avoid long questions.
- Abstract questions should be avoided in favor of factual questions.
- Questions should be asked in a neutral format to prevent the experts from being influenced.
- Sensitive topics may require stating questions with lead statements that establish supposedly accepted social norms in order to encourage subjects to answer the question truthfully.

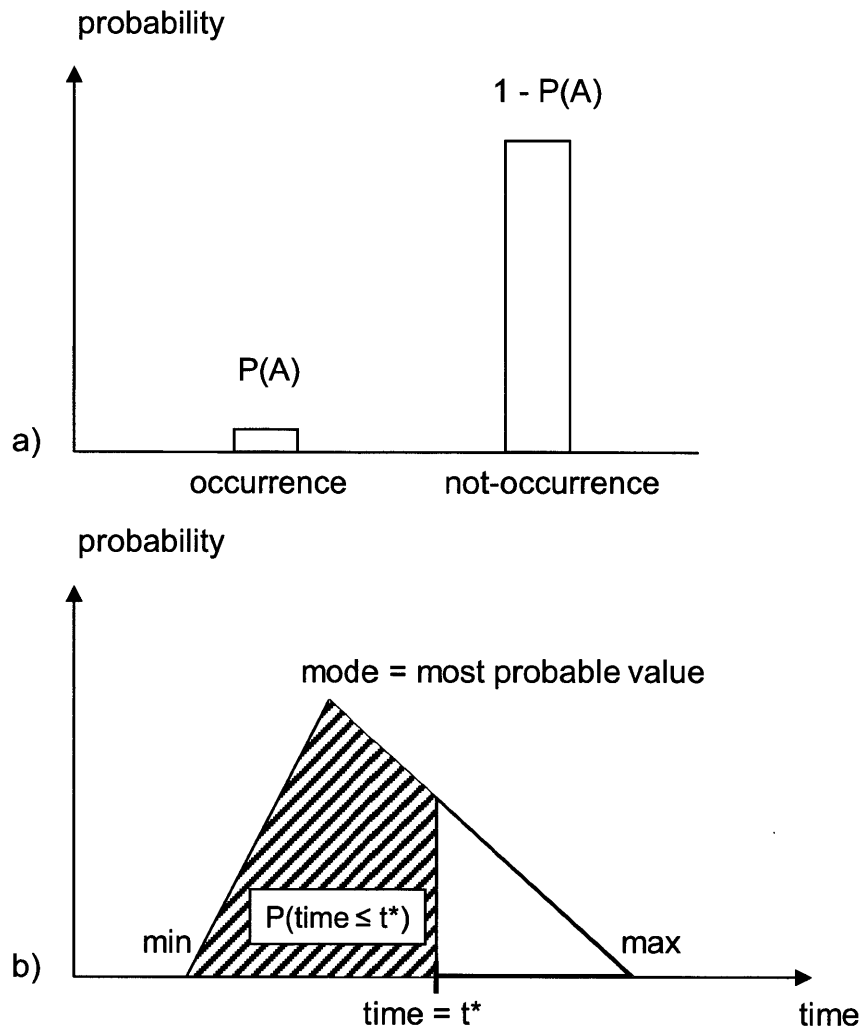


Figure 4-5: Concept visualization for a) the probability of occurrence and not-occurrence of event A, and for b) the triangular distribution of a time variable. These two probability concepts were reviewed with the experts in the calibration.

Two suggestions are given by Hallowell & Gambatese (2010) on the organization of the questions when asking for numerical values such as probabilities:

- When asking for the probability and the severity of an event, it is good practice to ask two separate questions, one addressing the probability of occurrence and one addressing the severity if the event occurs.
- In order to limit the contrast bias (influence by exposure to previous values), the question order should be randomized.

Following the just mentioned good practices, the questionnaire was prepared. The questionnaire consists of short, clearly-stated questions, each addressing one item. The estimation of the probability of occurrence of a disruptive event was separated from the estimation of the cost and time impact of the event. Finally, the sequence of the questions was randomized.

For the thesis work, the experts had to answer questions on the following items: the time of activities in the viaduct-, cut- and embankment construction, and for all structures the cost correlations as well as the disruptive events (Table 4.1). Note that information, for which no questions were asked, was available in the form of historical data. Examples of questions for the three items are given here (the entire questionnaire can be found in Appendix A):

- activity duration: "At the minimum how long does it take to construct a viaduct abutment?" Similarly, the most probable- and the maximum time to construct the abutment were asked. Analogous questions were asked about other activities in viaduct construction, and about activities in the construction of cuts and embankments. Since probability distributions of activity times of tunnel construction were available from the designers, the tunnel expert did not have to answer questions on activity duration.
- cost correlation: "What is the correlation between the cost of constructing a unit length (i) of tunnel and the cost of constructing the next unit length ($i + 1$) of tunnel, $\rho(i, i + 1)$? Similar questions were asked from the earthwork (cuts and embankments) expert about the cost correlations between activities such as excavating and filling. The viaduct expert excluded correlations in viaduct construction with the exception of correlations between the activities to construct the piers and between the activities to construct the deck sections.

- disruptive events: "Name the two most important disruptive events in cut and embankment construction." This introductory question was followed by the estimation of the probability of occurrence of the disruptive events. After the probability of occurrence had been estimated, the time and cost distributions were asked.

Two a priori assumptions were made in the questionnaire: the triangular distribution was assumed to be the probability distribution underlying the time variables as well as the cost and time impact of the disruptive events, and the correlation was estimated directly as a Spearman correlation coefficient. The first assumption has been discussed extensively in section 4.2. Regarding the correlation estimation, the direct estimation of the correlation value is preferred over indirect methods that estimate e.g. the conditional probability and the strength of relationships, and from these calculate the correlation coefficient (Clemen et al. 2000). Compared to the indirect methods, the direct estimation of the correlation coefficient presents two main advantages: it delivers the most accurate estimated correlation and the estimated value falls naturally between -1 and $+1$ (Clemen et al. 2000).

The estimation process consisted of four sessions. First, we conducted a feasibility check: each expert was asked if she/he thought the questions could be answered or should be asked differently. Second, the questionnaire was answered in two sessions of one hour length each, in order not to overwhelm the expert with too long sessions. Each of the two sessions included calibration and responding to the questions while the analyst was available for explanations. In a last session, the estimated values were clarified.

The following findings were obtained from the comments made by the experts and the discussions of the estimated values. First, when estimating probability distributions of either cost or time variables, the three experts followed radically different techniques. From past projects, the tunnel expert recalled the minimum and the maximum values of the cost/time variable, while he estimated the mode. He acknowledged to have more difficulties in estimating the mode rather than indicating the extremes. From experience, the viaduct expert calculated the minimum possible cost/time value and the most probable (mode) cost/time value, whilst she consistently estimated the maximum value as twice the mode value. She emphasized that the maximum was a rough estimate and it could be larger. The earthwork expert first estimated the mode value; then, from this he estimated the minimum and the maximum values.

Second, the randomization of the questions sequence was found to be unnecessary if not distracting by both the tunnel and the viaduct experts, so that it was not implemented in the questionnaire for the earthwork expert. Specifically, the tunnel and the viaduct experts navigated the questionnaire according to their line of thought: for example, after estimating the minimum of a variable, they would estimate the mode and the maximum of the same variable. This finding is related to the previous one: since the tunnel expert remembers the minimum and maximum values and the viaduct expert calculated the minimum and mode values, the contrast bias is not relevant to them. Thus, the randomization was not needed to prevent the contrast bias.

Third, estimating correlations between costs proved to be the most challenging task for all three experts. The estimation of correlations caused different degrees of difficulty although the experts had received the same introduction to the concept of correlation, including concept explanation and the visualization of different degrees of correlation (Figure 4-2). For the tunnel expert, some additional clarification was sufficient to estimate the requested correlations. With the viaduct expert, a discussion including some examples specific to her field helped in the estimation. The earthwork expert did not feel comfortable giving any estimation also after some detailed discussion of examples related to his field, finding the correlation concept too abstract.

The first and the second findings indicate that experts using estimation techniques rooted in the exact recording of past projects (tunnel expert) or in calculations of cost/time values (viaduct expert) 1) are not influenced by biases such as contrast, availability, anchoring and recency, and 2) do not benefit but rather are disturbed by the randomization of the question sequence. The third finding underscores the importance of the choice of experts: in correlation estimation, it is not sufficient for the estimator to be an expert in her/his field but also her/his knowledge of probability concepts is crucial for a successful estimation process.

It should be noted that the structure of the estimation process could be improved: for instance, an estimation involving more experts per field could lead to better results by integrating more opinions. In our specific case, it proved difficult to identify experts that 1) were experts in the field of interest, and 2) were acquainted with probability concepts. Another possibility to improve the accuracy of the expert estimations is to train the experts with a comparison of estimated values with available data (Ayyub 2001, Clemen et al. 2000). Again, in our case such training proved difficult to implement due to the data deficient environment, in particular concerning cost correlations and

disruptive events.

Aggregating the experience gained while questioning the experts and the good practices suggested in the literature (see above), the author developed some recommendations when seeking opinions on probability distributions of cost and time variables, on probability of occurrence of disruptive events and on correlations:

- selection of experts: it is paramount that the expert is knowledgeable on probability concepts besides possessing expertise in her/his field. It is advisable to select experts that are not acquainted with the project to avoid optimism bias.
- structure of estimation process: we found a four-session structure very valuable. The feasibility check (first session) allowed us to make changes in the questionnaire so that we asked questions that the experts could answer. The estimation sessions (second and third session) allowed the experts to go through the questions with no time pressure. Most importantly, the experts did not show any symptoms of growing tired of the estimation process and gave their best estimate until the very last question. The discussion session (fourth session) permitted us to clarify the estimated values if needed.
- debiasing techniques: calibration should be a cornerstone of every expert estimation. We found it satisfactory to have a calibration at the beginning of every estimation session. The calibration should explain and visualize probability concepts, and it should discuss biases in detail.
- questionnaire preparation: the guidelines for good practice in the literature (Ayyub 2001, Hallowell & Gambatese 2010) are extremely valuable. We encourage to write short, clearly-stated questions, to include one item per question, to separate the estimation of the probability of occurrence from the estimation of the severity. However, we disagree with the suggestion of randomizing the question sequence, since this does not seem to help but rather hinder the experts in their estimations.

The research work presented in this thesis makes use of available historical data and the estimations of experts. In the latter case, calibration of the experts and questionnaire design are crucial to avoid biased estimations. It has been found that 1) experts follow radically different estimation techniques, and 2) only experts with prior knowledge of probability concepts, could express

opinions on correlations. Based on these findings, recommendations on probability- and correlation estimations have been given.

4.4 Effectiveness and Advantages of the Proposed Uncertainty Model

In this section, the uncertainty model is compared against the causes and cures for cost underestimation presented in the literature review in chapter 2. Specifically, the effectiveness of the uncertainty model regarding the causes of cost underestimation suggested by Flyvbjerg et al. (2002) and regarding the cost escalation factors presented by Anderson et al. (2007) (sections 2.3 and 2.4) are described. Also, the uncertainty model is compared with the cost estimation practice predominant in transportation agencies (unidentified contingency) and with more advanced cost estimation techniques, such as reference class forecasting and the Cost Estimate Validation Process (sections 2.3 and 2.4). To conclude, the overall advantages of the uncertainty model are presented.

The uncertainty model satisfactorily addresses two out of three causes (technical and psychological but not economic-political) for the frequent cost underestimation suggested by Flyvbjerg et al. (2002) (section 2.3). The technical cause focuses on technical factors causing cost underestimation, such as imperfect estimation techniques, inadequate data, and inexperienced forecasters. By being a more advanced estimation technique than the available ones, the uncertainty model addresses the technical cause: it considers three sources of uncertainty (variability, cost correlations, and disruptive events; see section 4.2), and models them quantitatively with Monte Carlo simulations. It also addresses the problem of inadequate data and inexperienced forecasters: it uses historical data collected from similar projects and, when these are missing, it supplements them with estimations by experts (section 4.3). The psychological cause by Flyvbjerg et al. (2002) relates this cause of cost underestimation mostly to the optimism bias of the planners, that is the tendency by the planners to overestimate benefits and the ability to find solutions while underestimating costs and the occurrence of problems. The uncertainty model addresses optimism bias by taking an "outside view" (Lovallo & Kahneman 2003, Flyvbjerg 2006): in the outside view the project is not analyzed in its details but rather its cost and time are estimated based on data from similar projects. The uncertainty model generates its cost and time distributions based on historical data

from past projects and, when these are not available, from the estimations of experts. These do not know the project in detail but have a long experience in the field of transportation infrastructure construction. The experts must go through a calibration process before they estimate the required data (section 4.3). On the other hand, the uncertainty model does not offer any tool to counteract the economic-political cause for cost underestimation suggested by Flyvbjerg et al. (2002): it assumes that the available data are representative and that experts will act conscientiously and not follow some economic-political interests.

Out of the 11 internal factors and the seven external factors that cause cost underestimation according to Anderson et al. (2007) (section 2.4), the cure in the form of a risk strategy can address two external factors (poor estimations and inconsistent application of contingencies) and three internal factors (effects of inflation, unforeseen events, and unforeseen conditions). The uncertainty model addresses the two external factors and two of three internal factors. It addresses poor estimations since it can improve cost and time estimates by identifying and analyzing different sources of uncertainty. It addresses the inconsistent application of contingencies by presenting the results of the uncertainty analysis with a probability distribution rather than a single number. The uncertainty model also addresses two internal factors: unforeseen events and unforeseen conditions. It models these with the uncertainty sources of variability and uncertain events. In its current form, the uncertainty model does not address the effects of inflation. However, this could be included in the uncertainty model relatively easily. The uncertainty model also considers one external factor that is usually addressed by the integrity strategy (section 2.4): bias, or purposeful underestimation of project costs. It attempts to counteract this by selecting experts that do not know the project in detail and do not have stakes in it. Also, it addresses optimism bias by taking the outside view and with debiasing techniques such as the calibration of experts prior to estimation.

The uncertainty model offers several advantages, compared to the practice in most transportation agencies, to reference class forecasting and to the Cost Estimate Validation Process (CEVP) (sections 2.3 and 2.4).

Different from the practice in most state agencies that apply contingency percentages to cover some unknown risks (section 2.4), the uncertainty model represents a significant improvement: sources of uncertainty are identified and modeled quantitatively and, if historical data cannot be collected, experts in the field are asked to formally estimate such uncertainties.

Regarding reference class forecasting (section 2.3), the uncertainty model has adopted one of its cornerstones: the outside view. Additionally, it includes expert estimation, which considers the first caveat of reference class forecasting: the possible lack of historical data. The second caveat of reference class forecasting (extreme outcomes cannot be captured) is also considered in the uncertainty model since the probability distributions of cost and time include scenarios from the least probable ones (extremes) to the most probable one (mode of the distribution).

Finally, the uncertainty model can be considered an advancement of the Cost Estimate Validation Process (CEVP) (section 2.4). The first difference between the two methodologies is in the calculation of the project cost: while in the CEVP the risk cost (the probability distribution of a risk) is added to the base cost, in the uncertainty model the Monte Carlo simulation generates probability distributions for both the variability (the uncertainty of the base cost) and the disruptive event (corresponding to the risk cost in the CEVP). Thus, the uncertainty model can model the cumulated effect of variability and disruptive events. Second, a significant advancement of the uncertainty model is its capability to systematically model correlations between activity costs and their impact on the total cost distribution. Third, an important difference between the two methodologies is the view point: the CEVP takes the inside view (experts are made to learn the details of the project to estimate risk events), while the uncertainty model takes the outside view, i.e. experts are informed on the use of the data for project cost estimation but they are not informed on the details of the project, and are asked to judge uncertainties solely based on their experience.

Overall, the uncertainty model marks several advancements and makes several contributions to the field of cost and time estimation of infrastructure construction projects. It gives a clear definition of uncertainty, it covers positive and negative aspects of uncertainty, it improves the quality of the data used, it quantitatively analyzes the uncertainties, it is useful beyond just cost and time estimation, and it is tailor-made for linear/networked projects such as rail lines, the infrastructure construction project type most plagued by cost underestimation (Flyvbjerg et al. 2002). All these aspects are discussed in the next paragraphs.

The uncertainty model defines uncertainty as a condition whose outcome is not known and where the outcome can be positive or negative. This positive and negative description characterizes all three sources of uncertainty, that is variability, cost correlations and disruptive events. Variability

is modeled with a probability distribution: for a cost distribution, a value below the mean is a positive outcome (the activity costs less than the mean cost) while a value above the mean is a negative outcome (the activity costs more than the mean cost). For cost correlations, positive correlations increase the standard deviation of the total costs (negative outcome), while negative correlations decrease the standard deviation of the total costs (positive outcome). A disruptive event is modeled with a probability of occurrence and the probability distributions of the cost and time impacts if the event occurs. The disruptive event can be positive or negative: in the example in section 4.1, a sudden water inflow in a tunnel was the negative disruptive event, while the delivery of more resistant disks for the tunnel boring machine was the positive disruptive event. With the uncertainty model, uncertainties with negative as well as positive outcomes can be modeled with variability, cost correlations and disruptive events.

In the uncertainty model, uncertainties are modeled using two data sources that ensure the quality of the input data: historical data and expert opinions. Historical data stem from comparable projects while estimations are sought from experts with long experience in their fields. Modeling the uncertainties requires two steps: identification and quantification. Uncertainties are identified from the historical data of past projects and through brainstorming of the experts. They are quantified depending on the uncertainty type: variability is quantified by defining the probability distribution (lognormal, triangular, etc.) and the distribution parameters (minimum, mode and maximum, or mean and standard deviation, etc.); cost correlations are quantified with correlation coefficients and the matrix type (identity matrix, diagonal band matrix, etc.); a disruptive event is quantified with its probability of occurrence and the probability distributions of the cost impact and time impact (type and distribution parameters). The uncertainty model prevents biases in the quantification of the uncertainties with four main tools. First, historical data are not biased. Second, expert opinions are sought from individuals who do not know the project in detail and do not have stakes in it. Third, the experts are made to take the outside view, i.e. they draw their estimates from past experience. Fourth, the experts undergo a calibration process to inform them about biases and to aim at preventing them from giving biased estimations. Thus, historical data and expert opinions ensure the quality of the data used to quantify the uncertainties of the construction process.

The uncertainty model quantitatively analyzes the overall uncertainty of the construction project with a Monte Carlo simulation: 1) it calculates the distributions of the total cost and the total

time, 2) it aggregates uncertainties to model their cumulative impact on total cost and total time, 3) it visualizes the project cost and time, and 4) it gives insight on which uncertainties have the largest impact on project cost and time. The uncertainty model calculates the distributions of the total cost and the total time. From these, the following information on total cost and total time can be read: the range, the mode (most probable total cost and total time), the extremes (if interested in worst case scenarios), and specific percentiles of the distributions. The uncertainty model aggregates uncertainties to model the cumulative impact of all variabilities, of all cost correlations, and of all disruptive events. It visualizes the project cost and time in a cost-time scattergram that illustrates the cost-time relationship. Finally, it offers insight on the impact of the different uncertainties: e.g. the disruptive event with the largest impact is prioritized compared to other disruptive events with smaller impacts.

Besides cost and time estimation and evaluation of uncertainty impacts, the quantitative analysis has further uses. Its results are the starting point for a mitigation strategy, which will focus on the uncertainties with the largest impact. Its results are the starting point also for budget allocation, which is based on percentiles of the total cost distribution. The quantitative analysis can be used again during the rest of the project to analyze new uncertainties, the effectiveness of countermeasures to mitigate threats, or to take into consideration that some uncertainties have been eliminated. In principle, it can be used iteratively during the construction of the project to model the changing uncertainties.

Differently from other tools developed for construction projects in general, the uncertainty model is particularly well suited for projects characterized by complexity. This complexity is given by the different types of structures, the number of interconnected activities, the sources of uncertainty. The uncertainty model has been successfully applied to four alignments of a new rail line (chapter 7), where each consists of more than 70 structures, numerous activities (more than 300 activities for a sequence of cuts and embankments), variable costs and times, correlated costs and different disruptive events for each type of structure. Hence, the uncertainty model is well suited for complex linear or networked projects such as a rail line. Since rail lines are the transportation infrastructure projects most plagued by cost underestimation (section 2.2), the uncertainty model can be of great help in improving the estimation of the costs of such projects.

In the preceding, the effectiveness of the uncertainty model has been measured against the causes

of cost underestimation, and the advantages of the uncertainty model compared with available estimation techniques. Finally, the contributions of the uncertainty model have been highlighted.

4.5 Conclusions

An uncertainty model based on three sources of uncertainty (variability, cost correlations, and disruptive events) has been proposed. Besides available historical data, it makes use of expert opinions, which are gathered with questionnaires. The uncertainty model is effective against the causes of cost underestimation and it offers several advantages compared to current estimation techniques.

Chapter 5

Correlations in Rail Line Construction

The uncertainty model proposed in chapter 4 represents cost correlations among other sources of uncertainty. After reviewing correlation measures and correlation models, the author investigated the cost correlations that occur in rail line construction. This chapter describes different types of cost correlations in rail line construction and analyzes their impact on the total cost distribution. The correlation between costs of a repeated activity has a dramatic impact on the standard deviation of the total cost while the correlation between costs of different activities has a limited impact. On the other hand, the correlation between cost and time of an activity does not impact the total cost distribution. Due to the extremely large impact on the total cost distribution, the correlation between costs of a repeated activity is further pursued and modeled as a source of uncertainty in the uncertainty model.

The detailed structure of the chapter is as follows: first, examples of cost underestimations in building construction due to correlation are given (section 5.1). Second, correlation measures and correlation models are reviewed (sections 5.2 and 5.3). Then, cost correlations in rail line construction are identified and their impacts on the total cost analyzed in a viaduct case study and a tunnel case study (section 5.4).

5.1 Examples of Cost Underestimation in Building Construction due to Correlation

In construction, costs are often positively correlated. If one assumes the costs are independent (i.e. uncorrelated) when they are in fact positively correlated, the standard deviation of the sum of the costs is underestimated (probability concepts are explained in Appendix B, section B.1). Examples from the literature (Newton 1992, Touran 1993) of underestimation of the total cost standard deviation in building construction are reported in the following.

Touran (1993) gives an example of the underestimation of total cost standard deviation. He considered 25 projects of low-rise office buildings constructed in Massachusetts between 1981 and 1983, for which three cost elements (electrical, mechanical, and cost of roofing, insulation and waterproofing) are known. The standard deviation of the sum of the cost elements was calculated. Each cost element was assumed to be lognormally distributed. A 5,000-sample simulation was run assuming independence among variables (independent variables are uncorrelated variables; more details in Appendix B, section B.2), and the standard deviation of the simulation was calculated. The simulation underestimates the standard deviation by 30% (Table 5.1).

Table 5.1: Comparison of sample and simulation standard deviations of construction costs (Touran, 1993). Assuming independence in the simulation caused the underestimation of the construction cost standard deviation.

	Standard deviation [\$/ft ²]
sample	7.4
simulation	10.5

Another example of underestimation of the total cost standard deviation is given by Newton (1992). This sample consists of eight high quality office buildings built between 1985 and 1990 in Sidney. The projects were broken down into 19 cost elements. Again, the standard deviation of the sum was calculated. A triangular distribution was assumed to be the underlying distribution of the cost elements. Based on the data from the eight projects an expert was asked to estimate the minimum, the mode and the maximum values of each cost item distribution. A 500-sample simulation was run assuming independence among variables, and the standard deviation of the simulation was calculated. The simulation underestimates the standard deviation by 70% (Table 5.2).

Table 5.2: Comparison of sample and simulation standard deviations of construction costs (Newton, 1992). Assuming independence in the simulation caused the underestimation of the construction cost standard deviation.

	Standard deviation [AUD/m ²]
sample	132
simulation	468

From the two examples one can see that assuming independence yields an optimistic assessment of the total cost standard deviation. The difference in the underestimation of the standard deviation, 30% in the example by Touran (1993) and 70% in the example by Newton (1992), is due to the different magnitude of correlations between the cost elements in the two examples.

Two examples on the underestimation of the standard deviation of the sum of cost variables in building construction have been given. In order to capture the increase in standard deviation, the correlation between the costs must be modeled. In the following, correlation measures and correlation models are presented.

5.2 Correlation Measures

A correlation measure needs to be selected in order to model correlations between cost variables. Here, after a brief explanation of general requirements for a correlation matrix to be valid three correlation measures are presented and their pitfalls and advantages discussed.

When two random variables are correlated, the degree of correlation is measured with a correlation coefficient. A correlation matrix consists of correlation coefficients, ρ_{ij} , between random variables X_i and X_j , $i \leq n$, $j \leq n$, n random variables. These are requirements for the correlation matrix to be valid:

- $\rho_{ij} = 1$ for $i = j$, i.e. the entries in the matrix' diagonal are ones because the correlation between a random variable and itself is a perfect positive correlation.
- The correlation matrix is symmetric, i.e. $\rho_{ij} = \rho_{ji}$. This is due the fact that the correlation between random variable X_i and random variable X_j is equal to the correlation between random variable X_j and random variable X_i .
- correlation $\rho_{ij} = [-1, 1]$ because correlations are defined within this range.

The most commonly used correlation coefficient is the Pearson correlation coefficient, which measures the linear dependence between random variables. Other correlation coefficients are the Spearman and Kendall's tau correlation coefficients, which are correlation measures based on ranks. Pearson, Spearman and Kendall's correlation coefficients are discussed in the following.

The Pearson correlation coefficient, aka product moment correlation coefficient or linear correlation coefficient, is a measure of the strength of the linear dependence between random variables. The Pearson correlation coefficient between two random variables X and Y is defined as (see e.g. Ang & Tang 2007)

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (5.1)$$

where μ_X , μ_Y , σ_X and σ_Y are, respectively the means and the standard deviations of X and Y . Based on a set of samples (x_i, y_i) from X and Y , the Pearson correlation coefficient can be estimated by (see e.g. Ang & Tang 2007)

$$\hat{\rho} = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}, \quad (5.2)$$

where n is the number of samples, \bar{x} and \bar{y} are, respectively, the sample mean of X and Y estimated by (see e.g. Ang & Tang 2007)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (5.3)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad (5.4)$$

and s_X and s_Y are, respectively, the sample standard deviations of X and Y estimated by (see e.g. Ang & Tang 2007)

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (5.5)$$

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (5.6)$$

The value of ρ ranges from -1 and $+1$. If the estimated ρ is close to -1 or $+1$, there is strong linear relationship between X and Y . On the other hand, $\rho \simeq 0$ indicates a lack of linear relationship

between the variables. 1,000 samples from random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ are shown in Figure 5-1 with correlations $\rho = -0.9$, $\rho = 0$, and $\rho = 0.9$, respectively.

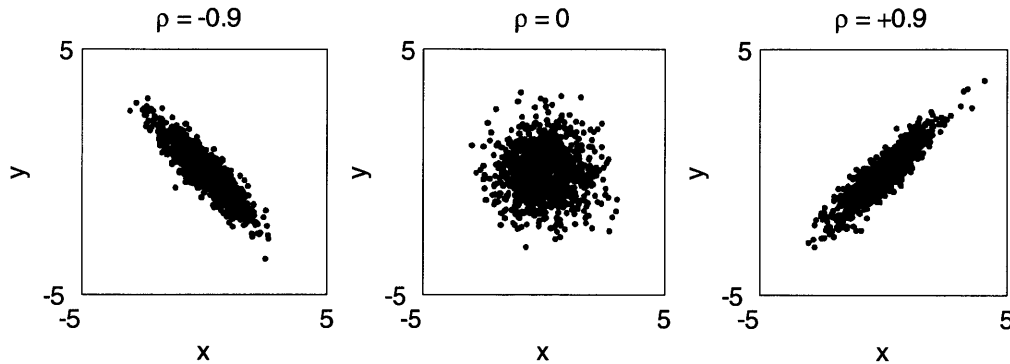


Figure 5-1: 1,000 samples from Gaussian distributed random variables $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ with different Pearson correlation coefficient, ρ . The value of this can vary between -1 and $+1$.

The Spearman correlation coefficient, aka rank correlation coefficient, is a measure of the strength of the monotonic dependence between random variables. The Spearman correlation coefficient between two random variables X and Y with cumulative distribution functions F_X and F_Y is defined as (see e.g. Embrechts et al. 2003)

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y)), \quad (5.7)$$

where ρ_r is the Spearman correlation coefficient, and ρ is the Pearson correlation coefficient. Based on n samples (x_i, y_i) from X and Y , the Spearman correlation coefficient can be estimated as (see e.g. Kurowicka & Cooke 2006)

$$\hat{\rho}_r = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}, \quad (5.8)$$

where R_i is the rank of x_i among the other x 's, S_i is the rank of y_i among the other y 's, and

$$\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i, \quad (5.9)$$

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i. \quad (5.10)$$

The value of ρ_r ranges from -1 and $+1$. 1,000 samples (x_i, y_i) of random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ are shown in Figure 5-2 with correlations $\rho_r = -0.9$, $\rho_r = 0$, and $\rho_r = 0.9$, respectively. The ranks, R_i and S_i , of the samples (x_i, y_i) are shown in Figure 5-3.

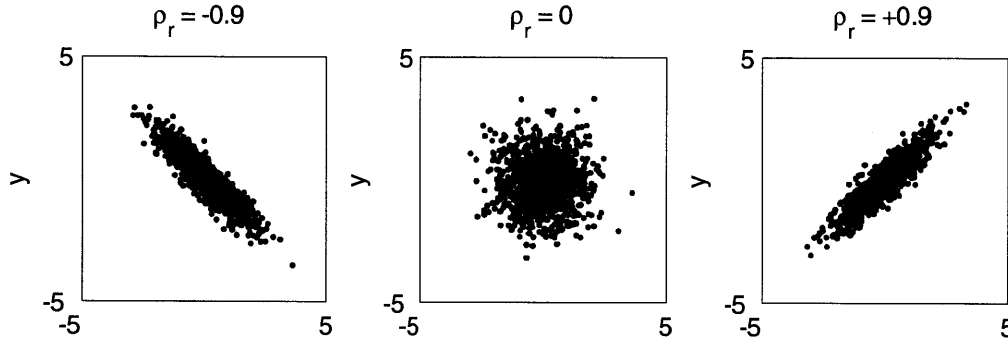


Figure 5-2: 1,000 samples from Gaussian distributed random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ with different Spearman correlation coefficient, ρ_r .

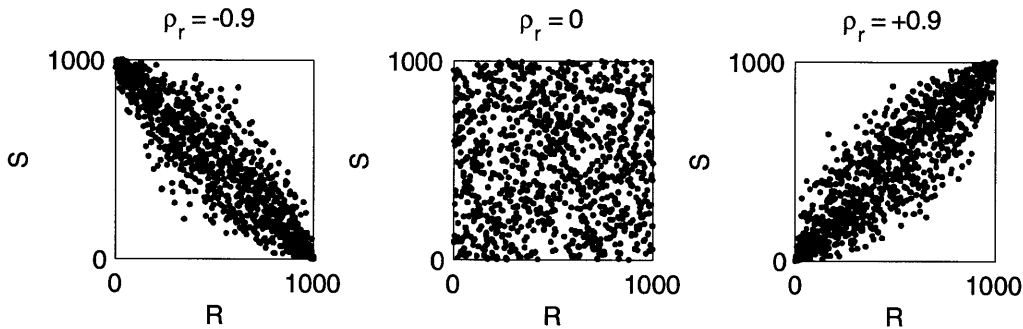


Figure 5-3: Ranks of 1,000 samples from Gaussian distributed random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ with different Spearman correlation coefficient, ρ_r .

The expressions to estimate the Pearson and the Spearman correlations show the differences between the two correlations. By combining expressions 5.2, 5.5 and 5.6, one obtains the following expression to estimate the Pearson correlation coefficient:

$$\hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (5.11)$$

The above expression can be compared with expression 5.8 to estimate the Spearman correlation

coefficients. As one can see the expressions to estimate the Pearson correlation and the Spearman correlation have identical operations, while they differ in the operands. The Pearson correlation is calculated with the samples (x_i, y_i) drawn from (X, Y) , whereas in the expression to calculate the Spearman correlation samples (x_i, y_i) are substituted by the samples' ranks (R_i, S_i) .

Kendall's tau is another measure of dependence based on ranks. Given random variables X and Y and an independent copy of random variables X and Y , \tilde{X} and \tilde{Y} , Kendall's tau is defined as (see e.g. Embrechts et al. 2003)

$$\tau(X, Y) = \mathbf{P} \left\{ (X - \tilde{X}) (Y - \tilde{Y}) > 0 \right\} - \mathbf{P} \left\{ (X - \tilde{X}) (Y - \tilde{Y}) < 0 \right\}. \quad (5.12)$$

From the expression one can see that Kendall's tau is the probability of concordance minus the probability of discordance, where two pairs, (x_i, y_i) and (x_j, y_j) , are concordant when $(x_i - x_j)(y_i - y_j) > 0$, and discordant when $(x_i - x_j)(y_i - y_j) < 0$. Given X and Y are continuous, ties $(x_i - x_j)(y_i - y_j) = 0$ occur with probability zero. Based on n samples (x_i, y_i) and n samples (x_j, y_j) , Kendall's tau can be estimated as (see e.g. Genest & Favre 2007)

$$\tau = \frac{P_n - Q_n}{\binom{n}{2}}, \quad (5.13)$$

where P_n and Q_n are the number of concordant and discordant pairs, respectively.

For the case of the bivariate normal distribution, direct relationships 1) between the Spearman correlation coefficient, ρ_r , and the Pearson correlation coefficient, ρ , and 2) between Kendall's tau, τ , and the Pearson correlation coefficient, ρ , exist (Kruskal 1958):

$$\rho_r = \frac{6}{\pi} \arcsin \left(\frac{\rho}{2} \right), \quad (5.14)$$

$$\tau = \frac{2}{\pi} \arcsin(\rho). \quad (5.15)$$

The relationships ($\rho - \rho_r$ dashed, $\rho - \tau$ dotted) are drawn in Figure 5-4. The diagonal (solid line) is depicted for reference to show the difference between the two rank correlations and the Pearson correlation.

The Spearman correlation is preferred over Kendall's tau for two reasons. First, the definition of Kendall's tau is a rather abstract concept (difference between the probability of concordance and

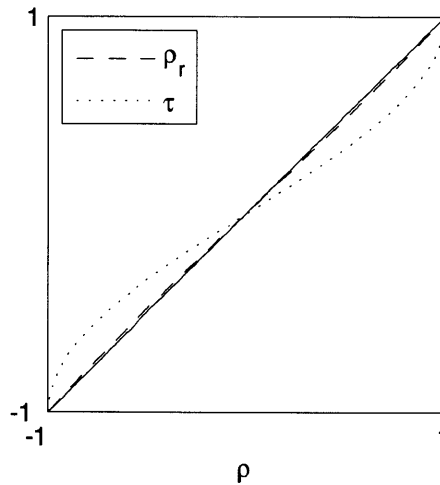


Figure 5-4: Spearman correlation coefficient, ρ_r (dashed), and Kendall's Tau, τ (dotted), as a function of Pearson correlation coefficient, ρ , for the bivariate normal distribution. For comparison the equality $\rho = \rho_r = \tau$ is shown (solid).

the probability of discordance). Second, the Spearman correlation is distinct from but similarly defined as the Pearson correlation: the Pearson is the correlation between variates, the Spearman is the correlation between the ranks of the variates. The relation between the Pearson correlation and the Spearman correlation of two random variables is thus explicit.

The Spearman correlation is preferred over the Pearson correlation since, with non-normal distributions, it is a more meaningful measure of dependence than the Pearson correlation. Embrechts et al. (2003) point out that although the Pearson correlation is a natural measure of dependence in elliptical distributions, such as the multivariate normal and the multivariate t-distribution, random variables are often not jointly elliptically distributed, so that using the Pearson correlation as a measure of dependence in these cases might be misleading. Embrechts et al. (2003) also suggest the use of rank correlations, such as the Spearman correlation, as the measure of dependence for non-elliptical distributions.

The preference of the Spearman correlation over the Pearson correlation is also motivated by the pitfalls of the Pearson correlation coefficient together with the advantages of the Spearman correlation coefficient. These are listed in the following.

The first pitfall of the Pearson correlation coefficient, ρ , is its range. In theory, the Pearson

correlation coefficient ranges from -1 to $+1$. In reality, depending on the probability distributions of the random variables, the actual range of ρ is a subset of the interval $[-1, +1]$ (examples 5.1 and 5.2). This is a considerable obstacle when one works with expert estimations because an expert can estimate a correlation, which for the probability distributions of the random variables lies outside the range of ρ .

Example 5.1 Let $X \sim \text{Lognormal}(0, 1)$ and $Y \sim \text{Lognormal}(0, \sigma^2)$, $\sigma > 0$. Then, $\rho_{\min} = \rho(e^Z, e^{-\sigma Z})$ and $\rho_{\max} = \rho(e^Z, e^{\sigma Z})$, where $Z \sim \mathcal{N}(0, 1)$. ρ_{\min} and ρ_{\max} can be calculated analytically:

$$\rho_{\min} = \frac{e^{-\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}, \quad (5.16)$$

$$\rho_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e - 1)(e^{\sigma^2} - 1)}}. \quad (5.17)$$

These minimum and maximum correlations are shown graphically in Figure 5-5. One can observe that $\lim_{\sigma \rightarrow \infty} \rho_{\min} = \lim_{\sigma \rightarrow \infty} \rho_{\max} = 0$ (Embrechts et al. 2002).

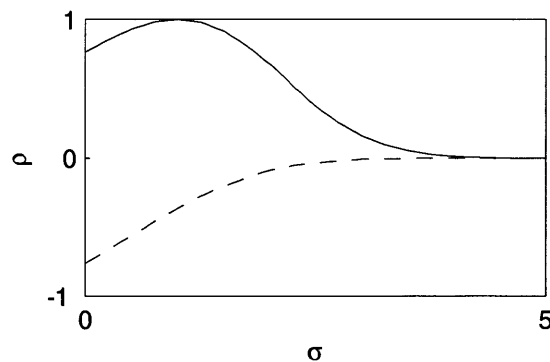


Figure 5-5: Minimum (dashed) and maximum (solid) Pearson correlation coefficients for $X \sim \text{Lognormal}(0, 1)$ and $Y \sim \text{Lognormal}(0, \sigma^2)$, $\sigma > 0$. For the given distributions, the actual range of ρ is a subset of the interval $[-1, +1]$.

Example 5.2 Let $X \sim \text{Gamma}(\alpha_1, \beta_1)$ and $Y \sim \text{Gamma}(\alpha_2, \beta_2)$. The shape parameters α_1 and α_2 , and the scale parameters β_1 and β_2 are real numbers. The value of the Pearson correlation coefficient ρ does not depend on the scale parameters β_i . The range of the Pearson correlation

coefficient is more limited than $-1 \leq \rho \leq +1$ (Figure 5-6). In fact, only when $\alpha_1 = \alpha_2$ is it possible for $\rho = +1$ (Schmeiser & Lal 1982).

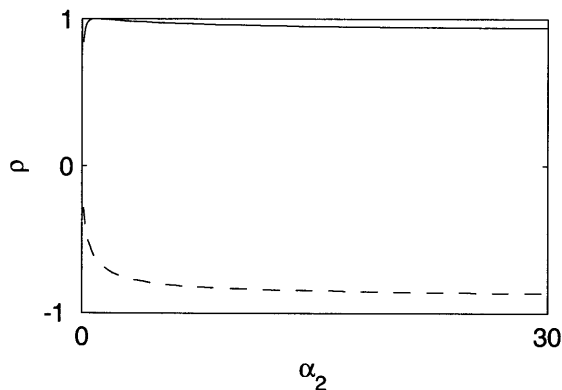


Figure 5-6: Minimum (dashed) and maximum (solid) Pearson correlation coefficient as a function of the distribution parameter α_2 , for $\alpha_1 = 1$, for a bivariate gamma distribution. For the given distributions, the actual range of ρ is a subset of the interval $[-1, +1]$.

The second pitfall of the Pearson correlation coefficient is that it is not invariant under non-linear strictly increasing transformations $T : \mathbb{R} \rightarrow \mathbb{R}$ (Li & Hammond 1975, Cario & Nelson 1997, Lurie & Goldberg 1998, Embrechts et al. 2002). For two random variables (X, Y) in general

$$\rho(X, Y) \neq \rho(T(X), T(Y)). \quad (5.18)$$

Since the cumulative distribution function is a non-linear strictly increasing transformation, for two random variables X and Y with cumulative distribution functions F_X and F_Y in general

$$\rho(X, Y) \neq \rho(F_X(X), F_Y(Y)). \quad (5.19)$$

This is relevant because some correlation models (section 5.3) require non-linear strictly increasing transformations. This means that random samples with the desired Pearson correlations cannot be generated with these methods. In fact, if the desired Pearson correlations are input in these methods, the random samples generated with these methods will generally not have the desired Pearson correlations. Despite this pitfall, correlation models have been developed to generate samples with the desired Pearson correlations (Appendix B, section B.3).

The first advantage of the Spearman correlation coefficient, ρ_r , is its range. In contrast to the Pearson correlation coefficient, the Spearman correlation coefficient, ρ_r , ranges from -1 to $+1$ independent of the probability distributions of random variables X and Y . The entire range of Spearman correlation coefficients between -1 and $+1$ exist given any marginal distributions F_X and F_Y for X and Y . Thus, it is guaranteed for the expert estimation to lie within the range of ρ_r .

The second advantage of the Spearman correlation coefficient is that it is invariant under non-linear strictly increasing transformations $T : \mathbb{R} \rightarrow \mathbb{R}$. For two random variables X and Y

$$\rho_r(X, Y) = \rho_r(T(X), T(Y)) \quad (5.20)$$

so that for two random variables X and Y with cumulative distribution functions F_X and F_Y

$$\rho_r(X, Y) = \rho_r(F_X(X), F_Y(Y)). \quad (5.21)$$

Thus, the correlation models that require non-linear strictly increasing transformations (section 5.3) can generate random samples with desired Spearman correlations. This is explained with the following example, adapted from Genest & Favre (2007).

Example 5.3 *Ten pairs are drawn from random variables $X \sim \text{Beta}(5, 5)$ and $Y \sim \text{Lognormal}(0, 1)$. The graphs in Figure 5-7 show*

- *Top left: the samples (x_i, y_i) .*
- *Top right: the cumulative distributions of the samples $(F_X(x_i), F_Y(y_i))$.*
- *Bottom left: the ranks (R_i, S_i) of the samples (x_i, y_i) .*
- *Bottom right: the ranks (T_i, V_i) of the cumulative distribution of the samples $(F_X(x_i), F_Y(y_i))$.*

From the graphs one can see that the Pearson correlation coefficients of the values (top graphs) will be different because $\rho(X, Y) \neq \rho(F_X(X), F_Y(Y))$, whereas the Pearson correlation coefficients of the ranks (bottom graphs), i.e. the Spearman correlation coefficients of the values, will be identical because $\rho_r(X, Y) = \rho(R, S) = \rho(V, T) = \rho_r(F_X(X), F_Y(Y))$.

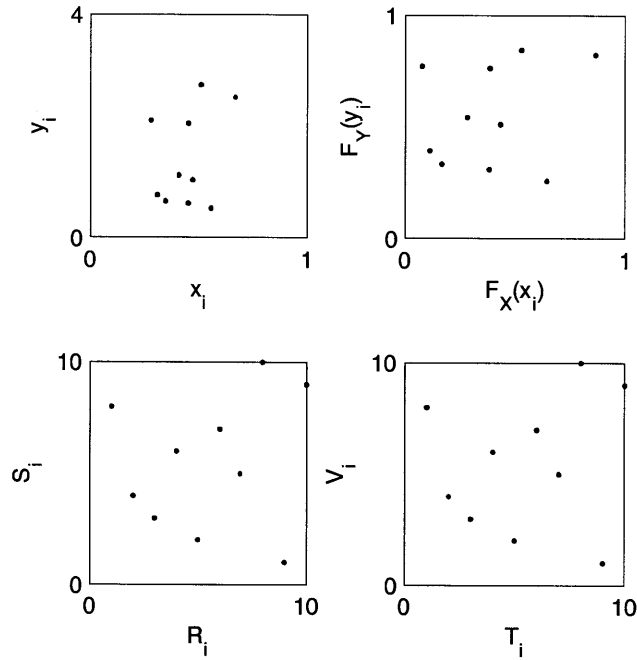


Figure 5-7: Given random variables $X \sim \text{beta}(5, 5)$ and $Y \sim \text{lognormal}(0, 1)$, depicted are the samples (x_i, y_i) (top left), the cumulative distributions of the samples $(F_X(x_i), F_Y(y_i))$ (top right), the ranks (r_i, s_i) of the pairs (x_i, y_i) , (bottom left), and the ranks (v_i, t_i) of the pairs $(F_X(x_i), F_Y(y_i))$ (bottom right). Differently from the Pearson correlation coefficient, the Spearman correlation coefficients of the values are the same since $\rho_r(X, Y) = \rho(R, S) = \rho(V, T) = \rho_r(F_X(X), F_Y(Y))$.

The Spearman correlation coefficient is chosen as measure of correlation to be used in the methods to model correlations in Monte Carlo simulations for the two reasons explained above, and reiterated here:

1. the Spearman correlation coefficient, ρ_r , ranges from -1 to $+1$ independent of the probability distributions of random variables so that it is guaranteed for the expert estimation to lie within the range of ρ_r .
2. the Spearman correlation coefficient is invariant under non-linear strictly increasing transformations so that random samples with desired Spearman correlations can be generated with the available methods to model correlations.

In this section, three correlation measures (Pearson, Spearman and Kendall's correlations)

have been presented. Due to its advantages when modeling correlated variables, the Spearman correlation has emerged as the best suited correlation measure. In the next section, the available correlation models are discussed.

5.3 Correlation Models

When modeling correlations between cost variables in rail line construction, a model of correlation and a measure of correlation are needed. The Spearman correlation coefficient has been identified as the best suited correlation measure. In this section, eight correlation methods are presented and compared against three requirements by the uncertainty model (chapter 4). One correlation method will emerge as the best suited to method correlation.

The correlation method must fulfill three requirements by the uncertainty model:

1. It must generate samples from correlated variables with the desired correlation and with desired marginal distributions.
2. Ideally, it should be distribution-free, i.e. should allow any marginal distribution. In particular, the lognormal and the triangular distribution must be distributions that can be modelled, since these are the distributions that model cost and time variables in the uncertainty model.
3. The number of variables, which are correlated at one time, should not be limited.

Eight correlation methods have been identified in the literature:

1. the limitation of disaggregation,
2. the pessimistic/optimistic approach,
3. sampling from joint distribution,
4. the Central Limit Theorem (CLT),
5. the lognormal marginal method,
6. Iman and Conover's method,
7. NORTA,

8. copulas.

In the following, these correlation methods are presented, and the advantages and disadvantages of each method are outlined. The correlation method best suited to model correlations in the rail line construction is then selected.

Limitation of Disaggregation and Pessimistic/Optimistic Approach

Pouliquen (1970) suggested two pragmatic approaches to correlation. The first method is centered on limiting the level of disaggregation. The level of disaggregation determines the degree of detail of the analysis, e.g. the cost of a rail line open track can be subdivided into the cost of clearing, earthwork, ballast, sub-ballast, and rail tracks. The cost of e.g. the ballast could be further disaggregated. Correlations can be avoided by limiting disaggregation, e.g. by working with the total cost of a rail line open track, the correlation between the cost of the ballast and the cost of the sub-ballast is irrelevant. Where to stop disaggregating requires judgment on two criteria. On the one hand, the more detailed the cost components, the higher the risk of correlations among cost components. On the other hand, a more detailed level (down to a certain limit) helps the estimator formulating a judgment on the cost.

The second method (Pouliquen 1970) is centered on the optimistic and/or pessimistic scenarios. If one suspects two variables to be correlated but the correlation cannot be quantified, a pessimistic scenario of the suspected correlation can be examined. For example if two variables are suspected to be positively correlated, a simulation with a correlation of +1 between the two variables can be run. Assuming positive correlation between two costs will increase the total cost of the project. Therefore, if the project is still acceptable, then a decision can be taken nonetheless.

The disadvantage of the first method is that it does not model correlations but rather identifies scenarios where correlations are not relevant. The second method allows one to model only extreme scenarios of correlations. Both approaches are not viable for a large number of variables and complex systems such as a rail line construction.

Sampling from Joint Distribution

Samples from correlated random variables with the desired Spearman correlation matrix and desired marginal distributions can be generated from a joint distribution. This method consists of two steps:

first the joint distribution is calculated, then the samples are drawn from the joint distribution.

The conventional approach to specifying a joint distribution of random variables is as the product of marginal and conditional distributions. The joint, the conditional and the marginal probability density functions (pdf's) of random variables X and Y are related to each other by the formulas (see e.g. Bertsekas & Tsitsiklis 2002)

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x) = f_Y(y) f_{X|Y}(x|y). \quad (5.22)$$

The expression 5.22 can be extended to the n -dimensional joint distribution

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = f_{X_n}(x_n) f_{X_1,\dots,X_{n-1}|X_n}(x_1,\dots,x_{n-1}|x_n). \quad (5.23)$$

An alternative approach is constructing the joint distribution using a copula, e.g. the normal copula (more on copulas on page 164). This approach requires only the random variables' marginal distributions, and a measure of dependence among the random variables. The random variables' joint distribution is calculated as (Clemen & Reilly 1999)

$$f(x_1,\dots,x_n|R_r) = f(x_1) \times \dots \times f(x_n) \times \exp\{-y^T(R^{-1} - I)y/2\} / |R|^{1/2}, \quad (5.24)$$

where

R_r is the Spearman correlation matrix,

R is the Pearson correlation matrix, calculated from R_r with expression 5.14,

I is the identity matrix,

$f_i(x_i)$ are the probability density functions (pdf's) of the univariate random variables,

$F_i(x_i)$ are the cumulative distribution functions (CDFs) corresponding to $f_i(x_i)$,

$y = (\Phi^{-1}[F_1(x_1)], \dots, \Phi^{-1}[F_n(x_n)])$, and

Φ^{-1} is the normal inverse transformation.

Once the joint distribution is calculated, samples are generated from it. Generating samples from a one-dimensional distribution is easily performed, whereas generating samples from a joint distribution is not as straightforward. There are two procedures that can generate samples from a joint distribution with any marginals without prior knowledge of the joint distribution: (1) an acceptance-rejection method, and (2) Markov-Chain-Monte-Carlo methods such as a random walk

Metropolis-Hastings algorithm or Gibbs sampling (see e.g. Robert & Casella 2004). The disadvantage of the acceptance-rejection method is that the acceptance rate decreases with increasing dimensionality making the method inefficient. Random walk Metropolis-Hastings algorithms are known to fail in converging to the desired joint distribution in large dimensions. Appropriate Metropolis-Hastings algorithms can be developed to simulate specific cases but off-the-shelf algorithms, such as a random walk Metropolis-Hastings algorithm, do not function in every possible setting. Gibbs sampling is applicable when the joint distribution is not known explicitly but the conditional distribution of each variable is available Robert & Casella (2004). Unfortunately, conditional distributions are also not known in the applications to rail line construction presented in this thesis.

Central Limit Theorem

The Central Limit Theorem (CLT) allows one to model the total cost with a normal distribution, given that the total cost is the sum of independent cost variables.

Definition 5.1 *Let X_1, \dots, X_n be a sequence of independent identically distributed random variables with common mean μ and variance σ^2 , and define*

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}. \quad (5.25)$$

Then, the cumulative distribution function of Z_n converges to the standard normal cumulative distribution function (CDF)

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx, \quad (5.26)$$

in the sense that

$$\lim_{n \rightarrow \infty} \mathbf{P}(Z_n \leq z) = \Phi(z), \quad (5.27)$$

for every z (see e.g. Bertsekas & Tsitsiklis 2002).

The distribution of the X_i can be discrete, continuous or mixed and have any shape, as long as it has finite mean and variance.

If X_1, \dots, X_n are the cost variables, $S_n = X_1 + \dots + X_n$ is the total cost. The probability $\mathbf{P}(S_n \leq c)$ can be approximated with the following procedure:

1. calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n .

2. Calculate the normalized value

$$z = \frac{(c - n\mu)}{\sigma\sqrt{n}}. \quad (5.28)$$

3. Use the approximation

$$\mathbf{P}(S_n \leq c) \approx \Phi(z), \quad (5.29)$$

where Φ is the standard normal CDF (see e.g. Bertsekas & Tsitsiklis 2002).

The CLT also indicates that the sum of a large number of independent variables is approximately normal (Bertsekas & Tsitsiklis 2002). This is an important concept in construction cost estimation where independent cost variables, but not necessarily identically distributed, are summed to obtain the total cost. Given a linear construction project subdivided into segments $i = 1, \dots, n$ of different length L_i with cost X_i with mean m_i and variance σ_i^2 , the mean and the variance of the total costs $S_n = X_1 + \dots + X_n$ are

$$m_{S_n} = \sum_{i=1}^n m_i L_i, \quad (5.30)$$

$$\sigma_{S_n}^2 = \sum_{i=1}^n \sigma_i^2 L_i^2, \quad (5.31)$$

so that probabilities of the type $\mathbf{P}(S_n \leq c)$ can be calculated with the procedure above.

The advantages of the CLT are that it allows the calculation of certain probabilities by referring to a standard normal CDF table, and only means and variances are required for such calculations. The disadvantage is that it requires independence between variables so that cases where the variables are correlated cannot be modeled.

Lognormal Marginals Method

Touran & Wiser (1992) propose a method to generate correlated cost variates with the desired Pearson correlation coefficients and with lognormal marginal distributions. The method uses the multivariate normal distributions and normal-lognormal transformations. The method by Touran & Wiser (1992) is named here the lognormal marginals method.

One wishes to generate n samples from k correlated lognormal random variables. Z is the $k \times k$ covariance matrix. The method is implemented in 5 steps:

1. Transform means $\tilde{\mu}_i$, variances $\tilde{\sigma}_i^2$, and correlation coefficients $\tilde{\rho}_{ij}$ of the lognormal distributions to means μ_i , variances σ_i^2 , and correlation coefficients ρ_{ij} of normal distributions (Johnson & Ramberg 1978):

$$\mu_i = \ln \left(\frac{\tilde{\mu}_i^2}{\sqrt{\tilde{\sigma}_i^2 + \tilde{\mu}_i^2}} \right), \quad (5.32)$$

$$\sigma_i^2 = \ln \left(1 + \frac{\tilde{\sigma}_i^2}{\tilde{\mu}_i^2} \right), \quad (5.33)$$

$$\rho_{ij} = \frac{1}{\tilde{\sigma}_i \tilde{\sigma}_j} \ln \left(1 + \tilde{\rho}_{ij} \left| \frac{\tilde{\sigma}_i \tilde{\sigma}_j}{\tilde{\mu}_i \tilde{\mu}_j} \right| \right). \quad (5.34)$$

2. Calculate the lower triangular $k \times k$ matrix X such that $Z = XX^T$.
3. Generate independently, identically distributed $N_j \sim \mathcal{N}(0, 1)$.
4. Calculate the logarithm of the cost random variates LC_i :

$$LC_i = \mu_i + \sum_{j=1}^i x_{ij} N_j \quad (5.35)$$

where x_{ij} are the entries of the lower triangular matrix X .

5. Transform the logarithm of the cost random variates, LC_i , to cost random variates, C_i :

$$C_i = e^{LC_i}. \quad (5.36)$$

The lognormal marginal method is simple to apply. However, the disadvantage of this method is that it is not distribution-free as it assumes lognormal marginal distributions. Still, there are studies in the literature showing that construction cost variables tend to be lognormally distributed (Touran & Wiser 1992, Touran & Suphot 1997).

Touran & Wiser (1992) analyzed 15 unit cost data for 1,014 low-rise office buildings, which were built between 1982 and 1992. Data were adjusted to account for cost escalation with time. Although the data stem from projects from across US, costs were not adjusted with location indexes. Fifteen major unit-cost items were considered: general and overhead; site work; concrete; masonry; metals; carpentry; moisture protection; doors, windows, and glass; finishes; specialties; equipment;

furnishings; conveying systems; mechanical; and electrical. Not all 15 cost items figured in each project; therefore, the sample sizes differed across cost items. Among the normal, lognormal, beta, and gamma distributions, the lognormal distribution fitted best according to the chi-squared test. Eleven samples were lognormal random variates at 1% level of significance.

Touran & Suphot (1997) analyzed 10 cost items of 131 low-rise office projects. The sample data are the same as for Touran & Wiser (1992) but this time the data were skimmed by considering only the buildings that included data for all of the following items: sitework; concrete; masonry; metals; carpentry; moisture protection; doors, windows, and glass; finishes; mechanical; and electrical. This time the lognormal distribution did not fit best the most sample data, as it had been the case in Touran & Wiser (1992). Four cost items followed the gamma distribution, three the lognormal distribution, one the beta distribution, one the Erlang distribution, and one the beta or the lognormal distribution equally well. However, one needs to note that the sample in Touran & Wiser's (1992) was ten times larger than the sample in Touran & Suphot (1997).

One can conclude that the lognormal distribution fits many but not all construction cost variables. Thus, when the construction cost variables follow the lognormal distribution, the lognormal marginals method is appropriate. When construction cost variables do not follow the lognormal distribution, correlation cannot be modeled with the lognormal marginals method.

Iman and Conover's Model

Iman & Conover (1982) present a method for generating correlated random variates with desired Spearman correlation matrix and desired marginal distributions.

One wishes to generate n samples from k correlated random variables. The Spearman correlation coefficients between the k random variables are given by the $k \times k$ matrix R_r . The method by Iman & Conover (1982) is implemented in 4 steps:

1. Generate an $n \times k$ matrix, X , of independent variates.
2. Generate a $n \times k$ matrix, C , whose columns represent independent permutations of an arbitrary set of n numbers. The row vector of C , C_i , has correlation matrix I , i.e. the rows are uncorrelated. In the example in Iman & Conover (1982), the van der Waerden scores $\Phi^{-1} \left(\frac{i}{n+1} \right)$ are used. Φ^{-1} is the inverse function of the standard normal distribution func-

tion. The van der Waerden scores, aka normal scores, are quantiles of the standard normal distribution.

3. Calculate the lower triangular matrix P such that $R_r = PP^T$, using e.g. the Cholesky factorization. The multiplication of $C_i P^T$ results in a vector which has the desired Spearman correlation matrix R_r . The multiplication of CP^T results in a matrix C^* , which has Spearman correlation matrix M , which is close to the desired correlation matrix R_r .
4. Rearrange the entries in each column of the matrix, X , so that they have the same order (rank) as the corresponding column of C^* . As the entries in each column of the matrix, X , have the same order (rank) as the entries in each column of the matrix, C^* , the two matrices have the same Spearman correlation matrix, which is M . As the entries in the columns of the input matrix, X , are ordered (ranked) but not changed, the input marginal distributions are preserved.

Iman & Conover's (1982) 4-step method is clarified with an example.

Example 5.4 *Two random variables $Y_1 \sim \text{Beta}(5, 5)$ and $Y_2 \sim \text{Lognormal}(0, 1)$ are correlated with Spearman correlation matrix*

$$R_r = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

1. *n independent gamma distributed variates and n independent lognormal distributed variates are generated. The two sets of variates are the first and the second columns of the matrix, X . The scatterplot of the independent samples, the frequency plots with the beta and lognormal pdf's are shown in Figure 5-8a.*
2. *An arbitrary set of n numbers is generated. Random numbers from the normal distribution $\mathcal{N}(0, 1)$ are used. The set of n numbers is randomly permuted. The scatterplot of the sets, the frequency plots with the normal pdf's are shown in Figure 5-8b.*
3. *The lower triangular matrix P is calculated:*

$$P = \begin{bmatrix} 1.0000 & 0 \\ 0.5000 & 0.8660 \end{bmatrix}.$$

$C^* = CP^T$ is calculated, as well as the Spearman correlation matrix of C^* , M :

$$M = \begin{bmatrix} 1 & 0.5094 \\ 0.5094 & 1 \end{bmatrix}$$

The scatterplot of C^* , the frequency plots with the normal pdf's are shown in Figure 5-8c.

4. Each column of X is rearranged so that it has the same order (rank) as the corresponding column of C^* . The Spearman correlation matrix of X is equal to M . The scatterplot of X , the frequency plots with the beta and the lognormal pdf's are shown in Figure 5-8d.

Iman & Conover's (1982) method has several advantages: the method is simple to apply, is distribution-free, and does not require the probability distributions to be continuous and invertible. The disadvantage is that the generated samples have only approximately the desired correlation matrix (step 3). Another disadvantage is that the method requires a search subroutine to find the rank of the entries in matrix C^* . Despite the limitations, the Iman & Conover's (1982) method is the most flexible of the methods presented here in terms of acceptable marginal distributions.

NORTA

Cario & Nelson (1997) present a method, "NORmal To Anything" (NORTA), to generate correlated random variables with desired marginal distributions and desired correlation. The NORTA method can be used with both Pearson or Spearman correlation coefficients (Cario & Nelson 1997). The method is explained here with Spearman correlation coefficients.

One wishes to generate n samples from k random variables (X_1, \dots, X_k) with continuous invertible univariate distribution functions (F_1, \dots, F_k) . The Spearman correlation coefficients between the k random variables are given by the $k \times k$ matrix R_r . The NORTA method is implemented in 3 steps:

1. Draw n samples, (y_1, \dots, y_k) , from a multivariate normal distribution with Spearman correlation matrix R_r . The Pearson correlation matrix R is calculated with expression 5.14. R must be positive definite (the concept of positive definiteness is explained in Appendix B, section B.4).

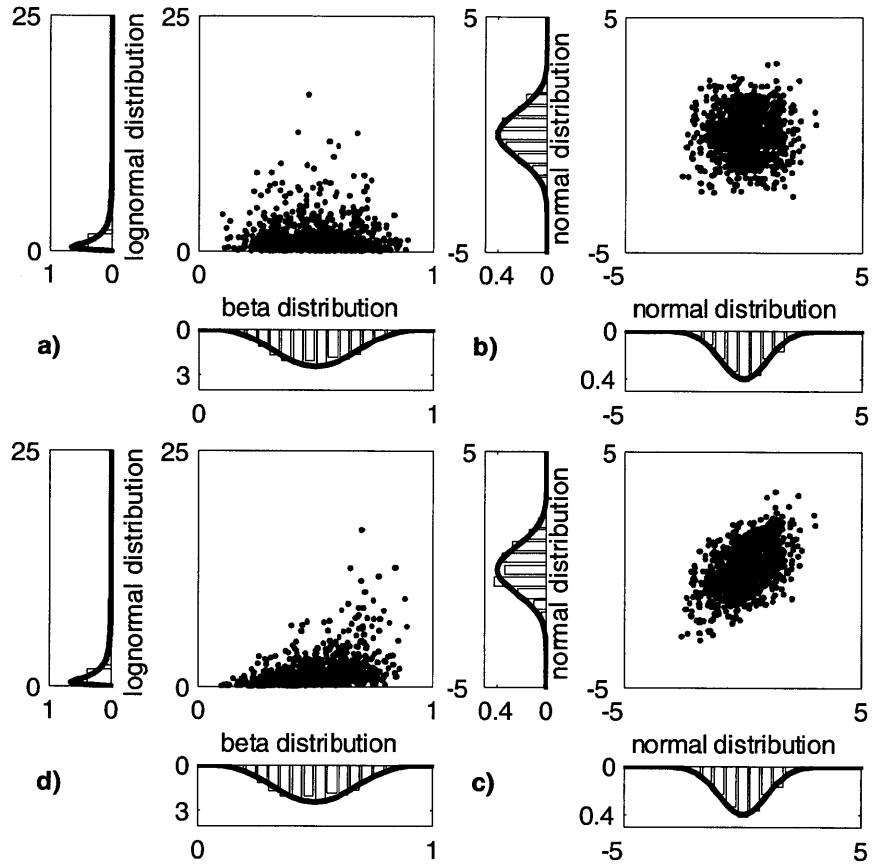


Figure 5-8: Simulation steps using Iman and Conover's correlation model to generate correlated variables with desired correlation matrix and desired marginal probability distributions. Clockwise from top left a) step 1: generate independent samples with beta and lognormal marginal distributions, b) step 2: generate independent samples with normal marginal distributions, c) step 3: generate correlated samples with normal marginal distributions, d) step 4: generate correlated samples with beta and lognormal distributions.

2. Calculate the normal cumulative distribution of the samples, $(\Phi(y_1), \dots, \Phi(y_k))$. $\Phi(y_i)$ is a uniform distribution on $[0, 1]$.
3. Apply the inverse cumulative distribution functions, $(F_1^{-1}, \dots, F_k^{-1})$, to the normal cumulative distributions of the samples, $(\Phi(y_1), \dots, \Phi(y_k))$, so that

$$(x_1, \dots, x_k) = F_1^{-1}(\Phi(y_1)), \dots, F_k^{-1}(\Phi(y_k)). \quad (5.37)$$

The random variables (X_1, \dots, X_k) have the Spearman correlation matrix R_r and the desired marginal distributions.

The 3-step method is clarified with the same example as for the Iman and Conover's model, and the results are compared.

Example 5.5 *Two random variables $X_1 \sim \text{Beta}(5, 5)$ and $X_2 \sim \text{Lognormal}(0, 1)$ are correlated with correlation matrix*

$$R_r = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

1. n samples, (y_1, y_2) , are drawn from a bivariate normal distribution with Spearman correlation matrix R_r . The Pearson correlation matrix R of (y_1, y_2) is

$$R = \begin{bmatrix} 1.0000 & 0.5176 \\ 0.5176 & 1.0000 \end{bmatrix}.$$

The scatterplot of the samples, the frequency plots with normal pdf's are shown in Figure 5-9a.

2. The normal cumulative distribution function of the samples, $(\Phi(y_1), \Phi(y_2))$, is calculated. The Spearman correlation matrix of $(\Phi(y_1), \Phi(y_2))$ is also R_r . The scatterplot of the cumulative distribution functions of the samples, the frequency plots with the uniform pdf's are shown in Figure 5-9b.
3. The correlated samples (x_1, x_2) are calculated by applying the inverse cumulative distribution functions, (F_1^{-1}, F_2^{-1}) , to the normal cumulative distributions function of the samples,

$(\Phi(y_1), \Phi(y_2))$. The Spearman correlation matrix of (x_1, x_2) is also R_r . The scatterplot of the correlated samples, the frequency plots with the beta and lognormal pdf's are shown in Figure 5-9c.

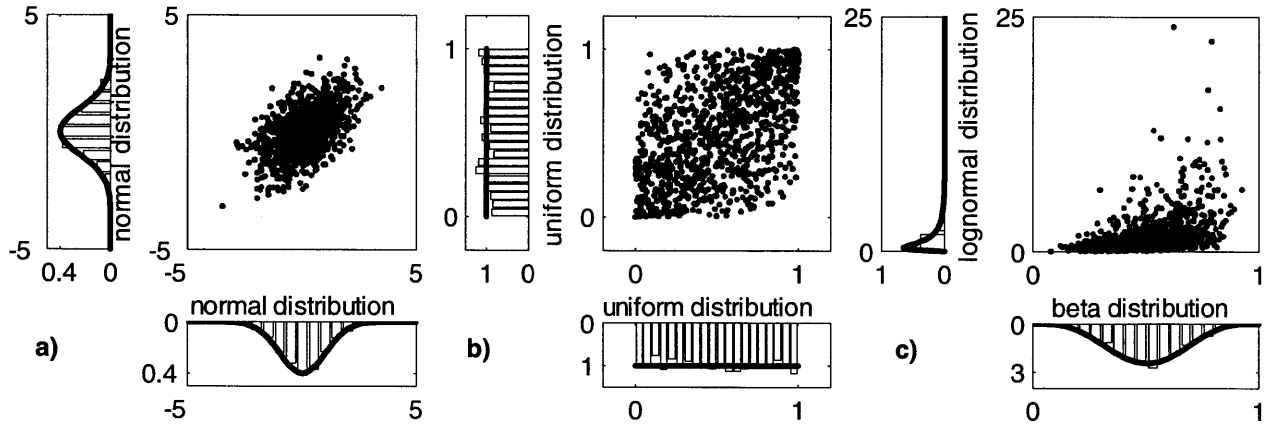


Figure 5-9: Simulation steps using the correlation model NORTA to generate correlated costs with desired correlation matrix and desired marginal probability distributions. a) step 1: generate correlated samples with normal marginal distributions, b) step 2: calculate the cumulative distributions of the correlated normal samples, that is generate correlated samples with uniform marginal distributions, c) step 3: calculate the inverse cumulative of the correlated uniform samples, that is generate correlated samples with desired marginal distributions.

NORTA requires the cumulative distribution function to be invertible (step 3). There is disagreement on whether NORTA also requires the probability distribution to be continuous. Cario & Nelson (1997) suggest that NORTA can be used with continuous, discrete and mixed marginals. Ghosh & Henderson (2002) reason that NORTA requires the cumulative distribution function to be continuous because if the distributions function F of a random variable X is not continuous, then $F(X)$ does not have a uniform distribution on $[0, 1]$, and so one will not obtain a normally distributed random variable using $\Phi^{-1}(F(X))$. From step 3 of the NORTA method, it is clear that the cumulative distribution function must be invertible. Also it needs to be continuous for the inverse F_i^{-1} of the normal cumulative distribution $\Phi(y_i)$ to be uniquely defined. If F_i^{-1} is not uniquely defined, x_i cannot be calculated. Thus, NORTA requires the probability distribution to be invertible and continuous. Note that although the triangular distribution includes a discontinuity at the mode of the function, it is in fact a continuous distribution. Thus, NORTA can be used with

triangular distributions.

In general terms, NORTA has one important advantage and one important disadvantage compared to Iman & Conover's (1982) method. The advantage is that NORTA generated samples have the desired correlation matrix. The disadvantage is that NORTA is distribution-free given two conditions: (1) the marginals are continuous, and (2) the marginals are invertible. In the comparison of examples 5.4 and 5.5, one can observe the following:

- Both generated samples are acceptable.
- Iman & Conover's (1982) method requires one step more than NORTA.
- NORTA has a more straightforward algorithm, whereas Iman & Conover's (1982) method requires a search subroutine and ordering of entries.
- Depending on the marginal distributions, the inversion subroutine in NORTA can be cumbersome (e.g. beta distribution).

Copulas

Copulas allow one to generate correlated variates with the desired Spearman correlation coefficients and with desired marginal distributions.

Definition 5.2 *A copula is a multivariate distribution function defined on the unit cube $[0, 1]^n$, with uniformly distributed marginals (see e.g. Embrechts et al. 2003).*

Example 5.6 *A copula defined on the unit square $[0, 1]^2$, with uniform marginal distributions is depicted in Figure 5-10.*

Generating correlated variates with copulas is based on the following 3 concepts (explained with two random variables without loss of generality):

1. A copula is a multivariate distribution with uniform marginal distributions.
2. Given two random variables X and Y with continuous, invertible cumulative distribution functions F_X and F_Y , $\tilde{X} = F_X(X)$ and $\tilde{Y} = F_Y(Y)$ have uniform distributions on $[0, 1]$.

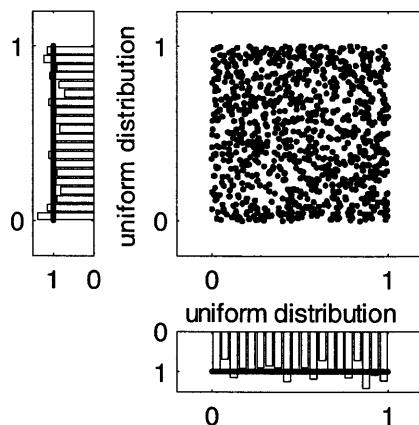


Figure 5-10: Copula on the unit square $[0, 1]^2$ with uniform marginal distributions. Copulas allow one to generate samples from correlated variables with the desired Spearman correlation coefficient and with desired marginal distributions.

3. If \tilde{X} and \tilde{Y} are correlated (Spearman correlation), X and Y are also correlated (the Spearman correlation is invariant under non-linear strictly increasing transformations such as a cumulative distribution function). Note that \tilde{X} and \tilde{Y} have uniform distributions, whereas X and Y can have any distribution.

Thus, if one generates a copula (i.e. a multivariate distribution) with a given Spearman matrix, the uniform marginal distributions have the given Spearman matrix. The uniform marginal distributions can be interpreted as the cumulative distributions of X and Y , $\tilde{X} = F_X(X)$ and $\tilde{Y} = F_Y(Y)$. Thus, the two random variables X and Y , calculated with the inverses of the cumulative distribution functions F_X^{-1} and F_Y^{-1} , have also the given Spearman matrix.

Let X_1, \dots, X_n be random variables with continuous cumulative distributions F_1, \dots, F_n and joint distribution H . There exists a copula C such that (see e.g. Clemen & Reilly 1999)

$$H(x_1, \dots, x_n) = C(F_1(x_1) \dots F_n(x_n)). \quad (5.38)$$

If X_1, \dots, X_n are independent (see e.g. Clemen & Reilly 1999):

$$H(x_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n). \quad (5.39)$$

The choice of copula impacts the dependence structure. If one knows the marginal distributions and the correlation matrix, there exist several possible multivariate distributions (Clemen & Reilly 1999, Embrechts et al. 2003, Kurowicka & Cooke 2006). Figure 5-11 shows two graphs with each 1,000 samples from a bivariate distribution with marginals $X \sim Uniform(0, 1)$ and $Y \sim Uniform(0, 1)$. The Spearman correlation coefficient is 0.7 in both cases. Despite having the same marginals and the same correlation coefficient, the two graphs show different dependence structures. In fact, the two graphs were generated with different copulas: the first graph was generated with the Frank copula, the second graph with the Clayton copula. Depending on the copula, the dependence structure between random variables can be different. Thus, if one knows the marginal distributions and the correlation matrix, the multivariate distribution is not uniquely defined.

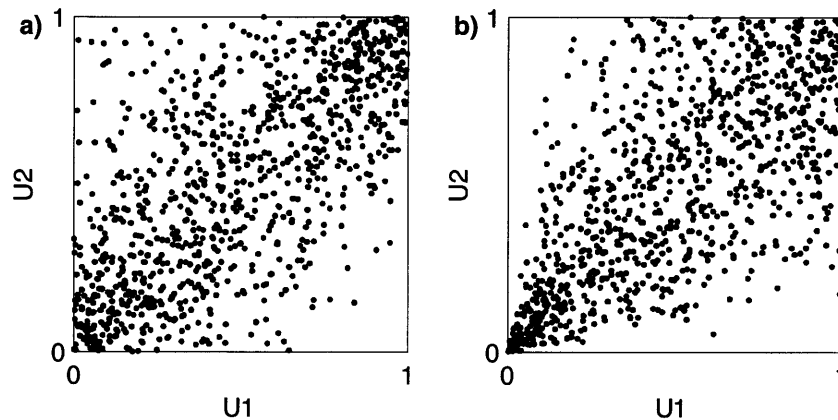


Figure 5-11: a) Frank and b) Clayton copulas with marginal uniform distributions $(0, 1)$ and Spearman correlation $\rho_r = 0.7$. Depending on the copula, the dependence structure between variables can be different. Thus, given the marginal distributions and the correlation, the multivariate distribution is not uniquely defined.

Without delving into the definitions and the properties of the different copulas, in the following qualitative characteristics of some copulas will be outlined. A more detailed description of copulas can be found in e.g. Nelsen (2006).

There are several families of copulas. Perhaps, the most used ones in simulation are the copulas from the elliptical and the Archimedean families. The Clayton and the Frank copulas are both in the Archimedean family. As shown in Figure 5-11, the Frank and the Clayton copulas are

qualitatively different: the Frank copula is characterized by symmetry, whereas the Clayton copula has a fan shape, with a concentration in the lower-left quadrant. These observations are confirmed by Figure 5-12, where three Archimedean copulas (Frank, Clayton and Gumbel) are compared in the case of two correlated uniform distributions $(0, 1)$ at three Spearman correlations. One can see that the Frank copula (top row) is symmetric, whereas the Clayton copula (middle row) and the Gumbel copula (bottom row) are characterized by a fan shape. The fan shape means that the correlation is not constant across the sample. In fact, correlation is larger close to the fan's vertex, where points are highly concentrated, and is lower in the rest of the fan, where points are more spread. Both the Clayton and the Gumbel copula are defined for positive correlations only. For large correlations, the Clayton copula shows a concentration in the lower-left quadrant, whereas the Gumbel copula shows a concentration in the upper-right quadrant. This behavior is named lower tail dependence (Clayton copula), and upper tail dependence (Gumbel copula).

Elliptical copulas include among others the normal copula and the t-copula. In Figure 5-13 the two elliptical copulas are compared in the case of two correlated uniform distributions $(0, 1)$ at three Spearman correlations. The normal copula is in the top row, whereas the t-copula is in the middle row with two degrees of freedom, and in the bottom row with twenty degrees of freedom. The t-copula approaches the normal copula for increasing degrees of freedom similarly to the t-distribution approaching the normal distribution for increasing degrees of freedom. Both the normal and the t-copulas are symmetric, and allow positive as well as negative correlations.

The normal copula, aka Gaussian copula, is often used because it is easy to simulate as it shares many of the tractability properties of the multivariate normal distribution, and because it allows one to model non-normal probability distributions. The copula of the multivariate normal distribution is defined as (see e.g. Embrechts et al. 2003)

$$C_R^{Ga}(u) = \Phi_R^n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)), \quad (5.40)$$

where Φ^{-1} is the inverse of the univariate standard normal pdf, and Φ_R^n is the multivariate standard normal pdf with Pearson correlation matrix, R . The multivariate normal distribution is parametrized in terms of Pearson correlations so that the Pearson correlation matrix, R , is calculated from the Spearman correlation matrix, R_r , with expression 5.14. Parametrizing the multivariate normal distribution with Pearson correlations is necessary; however, the Spearman correlations are the

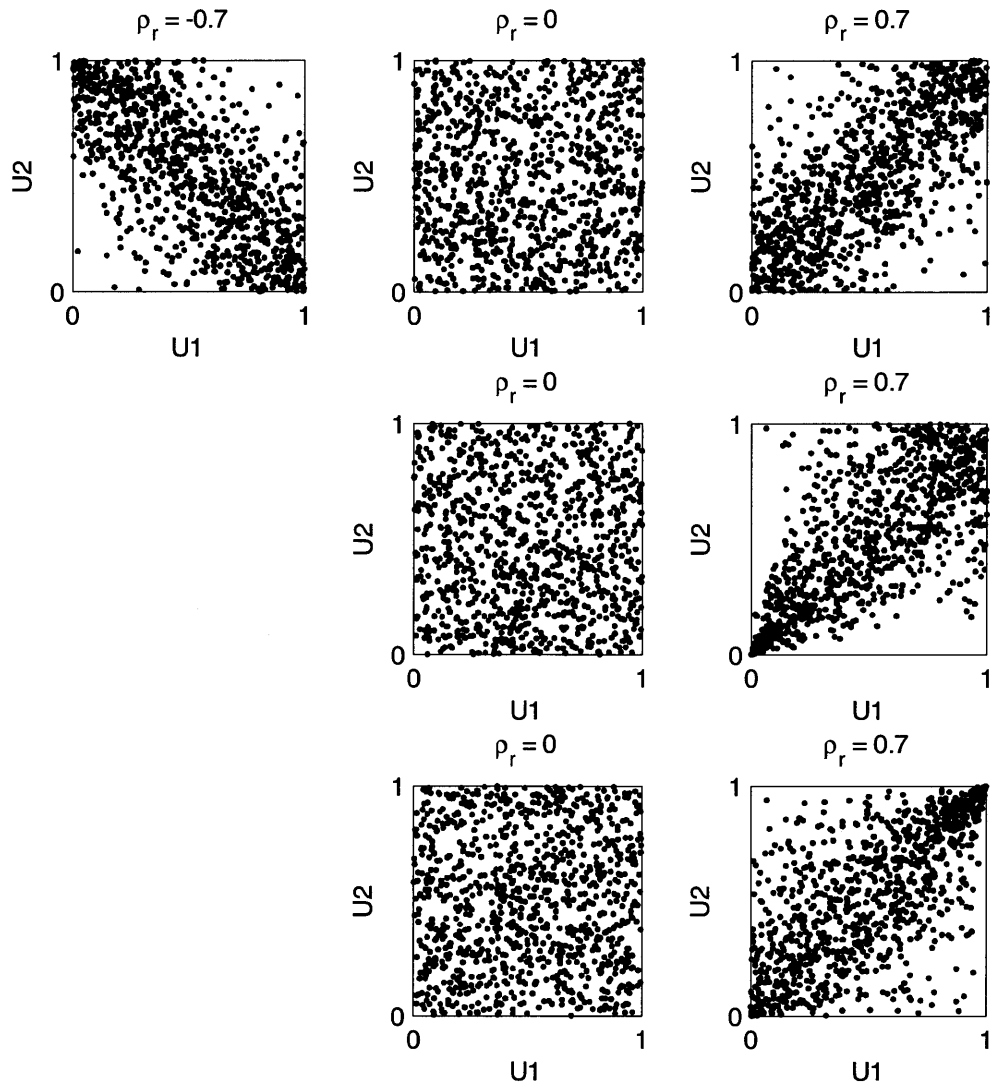


Figure 5-12: Samples from uniform distributions with different Spearman correlations (left column: $\rho_r = -0.7$, center column: $\rho_r = 0$, right column: $\rho_r = 0.7$) generated with different Archimedean copulas (top row: Frank copula, middle row: Clayton copula, bottom row: Gumbel copula). Depending on the copula, the dependence structure between variables changes. The Clayton and the Gumbel copulas exist only for positive correlations.

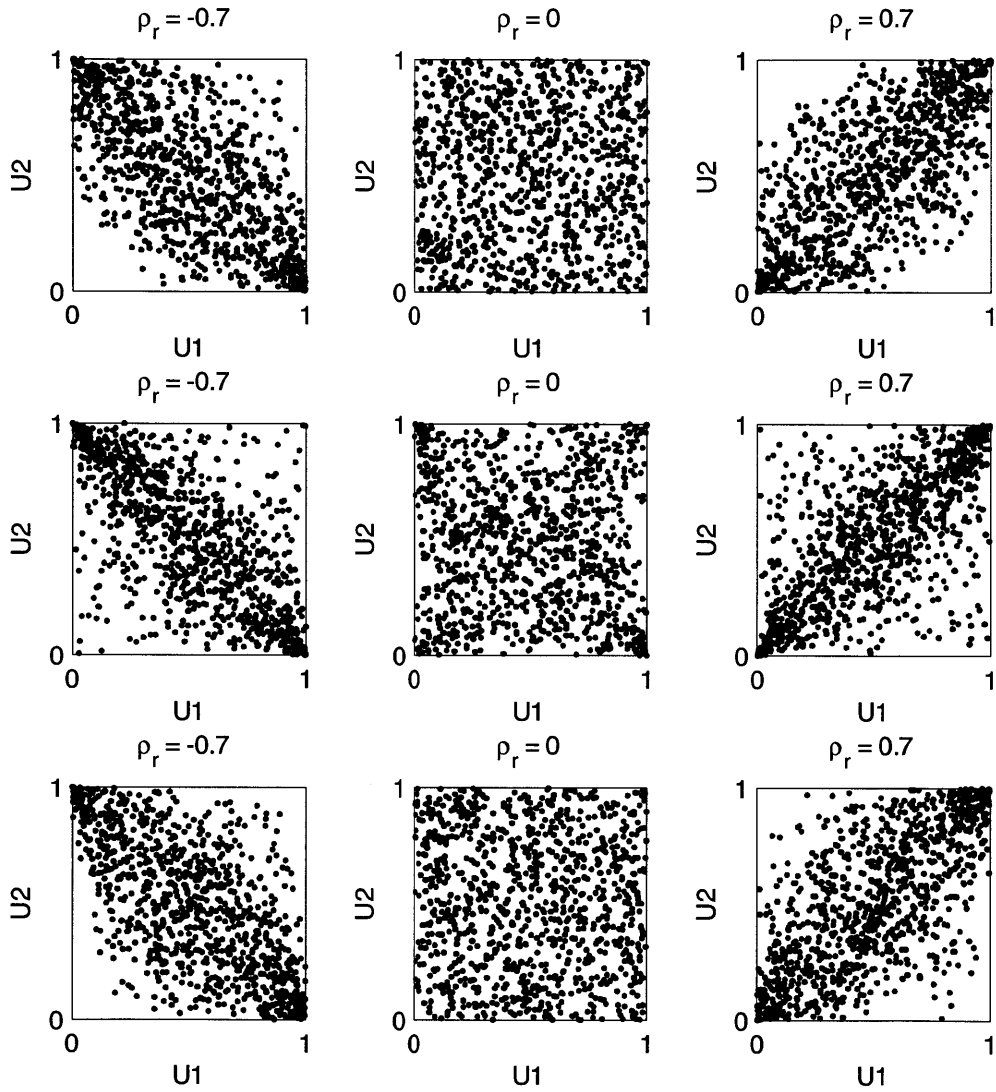


Figure 5-13: Samples from uniform distributions with different Spearman correlations (left column: $\rho_r = -0.7$, center column: $\rho_r = 0$, right column: $\rho_r = 0.7$) generated with different elliptical copulas (top row: normal copula, middle row: t-copula with 2 degrees of freedom, bottom row: t-copula with 20 degrees of freedom). Depending on the copula, the dependence structure between variables changes.

relevant correlations, as it is explained below with the help of the algorithm to generate random variates from the Gaussian copula.

Samples of correlated random variables can be generated from the Gaussian copula. One wishes to generate a sample from k correlated random variables, X_1, \dots, X_k . The Spearman correlation coefficients between the k random variables are given by the $k \times k$ matrix R_r . The Pearson correlation matrix, R , is calculated from the Spearman correlation matrix, R_r , with expression 5.14.

The algorithm to generate random variates from the Gaussian copula is implemented in 5 steps (see e.g. Embrechts et al. 2003):

1. Calculate the lower triangular matrix P such that $R = PP^T$, using e.g. the Cholesky factorization.
2. Draw a sample, z_1, \dots, z_k , from $\mathcal{N}(0, 1)$.
3. Set $y = Pz$.
4. Set $u_i = \Phi(y_i)$, $i = 1, \dots, k$. Note $(u_1, \dots, u_k) \sim C_R^{Ga}$.
5. $(x_1, \dots, x_k) = (F_1^{-1}(u_1), \dots, F_k^{-1}(u_k))$.

To generate n samples, steps 2-5 are repeated n times.

As mentioned above, the relevant correlation matrix is the Spearman correlation matrix, R_r , whereas the Pearson correlation matrix, R , should be considered as the means to the goal. In fact, the Pearson correlation matrix, R , allows one to generate normal random variates with Pearson correlation matrix, R , which also have Spearman correlation matrix, R_r (step 3 in the algorithm above). In steps 4 and 5, two non-linear strictly increasing transformations are applied to the generated sample, so that the output sample will have Spearman correlation matrix, R_r , and will generally not have Pearson correlation matrix, R , because the latter is not invariant under non-linear strictly increasing transformations.

One can observe that the algorithm above coincides with NORTA (previous correlation model). In fact, the first three steps of the algorithm allow one to generate normal random variates, which is the outcome of the first step in the NORTA method. The last two steps of the algorithm are

identical to the last two steps of NORTA. Thus, the NORTA method is based on the Gaussian copula.

It is not clear at this point in time what the criteria are to select a copula. Currently there is not agreement on a general procedure to select a copula (Genest et al. 2006, Schoelzel & Friederichs 2008). Copula selections is an ongoing topic of research. Several methods, including copula-specific goodness-of-fit tests, have been suggested (Genest et al. 2006, Genest & Favre 2007).

Selected Method for Modeling Correlations

In the preceding, eight correlation models have been presented. Here, they are compared against the requirements they must fulfill. The requirements on the correlation model, formulated at the beginning of section 5.3, are summarized here:

1. The desired correlation should be achieved.
2. The model should be distribution free, i.e. should allow one to use any marginal distribution.
3. The number of variables, which are correlated at one time, should not be limited.

The eight correlation models have been checked against the three requirements (Table 5.3). The first five methods are not further pursued because of the their disadvantages, which are reiterated here:

- The limitation of disaggregation does not model correlations but rather identify scenarios where correlations are not relevant and the pessimistic/optimistic approach allows one to model only extreme scenarios of correlations so that both approaches are not viable for large number of variables and complex systems such as a rail line construction.
- Sampling from a joint distribution satisfies the three requirements. However, the sampling method is either inefficient (acceptance-rejection algorithm) or not converging (Metropolis-Hastings).
- The CLT requires independence between variables so that cases where the variables are correlated cannot be modeled.

- The lognormal marginals method cannot model correlation between not lognormally distributed variables.

Iman and Conover's method and NORTA have been considered in more detail. In examples 5.4 and 5.5, correlated samples from a bivariate distribution with beta and lognormal marginals have been generated. The examples showed that both methods are relatively easy to implement. NORTA has one disadvantage compared to Iman and Conover's method: the marginal distributions must be continuous and invertible. Compared to NORTA, Iman and Conover's method needs one step more in the random generation algorithm, and requires a search subroutine to identify ranks. However, the main disadvantage, and in this sense the limitation, of Iman and Conover's method is that the generated samples have only approximately the desired correlation matrix. Due to this limitation Iman and Conover's method is not further pursued.

NORTA is based on a copula, the Gaussian copula. The question how to identify the most appropriate copula has also been raised. It is not clear at this point in time what the criteria are to select a copula. Until criteria on the procedure to select a copula are widely accepted, from the practical point of view selecting the Gaussian copula is no less appropriate than another choice of copula. Thus, the NORTA method, which is based on the Gaussian copula, is selected to generate samples from random variables with the desired correlation and with desired marginal distributions.

Table 5.3: The correlation models are compared against the requirements. Although sampling from a joint distribution satisfies all the requirements, it is not selected since it is either inefficient or does not converge. The NORTA method, which is based on the Gaussian copula is the selected correlation model.

Methods	requirements		
	desired correlation	distribution free	unlimited number of variables
limitation of disaggregation	no	no	no
pessimistic/optimistic approach	no	no	no
sampling from joint distribution	yes	yes	yes
CLT	no	yes	yes
lognormal marginal method	yes	no	yes
Iman and Conover's method	not exact	yes	yes
NORTA	yes	continuous & invertible	yes
copula (Gaussian)	yes	continuous & invertible	yes

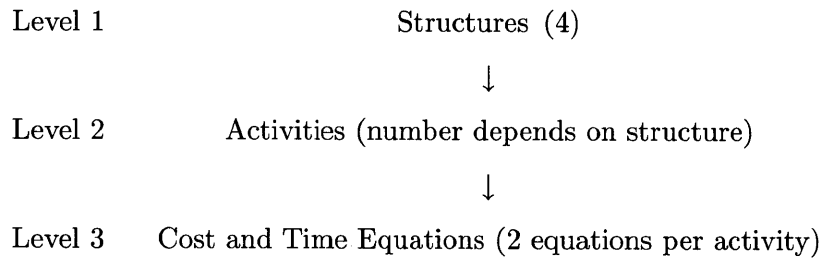
In the preceding, eight correlation methods have been presented. The correlation method NORTA complies with the requirements by the uncertainty model and is the best suited method

to represent correlations. With a correlation measure (Spearman correlation) and a correlation method (NORTA), correlations between cost variables in the construction of rail lines can be now modeled as shown below.

5.4 Correlations in the Construction of a Rail Line

After explaining how the total construction cost and time of a rail line are calculated, this section discusses pertinent correlations between costs in rail line construction. These are analyzed with two case studies and their impacts on the total cost distribution are assessed.

The construction of a rail line can be modeled as the construction of a sequence of four main structures (chapter 3): tunnels, viaducts, cuts and embankments. Every structure is modeled as a sequence of activities, which are characterized by cost and time equations. A hierarchy can be identified:



The total construction cost of a rail line is given by the sum of the costs of tunnel activities, $t = 1, \dots, n_t$, the costs of viaduct activities, $v = 1, \dots, n_v$, the costs of cut activities, $c = 1, \dots, n_c$, the costs of embankment activities, $e = 1, \dots, n_e$:

$$TotalCost = \sum_{t=1}^{n_t} Cost_t + \sum_{v=1}^{n_v} Cost_v + \sum_{c=1}^{n_c} Cost_c + \sum_{e=1}^{n_e} Cost_e. \quad (5.41)$$

Differently from the total cost, the total time is not the sum of all activities time because some activities are performed in parallel, so that the total construction time of a rail line is given by the sum of the activities, $j = 1, \dots, m$, on the longest path:

$$TotalTime = \sum_{j=1}^m Time_j. \quad (5.42)$$

It will now be shown how the correlations are considered in the rail line application.

5.4.1 Identification of Correlations between Costs

Correlations are identified given the following assumptions:

- The crew (equipment and labor) is assumed to be performing similarly at all times. If needed, an initial learning effect is modeled.
- The material, such as concrete, steel, blasting material, etc. is assumed to be available and the cost fluctuations are included in the cost distributions of the activities.
- Minor weather-related events, such as rain, are assumed to be included in the cost and time distributions of the activities. Exceptional weather events are modelled as disruptive events (section 4.2).

Given the above assumptions, 4 types of correlation are identified in the rail line construction model, which is described in chapter 3:

1. Correlation between the costs of a repeated activity in a structure,
2. Correlation between the costs of different activities in a structure,
3. Correlation between the costs of activities in adjacent structures,
4. Correlation between the costs of same activities in same type of structures.

In the following, the identified correlations for the four main structures are described in detail and different correlation scenarios are discussed.

Correlation between the Costs of a Repeated Activity in a Structure

The costs of a repeated activity are expected to be positively correlated because of the repetitiveness of the processes. Activities may change from structure to structure so that this type of correlation is analyzed one structure at a time.

- Tunnel: The construction of a tunnel is modeled with one activity. The activity may change along the tunnel if geology changes. Here, the case of the same activity being repeated is considered. The case of different activities in a tunnel is considered on page 176.

The tunnel activity is repeated for every meter of tunnel length. The scenarios are:

1. Every meter is independent, $\rho_r = 0$, so that the cost per meter is randomly selected for each meter.
 2. Every meter is perfectly correlated, $\rho_r = 1$, i.e. once a cost per meter is randomly selected, this remains constant for the entire tunnel.
 3. Intermediate scenarios include all positive correlations, $\rho_r = (0, 1)$, which correspond to the case where the cost per meter is randomly selected for each meter, and if one cost per meter is above average, the next cost per meter will tend to be also above average.
- Viaduct: The activities deck section, pier, foundation, abutment, and technical block are repeated in the construction of a viaduct. Although the crew is assumed to perform similarly at all times and despite the repetitiveness of the process, it is assumed that the costs of repeated activities in a viaduct are independent for the following reason. The deck section length can vary from deck section to deck section, the pier height can vary from pier to pier, the foundation depth can vary from foundation to foundation, the two abutments and the two technical blocks are on opposite ends of the viaduct. Thus, it is expected that e.g. the cost of pier i is uncorrelated from the cost of pier $(i + 1)$.
 - Cut: The activities clearing, excavating, capping, and sub-ballast are repeated in the construction of a cut. The activity excavating may change along the cut if geology changes. Here, the case of the same activity being repeated is considered. The case of different excavating activities in a cut is considered on page 177.

The activities clearing, excavating, capping, and sub-ballast are repeated for every meter of cut length so that for all activities the scenarios are:

1. Every meter is independent, $\rho_r = 0$, so that the cost per meter is randomly selected for each meter.
2. Every meter is perfectly correlated, $\rho_r = 1$, i.e. once a cost per meter is randomly selected, this remains constant for the entire cut.
3. Intermediate scenarios include all positive correlations, $\rho_r = (0, 1)$, which correspond to the case where the cost per meter is randomly selected for each meter, and if one cost per meter is above average, the next cost per meter will tend to be also above average.

- Embankment: The activities clearing, improving, filling, capping, and sub-ballast are repeated in the construction of an embankment. The activity improving may change along the embankment if geology changes. Here, the case of the same activity being repeated is considered. The case of different improving activities in an embankment is considered on page 177.

The same reasoning on costs of a repeated activity applies similarly to the cut.

Correlation between the Costs of Different Activities in a Structure

The costs of different activities in a structure are expected to be positively correlated because these activities might be subject to the same type of constraints. Activities change from structure to structure so that this correlation is analyzed one structure at a time.

- Tunnel: The construction of a tunnel is modeled with one activity. The tunnel activity may change if geology changes. Tunnel activity i is repeated n_i times, tunnel activity j is repeated n_j times. As the geology changes, tunnel activity i and tunnel activity j are considered independent.
- Viaduct: The construction of a viaduct is modeled with the activities deck section, pier, foundation, abutment, technical block, and finishing. It is assumed that the cost of the activities abutment, technical blocks, and finishing are not subject to site access constraints. Deck section, pier, foundation form a unit: first the foundation is constructed, followed by the pier and eventually the deck section. The construction of the foundation and the pier might be restricted by site access. If the site access for equipment and labor is difficult when constructing the foundation, similarly it will be difficult when constructing the pier. Depending on the construction method (balanced cantilever, span-by-span, launching; see section 3.2.2), the construction of the deck section might also be restricted by the site access. Thus, the cost of the construction of foundation, pier and deck section can be correlated due to site access. The scenarios are:
 1. The activities are independent, $\rho_r = 0$, so that the cost per foundation, pier, and deck section, respectively, are randomly selected for each activity.

2. The activities are correlated, $\rho_r = (0, 1]$, so that the cost per activity is randomly selected for each activity, and if the cost of an activity, e.g. foundation, is above average, the cost of the other activity, e.g. pier, will tend to be also above average.
- Cut: The construction of a cut is modeled with the activities clearing, excavating, capping, and sub-ballast. These activities might also be constrained by the site access so that similar scenarios as for the viaduct are considered.
 1. The activities are independent, $\rho_r = 0$, so that the cost of clearing, excavating, capping, and sub-ballast, respectively, are randomly selected for each activity.
 2. The activities are correlated, $\rho_r = (0, 1]$, so that the cost per activity is randomly selected for each activity, and if the cost of an activity, e.g. clearing, is above average, the cost of the other activity, e.g. excavating, will tend to be also above average.

The excavating activity may change if geology changes. Excavating activity i is repeated n_i times, excavating activity j is repeated n_j times. As the geology changes, excavating activity i and excavating activity j are considered independent.

- Embankment: The construction of an embankment is modeled with the activities clearing, improving, filling, capping, and sub-ballast. The same reasoning as for the cut applies.

The improving activity may change if geology changes. The same reasoning as for the excavating activity in a cut applies to the improving activity in an embankment.

Correlation between the Costs of Activities in Adjacent Structures

Typical adjacent structures are cuts and embankments in a cut and embankment sequence, cuts and tunnels at the tunnel's portals, and embankments and viaducts at the viaduct's ends. It is possible but unusual to have adjacent cuts and viaducts, adjacent embankments and tunnels, and adjacent viaducts and tunnels. Activities change from structure to structure so that this correlation is analyzed for a pair of typical adjacent structures at a time.

- Cut-embankment: The construction of a cut and the construction of an embankment share three activities: clearing, capping, and sub-ballast. If e.g. the clearing of the cuts is done

separately of the clearing of the embankments so that clearing is interrupted between the cut and the embankment, the cost of clearing the cut can be considered independent of the cost of clearing the embankment. The same applies to capping and sub-ballast. If these activities are repeated without interruption from the beginning to the end of the cut and embankment sequence, the costs can be modelled as correlated costs of repeated activities as it is explained under "Correlation between the Costs of a Repeated Activity in a Structure", where the structure is no longer the cut or the embankment but rather the cut and embankment sequence. The activities clearing, capping, and sub-ballast are repeated for every meter of cut and embankment sequence so that for all activities the scenarios are:

1. Every meter is independent, $\rho_r = 0$, so that the cost per meter is randomly selected for each meter.
 2. Every meter is perfectly correlated, $\rho_r = 1$, i.e. once a cost per meter is randomly selected, this remains constant for the entire cut and embankment sequence.
 3. Intermediate scenarios include all positive correlations, $\rho_r = (0, 1)$, which correspond to the case where the cost per meter is randomly selected for each meter, and if one cost per meter is above average, the next cost per meter will tend to be also above average.
- Cut-tunnel: The cut at the tunnel's portal and the tunnel are excavated in the same geology so that positive correlations between the cost of the cut activity excavating and the cost of the tunnel activity are expected. The cost distribution of a cut and the cost distribution of a tunnel are functions of the geology, in the sense that different cost distributions apply for different geologies. Thus, similar geologies determine similar cost distributions. Although this type of correlation is not formally modelled with correlation coefficients, it is addressed with similar cost distributions for similar geology.
 - Embankment-viaduct: Although an embankment is adjacent to a viaduct, in the model (chapter 3) embankments and viaducts do not share activities. In fact, embankment activities (clearing, improving, filling, capping, sub-ballast) are not viaduct activities, and vice versa. Also the geology can be considered different because the embankment geology depends mostly on the filled in material.

Correlation between the Costs of Same Activities in Same Type of Structures

Positive correlations between the costs of the same type of structures (e.g. between tunnel i and tunnel j) are expected if the geology is similar. As explained above, the cost distribution of a structure is a function of the geology so that similar geologies determine similar cost distributions. It follows that the costs of the same activities in the same type of structures are assumed independent since the simulation draws new numbers from the input cost distributions when the construction of a new structure is started. Thus, the underlying assumption is that structures are in different locations and/or using different construction means and methods. Note that this assumption of independence may not always apply.

Correlation between the Cost and the Time of an Activity

In this chapter correlations between costs only have been addressed. This focus is motivated by the observation that the sum of positively correlated costs has a larger standard deviation than the sum of independent costs (Newton 1992, Touran 1993). However, correlations between times and correlations between cost and time might occur. Correlations between time variables of repetitive activities in building construction have been modeled by Yang (2006) using NORTA, showing that correlation between time variables also causes an increase in total time standard deviation. Correlations between time variables can be modeled with the same methodology as for correlations between cost variables:

1. Identify possible correlations between time variables in the construction of rail lines.
2. Model correlations with NORTA.
3. Assess the impact of the correlations on the variance of the total time.

Correlations between cost and time of an activity might also occur. It is expected that the cost and the time of an activity are positively correlated. In fact, if an activity lasts longer than the average time, one can expect the cost of the activity to be also larger than average. Correlations between cost and time are also modeled here. Three scenarios are identified:

1. Cost and time are independent, $\rho_r = 0$, so that cost and time are selected independently.

2. Cost and time are perfectly correlated, $\rho_r = 1$, so that if the cost corresponds to the 80th percentile of the cost distribution, the time also corresponds to the 80th percentile of the time distribution.
3. Intermediate scenarios include all positive correlations, $\rho_r = (0, 1)$, which correspond to the case where the cost or the time is randomly selected and if e.g. the cost is above average, the time will tend to be also above average.

For the case of transportation infrastructure projects, it should be noted that activity cost and activity time are often not correlated.

Summary

In this section, five types of correlations in the construction of a rail line have been identified. The five types of correlations are summarized here:

1. Correlation between the costs of a repeated activity in a structure,
2. Correlation between the costs of different activities in a structure,
3. Correlation between the costs of activities in adjacent structures,
4. Correlation between the costs of same activities in same type of structures,
5. Correlation between the cost and the time of an activity.

The five types of correlations have been considered for the four types of structures. The types of correlations that are modeled with correlation coefficients are indicated with an 'x' in Table 5.4.

Table 5.4: Summary of types of correlation and structures considered in this chapter.

Structures	correlation type				
	1	2	3	4	5
tunnel	<i>x</i>				<i>x</i>
viaduct		<i>x</i>			<i>x</i>
cut	<i>x</i>	<i>x</i>	<i>x</i>		<i>x</i>
embankment	<i>x</i>	<i>x</i>	<i>x</i>		<i>x</i>

The impact of the correlation types are assessed with case studies. Each correlation type is investigated by modeling the construction of one of the structures. This choice assumes that the

impact of the correlation type on the total cost distribution is similar in the different structures. Correlation type 1 is investigated in a tunnel case study, correlation type 2 and type 5 are investigated in a viaduct case study. Correlation type 3 occurs in cuts and embankments, where activities are repeated. As explained above this can be modeled as a correlation type 1 (correlation between the costs of a repeated activity in a structure), where the structure is a cut and embankment sequence. For this reason no case study on correlation type 3 is presented here. Correlation type 4 is not further investigated.

In this section, five types of correlations have been identified and discussed in detail. Correlation types 1, 2, and 5 are further investigated with case studies. In the next section, correlation types 2 and 5 are analyzed with a viaduct case study. Correlation type 1 will be investigated in section 5.4.3.

5.4.2 Assessment of the Correlations' Impact: Viaduct Case Study (Correlation Type 2 and Type 5)

In this section, correlation type 2 (correlation between the costs of different activities) and correlation type 5 (correlation between the cost and the time of an activity) are investigated with a viaduct case study.

The case study is a viaduct of the new High Speed Rail in Portugal. The viaduct is 395m long and consists of the following elements (Figure 5-14a): 2 abutments, 2 technical blocks, 10 deck sections, 9 piers, 9 foundation footings, 9 foundation pile sets, the finishing and 1 set of sloped piers (the sloped pier cost includes the additional cost of having 2 sloped piers instead of 1 vertical pier). The construction of the viaduct's elements is modeled with activities including cost and time equations. The activities are named with the element whose construction they model, e.g. the activity abutment is the activity modeling the construction of an abutment.

The construction of the viaduct requires some additional activities: pre-work, activity to model the construction site setup time with zero cost; pre-deck, activity to model the deck construction setup time with zero cost; construction method, activity to model the cost of the selected deck construction method (cast-in-place concrete with gantry and staging) with zero time. All activities are included in the viaduct's activity network (Figure 5-14b).

The activities are summarized in Table 5.5. The time of the activities pier, foundation footing,

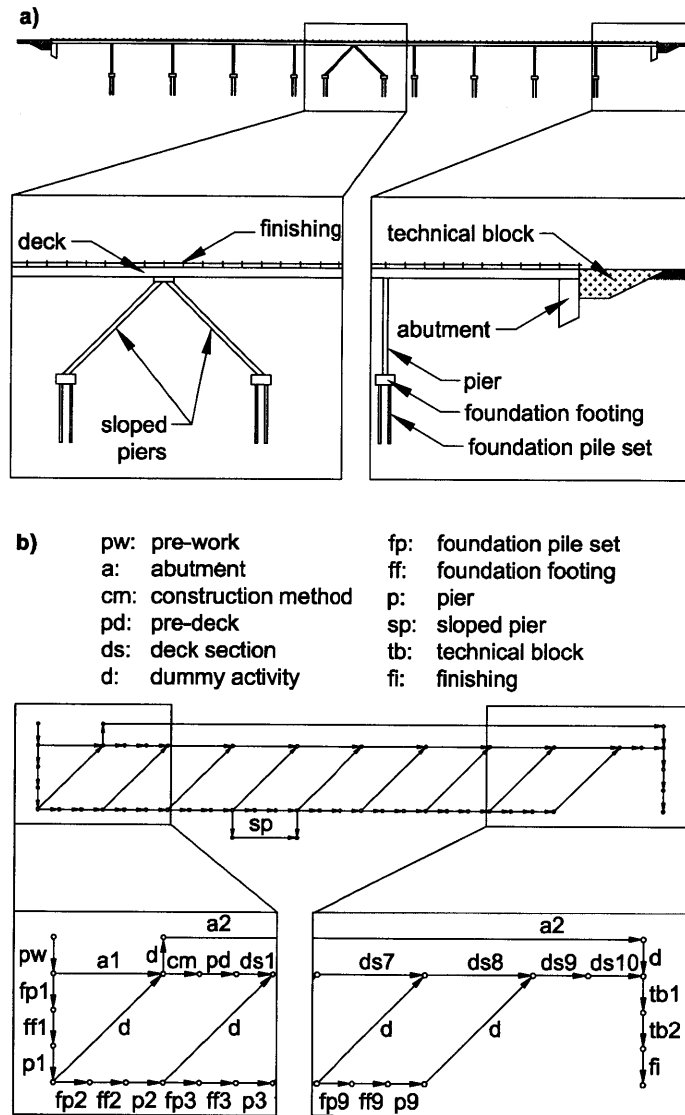


Figure 5-14: Viaduct case study. a) the viaduct consists of several elements, whose construction is modeled with b) the activity network. The total cost is the sum of all activity costs, while the total time is the sum of the activity times on the critical path in the activity network. A dummy activity, which is used to model dependencies between activities, has no cost and no time.

and foundation pile set are zero because they are in parallel with the activity deck section. The latter has been identified by the bridge engineer as being the more time consuming so that constructing a foundation pile set, a foundation footing and a pier takes less time than constructing a deck section, which requires 14 days. Note that this might not be the case in other projects.

Table 5.5: Viaduct case study. Estimates of the cost and time for the viaduct activities (after RAVE (2006b)). The estimates are the input for the cost and time of activities in the activity network.

Activity	abbreviation	repetitions	cost [euro]	time [days]
abutment	a	2	250,000	56
technical block	tb	2	150,000	14
deck section	d	10	163,925	14
pier	p	9	50,784	0
sloped pier	sp	1	240,992	0
foundation footing	ff	9	50,400	0
foundation pile set	fp	9	57,024	0
finishing	fi	1	177,750	6
pre-work	pw	1	0	42
pre-deck	pd	1	0	56
construction method	cm	1	370,000	0

In order to evaluate the impact of correlations on the distributions of total cost and total time, a probabilistic analysis is needed. The cost and time data available are deterministic. Initially, no information on the probabilistic distributions of the cost and time variables were available. Due to the lack of probabilistic data, assumptions were to construct cost and time probability distributions of the activities:

- The underlying distribution of cost variables is lognormal.
- The mode of the cost distribution is assumed equal to the deterministic estimate.
- The minimum value of the cost distributions is 80% of the mode.
- There is a probability of 2% of exceeding the High Value, which is assumed to be 150% of the mode, of the cost distributions. As an example, the probability distribution of the abutment's cost variable is shown in Figure 5-15a.
- The underlying distribution of time variables is approximated with the triangular distribution.

- The mode of the time distribution is assumed equal to the deterministic estimate.
- The minimum and the maximum of the time distributions are respectively 80% and 130% of the mode. As an example, the probability distribution of the abutment's time variable is shown in Figure 5-15b.

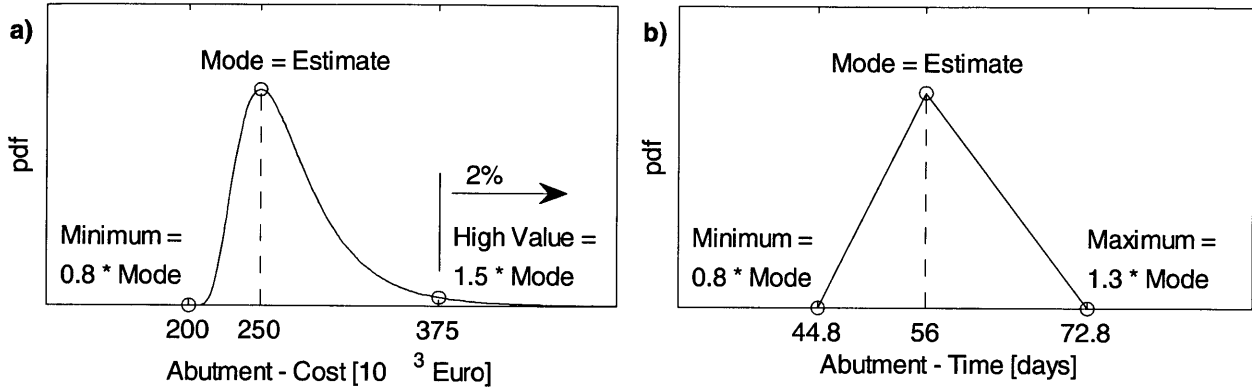


Figure 5-15: a) example of a lognormal cost distribution (abutment's cost), whose minimum, mode, and high value (exceeded with 2% probability) are determined based on the estimated cost (Table 5.5), b) example of a time triangular distribution (abutment's time), whose minimum, mode, and maximum value are determined based on the estimated time (Table 5.5).

The cost distributions C_i are defined as

$$C_i \sim \text{Lognormal}(\mu, \sigma), \quad (5.43)$$

where μ and σ are the mean and the standard deviation of the corresponding normal distribution.

Parameters μ and σ are calculated by solving the following system of equations

$$\begin{cases} \text{Mode} - \text{Min} = \exp(\mu - \sigma^2) \\ 1 - 0.02 = 0.5 + 0.5 \operatorname{erf}\left(\frac{\ln(\text{HighValue} - \text{Min}) - \mu}{\sigma\sqrt{2}}\right) \end{cases} \quad (5.44)$$

where Min , Mode , HighValue are the minimum, the mode and the value that can be exceeded with 2% probability, respectively, of the lognormal distribution.

The time distributions T_i are defined as

$$T_i \sim \text{Triangular}(\text{Min}, \text{Mode}, \text{Max}), \quad (5.45)$$

where Min , Mode , Max are the minimum, the mode and the maximum, respectively, of the triangular distribution.

In the following, the total cost and the total time for the construction of the viaduct are calculated, first deterministically and then probabilistically. In the probabilistic calculations, correlation type 2 (correlation between the costs of different activities) and correlation type 5 (correlation between the cost and the time of an activity) are modeled with particular scenarios and analyzed in detail with sensitivity analyses.

Deterministic Calculation

Table 5.5 lists the cost and the time required for the construction of the viaduct elements. The total cost is given by the sum of the element costs according to expression 5.41, and is equal to €4,651,864. The total time is given by the sum of the element times along the longest path according to expression 5.42. As shown in Table 5.5, the construction time of the foundation pile sets, foundation footings and piers is assumed to be equal to zero because the bridge engineer has identified the longest time path in the sequence of activities to be the construction of the deck sections (viaduct network in Figure 5-14). The longest path yields a total construction time of 328 days.

Probabilistic Calculation

The range of total cost and total for the construction of the viaduct is estimated for three different correlation scenarios. Depending on the impact, they are further investigated with sensitivity analyses.

The probabilistic calculation is based on Monte Carlo simulations with 5,000-sample runs. The number of 5,000 samples has been selected because it bounds the total cost standard deviation within $\pm 1\%$ in 10 out of 10 simulations.

The types of correlation described in section 5.4.1 for a viaduct are translated into scenarios for the construction of the 395m long viaduct:

- Scenario 1 is the reference scenario. Cost and time variables of each viaduct element are independent.
- Scenario 2 represents a correlation between the costs of different activities in the viaduct (correlation type 2 in Table 5.4). It is assumed that the construction of the pile set, the footing and the pier is constrained by the site access so that the pile set cost, the footing cost and the pier cost are correlated. A pile set, a footing and a pier form a unit (Figure 5-16). As the viaduct is constructed with the span-by-span construction method (section 3.2.2), it is assumed that the costs of deck sections are independent of the costs of the pile sets, the footing and the piers. The 395m viaduct includes nine units (Figure 5-14). The correlation is assumed to be $\rho_r = 0.8$ so that the correlation matrix of the elements in unit i is

$$R_i = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}. \quad (5.46)$$

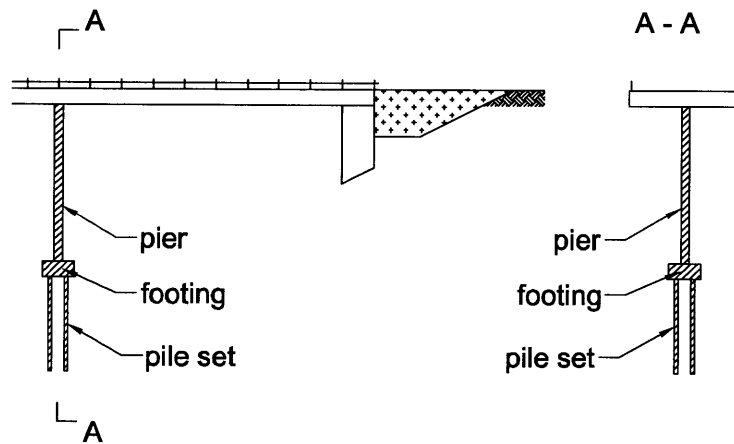


Figure 5-16: A unit (dashed) of the viaduct consists of a pile set, a footing and a pier. In scenario 2, the costs of these are correlated.

- Scenario 3 represents the correlation between the cost and the time of an activity (correlation type 5 in Table 5.4). Costs are independent of one another, times are independent of one

another but the cost and the time of each element i are correlated with correlation matrix

$$R_i = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}. \quad (5.47)$$

The mean of the total cost and the mean of the total time are equal to the sum of the cost means and the time means along the critical path, respectively. The cost means and the time means do not change from scenario to scenario. Thus, the mean of the total cost and the mean of the total time are not expected to vary in the three scenarios. The standard deviation of the total time is also not expected to vary because the time variables are assumed independent in the three scenarios. The standard deviation of the total cost is expected to increase in scenario 2 due to the correlation between the costs of the pile sets, the footings and the piers. The correlation between total cost and total time is expected to deviate from zero only in scenario 3, where the cost and the time of each activity are correlated. These expectations are confirmed by the results detailed below.

The means, the standard deviations of total cost and total time, and correlations between total cost and total time in the three scenarios are summarized in Table 5.6. Also in Table 5.6 are reported the normalized differences, Δ , in mean and standard deviation. The normalized difference in mean is calculated as

$$\Delta = \frac{m - m_r}{m_r}, \quad (5.48)$$

where m is the mean of the scenario and m_r is the mean of the reference scenario (scenario 1).

The normalized difference in standard deviation is calculated as

$$\Delta = \frac{s - s_r}{s_r}, \quad (5.49)$$

where s is the standard deviation of the scenario and s_r is the standard deviation of the reference scenario (scenario 1).

The scatterplot, the frequency plots of cost and time in scenario 1, scenario 2, scenario 3 are shown in Figures 5-17a, 5-17b, and 5-17c, respectively.

One needs to note that the results vary slightly from simulation to simulation due to the random generation. In fact, the number of 5,000 samples per simulation bounds the total cost standard

Table 5.6: Viaduct case study, scenarios 1, 2 and 3. Means, standard deviations of total cost and total time, and correlations between total cost and total time in the scenarios 1, 2 and 3. The standard deviation increases by 8.3% in scenario 2, while the total cost - total time correlation is equal to 0.38.

Scenario	cost			time		total cost – total time	
	mean [euro]	Δ [%]	st. dev. [euro]	Δ [%]	mean [days]	st. dev [days]	correlation [-]
1	5,059,400	–	131,200	–	339	10.4	0.00
2	5,060,000	0.0	142,100	8.3	339	10.4	0.00
3	5,059,600	0.0	131,300	0.1	339	10.5	0.38

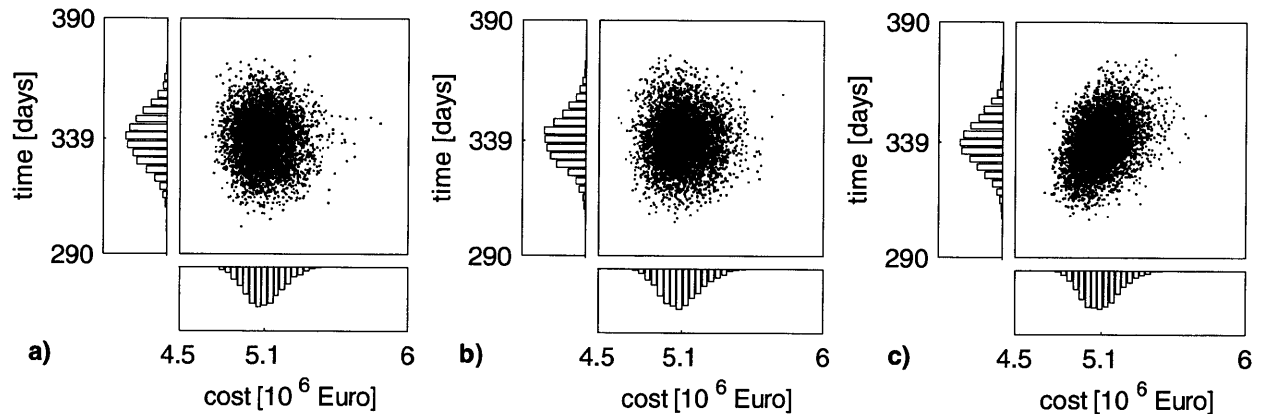


Figure 5-17: Viaduct case study, scenarios 1, 2 and 3. Scatterplots and frequency plots of the total cost and total time for a) scenario 1 (independent), b) scenario 2 (correlation type 1), and c) scenario 3 (correlation type 5). While correlation type 1 increases the spread of the total cost, correlation type 5 causes the correlation between total cost and total time.

deviation within $\pm 1\%$ in 10 out of 10 simulations. The following observations can be made:

- The mean of the total cost can be considered constant from scenario to scenario (Table 5.6). This is confirmed by a constant center of gravity in the cost frequency plots (Figure 5-17).
- The standard deviation of the total cost increases by 8.3% in scenario 2 (Table 5.6), where the costs of unit elements (pile set, footing, pier) were assumed to be correlated with correlation 0.8. In Figure 5-17b the increase in total cost standard deviation is barely visible in the scatterplot and in the frequency plot. It is likely that the total cost standard deviation depends on the viaduct length. This aspect is further investigated in "Scenario 2 - Sensitivity Analysis".
- The mean and the standard deviation of the total time remain constant (Table 5.6). This is confirmed by constant center of gravity and spread of the time frequency plots (Figure 5-17).
- Total cost and total time in scenarios 1 and 2 are uncorrelated (Table 5.6). The correlation between total cost and total time is 0.38 in scenario 3, where the correlation between the cost and the time of each element was assumed to be 0.8. The correlation between total cost and total time in scenario 3 is visible in the inclination of the cloud of data points in Figure 5-17c. It is likely that the correlation between total cost and total time depends on the viaduct length. This aspect is further investigated in "Scenario 3 - Sensitivity Analysis".
- The means of total cost and total time are larger than the deterministic total cost and total time ("Deterministic Calculation"). The means are larger than the deterministic results for the following reasons:
 1. Cost input distributions and time input distributions are skewed to the right, i.e. the mean is larger than the mode (Figure 5-15).
 2. The deterministic total cost and total time are the sum of the modes of the input distributions (Figure 5-15).
 3. The probabilistic means of total cost and total time are equal to the sum of the means of the input distributions.

Three correlation scenarios in the construction of a viaduct have been analyzed. It has been found that scenario 2 causes the standard deviation to increase while scenario 3 causes the correlation between total cost and total time to increase. Scenario 2 and scenario 3 are further investigated with sensitivity analyses.

Scenario 2 - Sensitivity Analysis

Scenario 2 (correlation between the costs of different activities - correlation type 2) has confirmed that when some input costs are correlated, the standard deviation of the total costs is larger than when all input costs are independent. The goal of this sensitivity analysis is to further investigate the change in total cost standard deviation as a function of viaduct length.

The viaduct length increases with the number of units in the viaduct. A unit consists of a pile set, a footing and a pier (Figure 5-14). In scenario 2, it is assumed that the pile set cost, the footing cost and the pier cost are correlated, whereas the rest of the costs are independent. The correlation between pile set cost, footing cost and pier cost is assumed to be $\rho_r = 0.8$ so that the correlation matrix of the elements in unit i is (as in expression 5.46):

$$R_i = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}. \quad (5.50)$$

5,000-sample simulations were run for viaducts with 2, 9, 30, 50 and 100 units, respectively. Again, the number of 5,000 samples has been selected because it bounds the total cost standard deviation within $\pm 1\%$ in 10 out of 10 simulations. The means and the standard deviations of total cost and total time for different number of units, u , are summarized in Table 5.7. The means and the standard deviations of the total cost assuming independence are also entered in Table 5.7 for reference. The change in total cost standard deviation is quantified in terms of normalized difference between the total cost standard deviation for correlated pile set cost, footing cost and pier cost, and the total cost standard deviation for independent pile set cost, footing cost and pier cost. The normalized difference in total cost standard deviations is calculated as

$$\Delta = \frac{s - s_r}{s_r}, \quad (5.51)$$

where s is the total cost standard deviation for correlated pile set cost, footing cost and pier cost, and s_r is the standard deviation for independent pile set cost, footing cost and pier cost. The normalized differences in total cost standard deviation, Δ , are also entered in Table 5.7.

The scatterplots, the frequency plots of cost and time are shown in Figure 5-18. The total cost and the total time change with the number of units. Thus, the scale on the axes in the five graphs are different. Nevertheless, comparability is obtained by plotting results with the same scale ratio, i.e. with ratio $x - axis$ to $y - axis$ of 15,000 euro to 1 day.

Table 5.7: Viaduct case study, scenario 2, sensitivity analysis. Means and standard deviations of the total cost assuming independence, means and standard deviations of total cost and total time assuming correlation and normalized differences in total cost standard deviation for different numbers of units. With increasing number of units, the normalized difference in total cost standard deviation increases to a maximum of 15.5%.

Units	independent cost [euro]		correlated cost [euro]		Δ [%]	time [days]	
	mean	st. dev.	mean	st. dev.		mean	st. dev.
2	2,608,000	106,200	2,608,200	108,400	2.1%	238	9.6
9	5,059,400	131,200	5,060,000	142,100	8.3%	339	10.4
30	12,414,000	182,000	12,414,000	203,000	11.5%	643	12.6
50	19,418,000	221,000	19,420,000	254,000	14.9%	932	13.9
100	36,940,000	304,000	36,938,000	351,000	15.5%	1,656	17.3

One needs to note that the results vary slightly from simulation to simulation due to the random generation. The number of 5,000 samples per simulation bounds the total cost standard deviation within $\pm 1\%$ in 10 out of 10 simulations. The following observations can be made:

- As expected the mean and the standard deviation of the total cost (independent and correlated) and the total time increase with increasing number of units, i.e. increasing viaduct length (Table 5.7).
- With increasing number of units, i.e. increasing viaduct length, the normalized difference between the total cost standard deviation for correlated pile set cost, footing cost and pier cost, and the total cost standard deviation for independent pile set cost, footing cost and pier cost, Δ , increases (Table 5.7). This is visible in the total cost frequency plots, which with increasing number of units, u , become flatter to indicate a larger spread of the sample (Figure 5-18).
- The normalized difference between the total cost standard deviations, Δ , as a function of the

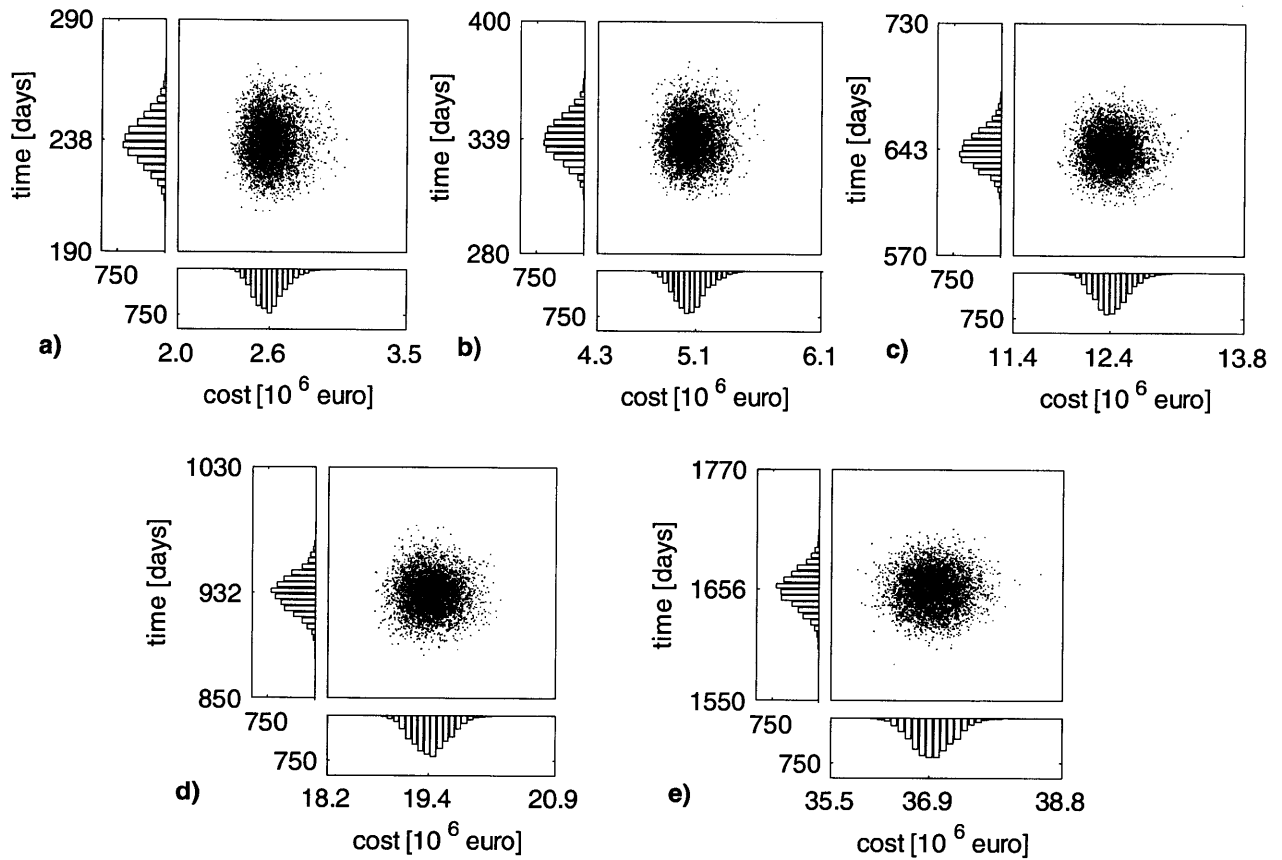


Figure 5-18: Viaduct case study, scenario 2, sensitivity analysis. Scatterplots and frequency plots of the total cost and total time for a) 2-unit, b) 9-unit, c) 30-unit, d) 50-unit, and e) 100-unit viaduct. With increasing number of units, the spread of the total cost increases. Note that the graphs are plotted with ratio $x - axis$ to $y - axis$ of 15,000 euro to 1 day to ensure comparability.

number of viaduct units, u , is shown in Figure 5-19. The data shown in Figure 5-19 are from simulations involving 100,000 samples. The larger number of samples is required to make certain that the normalized difference in standard deviations, Δ , is within $\pm 0.5\%$ in 10 out of 10 simulations. Despite this variation, the following observations can be made:

1. The normalized difference in total cost standard deviations, Δ , increases with increasing number of units, u .
2. The rate of increase of Δ reduces with increasing number of units, u .

These two observations can be explained with the following logic:

$$\uparrow \frac{\text{Correlated Input Costs}}{\text{Independent Input Costs}} \implies \uparrow \text{Standard Deviation of Total Costs, } s, \implies \uparrow \Delta = \frac{s - s_r}{s_r},$$

i.e. the larger the ratio of correlated input costs to independent input costs is, the larger the standard deviation of the total cost is and the larger the normalized difference between the total cost standard deviations for correlated pile set cost, footing cost and pier cost, and the total cost standard deviation for independent pile set cost, footing cost and pier cost, Δ .

The viaduct total cost consists of the costs of 2 abutments, 2 technical blocks, $(u + 1)$ deck sections, u piers, u foundation footings, u foundation pile sets, the finishing, and the construction process (Table 5.5). As described above, scenario 2 assumes that the pile sets costs, the footings costs and the piers costs are correlated, whereas the rest of the costs (abutments, technical blocks, deck sections, finishing, construction process) are independent. Thus, the ratio of correlated input costs to independent input costs is equal to the ratio of the cost of u piers, u foundation footings and u foundation pile sets to the cost of 2 abutments, 2 technical blocks, $(u + 1)$ deck sections, the finishing, and the construction process. Following the logic detailed above, with increasing u the ratio of correlated input costs to independent input costs increases, so that the normalized difference between the total cost standard deviations increases. This explains observation 1.

With u further increasing, the ratio of correlated input costs to independent input costs approaches some number determined by the ratio of the cost of u piers, u foundation footings, u foundation pile sets (correlated) to the cost of $(u + 1)$ deck sections (independent). Thus, as the ratio of correlated input costs to independent input costs approaches this number, the increase rate of the normalized difference in total cost standard deviations is expected to reduce. This explains

observation 2.

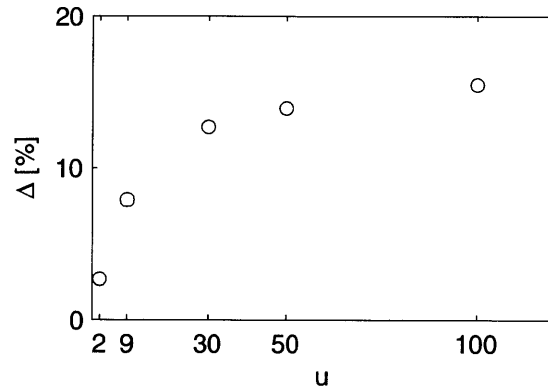


Figure 5-19: Normalized difference between the total cost standard deviation for correlated pile set cost, footing cost and pier cost, and the total cost standard deviation for independent pile set cost, Δ , as a function of the viaduct unit, u . Δ increases with increasing number of units, u ; however, the rate of increase of Δ reduces with increasing number of units, u .

The sensitivity analysis has shown that in scenario 2 (correlation between the costs of different activities - correlation type 2) the normalized difference in standard deviation further increases with increasing viaduct length. However, this increase is not very large. In the next section, a similar sensitivity analysis is conducted to investigate scenario 3.

Scenario 3 - Sensitivity Analysis

Scenario 3 (correlation between the cost and the time of an activity - correlation type 5) has shown that the input cost and the input time are correlated, the total cost and the total time will also be correlated, albeit with different correlation coefficient. The goal of this sensitivity analysis is to further investigate the change in correlation between total cost and total time as a function of viaduct length, i.e. the number of units in the viaduct. A unit consists of a pile set, a footing and a pier (Figure 5-16).

In scenario 3, costs are independent, times are independent but the cost and the time of each element i are correlated. The correlation between the cost and the time of each element i is assumed to be 0.8 so that the correlation matrix of each element is (as in expression 5.47):

$$R_i = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}. \quad (5.52)$$

6,000-sample simulations were run for viaducts with 2, 9, 30, 50 and 100 units, respectively. The number of 6,000 samples has been selected because it bounds the correlation between total cost and total time within $\pm 2\%$ in 10 out of 10 simulations. The means, the standard deviations of total cost and total time, and correlations between total cost and total time for different number of units, u , are summarized in Table 5.8. The scatterplot, the frequency plots of cost and time are shown in Figure 5-20. The total cost and the total time vary with the number of units. Thus, the scale on the axes in the five graphs are different. Nevertheless, comparability is obtained by plotting results with the same scale ratio, i.e. with ratio $x - axis$ to $y - axis$ of 15,000 euro to 1 day.

Table 5.8: Viaduct case study, scenario 3, sensitivity analysis. Means and standard deviations of total cost and total time, and correlations between total cost and total time for different numbers of units. With increasing number of units, the correlation between total cost and total time increases to a maximum of 0.56.

Units	total cost [euro]		total time [days]		total cost – total time correlation [-]
	mean	st. dev.	mean	st. dev.	
2	2,608,800	106,000	237	9.8	0.28
9	5,058,200	129,200	339	10.5	0.37
30	12,413,000	182,000	643	12.4	0.45
50	19,427,000	223,000	932	14.0	0.52
100	36,936,000	299,000	1,655	17.1	0.56

One needs to note that the results vary slightly from simulation to simulation due to the random generation. However, the number of 6,000 samples per simulation bounds the correlation between total cost and total time within $\pm 2\%$ in 10 out of 10 simulations. The following observations can be made:

- As expected the mean and the standard deviation of the cost and the time increase with increasing number of units, i.e. increasing length of the viaduct (Table 5.8).
- With increasing number of units in the viaduct, i.e. increasing length of the viaduct, the correlation between total cost and total time increases (Table 5.8). This is also visible in Figure 5-20, where the cloud of data points becomes more aligned along a more inclined

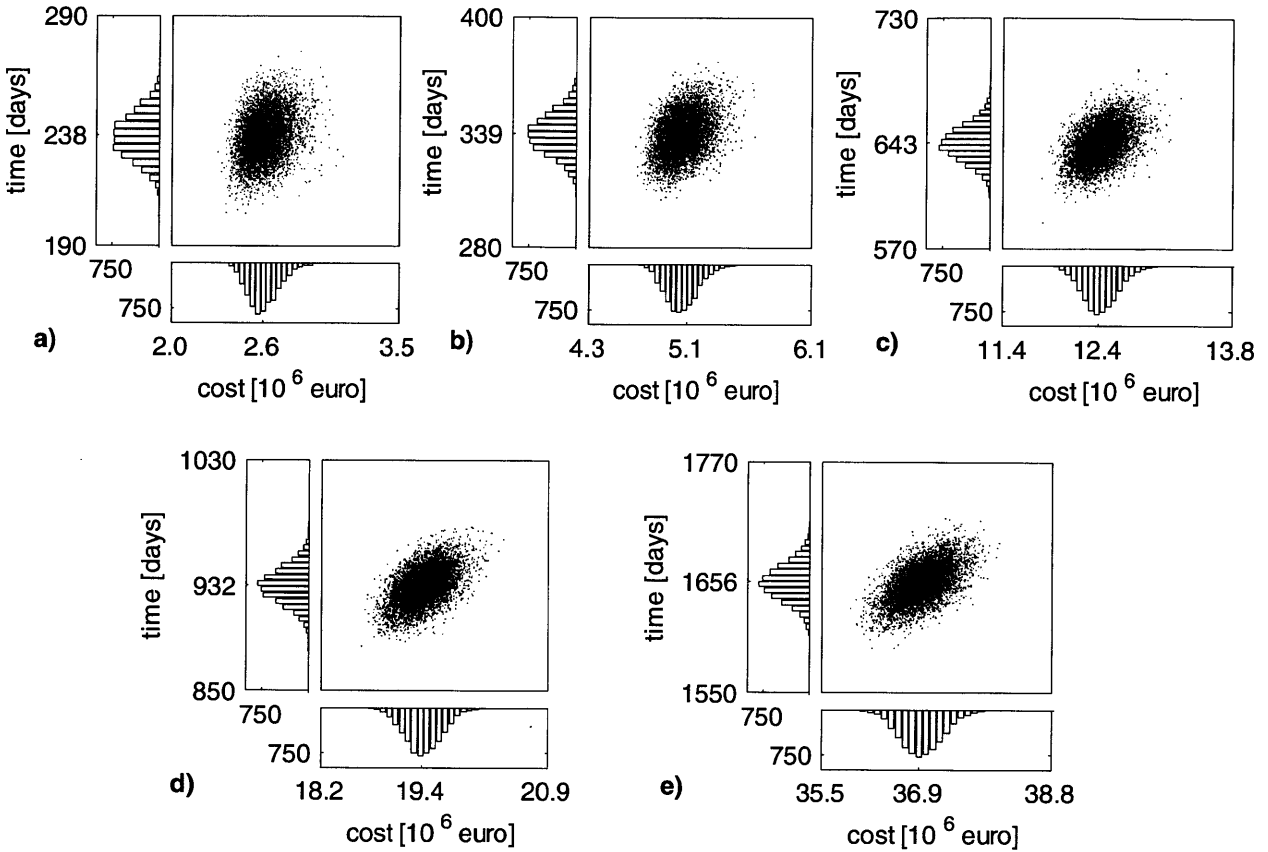


Figure 5-20: Viaduct case study, scenario 3, sensitivity analysis. Scatterplots and frequency plots of the total cost and total time for a) 2-unit, b) 9-unit, c) 30-unit, d) 50-unit, and e) 100-unit viaduct. With increasing number of units, the correlation between total cost and total time increases. Note that the graphs are plotted with ratio $x - axis$ to $y - axis$ of 15,000 euro to 1 day to ensure comparability.

direction with increasing number of units, u .

- The correlation between total cost and total time, ρ , as a function of the number of viaduct units, u , is shown in Figure 5-21. One can see that the correlation between total cost and total time increases with number of units. The correlation between total cost and total time at $u = 100$ does not reach 0.8, the value of the correlation between activity cost and activity time. However, the correlation between total cost and total time might reach 0.8 for a larger number of units, u .

Scenario 1 and scenario 3 are compared again in Table 5.9. Also in Table 5.9 the normalized

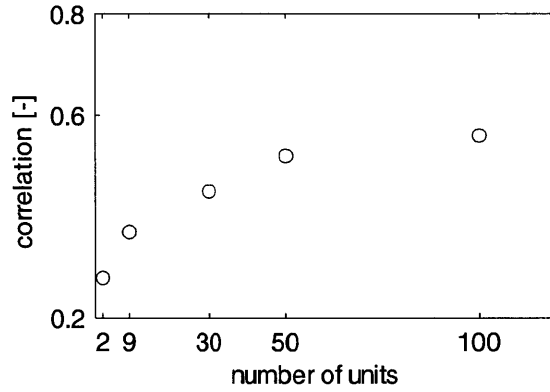


Figure 5-21: Correlation between total cost and total time, ρ , as a function of the viaduct unit, u . The correlation increases with number of units; however, for $u = 100$ it does not reach 0.8, the correlation between activity cost and activity time.

differences, Δ , in mean and percentiles are reported. The normalized difference is calculated as

$$\Delta = \frac{p - p_r}{p_r}, \quad (5.53)$$

where p is the mean or percentile of scenario 3 and p_r is the mean or percentile of scenario 1.

Table 5.9: Viaduct case study, scenarios 1 and 3. Means, 80th and 95th percentiles of total cost and total time, and correlations between total cost and total time for scenario 1 and scenario 3. The correlation between total cost and total time does not impact the marginal distributions of the total cost and the total time.

Scenario	cost percentiles [euro]			time percentiles [days]			total cost – total time correlation [-]
	mean	80 th	95 th	mean	80 th	95 th	
1	5,060,500	5,160,700	5,286,900	339	348	357	0.00
3	5,059,500	5,165,100	5,280,500	339	348	357	0.38
Δ [%]	0.0	0.1	-0.1	0.0	0.0	0.0	-

The results show that despite the correlation between the total cost and the total time of 0.38 in scenario 3, the means, the 80th and the 95th percentiles of the total cost and the total time are very similar to the ones of scenario 1, which has uncorrelated total cost and total time (Table 5.9). In fact, the correlation between cost and time of each activity impacts the cumulative distribution of total cost and total time but it does not impact the marginal distribution of the total cost and the marginal distribution of the total time. Thus, if one is interested in the means and/or percentiles

of the total cost and/or the total time, modeling correlation between activities cost and time is not needed. This is confirmed by the total cost and total time frequency plots in Figure 5-17a and Figure 5-17c. Compared to scenario 1, the additional information scenario 3 produces is that large total costs occur frequently with large total time.

For the case of transportation infrastructure projects, it should be noted that activity cost and activity time are often not correlated.

The sensitivity analysis has shown that in scenario 3 (correlation between the cost and the time of an activity - correlation type 5) the correlation between total cost and total time further increases with increasing viaduct length. Despite the increase, the effect of the correlation between total cost and total time is not as large as for the correlation between activity cost and activity time. It has also been shown that the correlation between the total cost and the total time does not impact the marginal distributions of total cost and total time.

Summary

The main observations from the viaduct case study are:

1. Correlation between costs of activities causes the standard deviation of the total cost to increase. This had been pointed out by Touran (1993) and Newton (1992), and it has been confirmed here. A new observation is that with increasing number of units including activities with correlated costs, the total cost standard deviation increases. For an increasing number of units, the increase rate of the total cost standard deviation reduces. For a correlation coefficient of 0.8 between the activities costs, the total cost standard deviation in the correlated scenario is 17% larger than the total cost standard deviation in the independent scenario ($u = 100$).
2. Correlation between cost and time of the activities is reflected in the correlation between total cost and total time. The correlation between total cost and total time increases with increasing length (i.e. increasing number of activities with correlated cost and time) but does not reach the value of the correlation between cost and time of the activities. However, it might reach the correlation between cost and time of the activities for a larger number of units.

- Correlation between cost and time of each activity does not impact the marginal distributions of the total cost and total time. Thus, modeling correlation between activities cost and time is not needed if one is interested in the means and/or percentiles of the total cost and the total time.

5.4.3 Assessment of the Correlations' Impact: Tunnel Case Study (Correlation Type 1)

In this section, correlation type 1 (correlation between the costs of a repeated activity) is investigated with a tunnel case study. The impact of correlation type 1 on the total cost for the construction of the tunnel will be analyzed with different scenarios and detailed with a sensitivity analysis.

The case study is a tunnel of the new High Speed Rail in Portugal. The tunnel is 500m long. The construction of the tunnel is modeled with one activity that includes all processes of constructing one meter of tunnel. The tunnel activity is characterized by a cost and an advance rate. Along the tunnel, cost and advance rate vary according to the geological conditions. The geological conditions are distinguished between good, medium and poor. The costs per meter, the advance rates and the lengths along the tunnel where these geological conditions are encountered are given in Table 5.10. The costs per meter and the advance rates have been estimated by a tunnel engineer. The lengths along the tunnel where the geological conditions are encountered have been estimated by a geologist.

Table 5.10: Cost and advance rate of tunnel excavation depending on the geological conditions. They are used to calculate the deterministic total cost and total time of the tunnel construction.

Geological conditions	length [m]	percentage [%]	cost [euro/m]	advance rate [m/day]
good	200	40	7,250	4
medium	250	50	11,000	2
poor	50	10	17,000	1

In order to evaluate the impact of correlations on the distributions of total cost and total time, a probabilistic analysis is needed. The cost and advance rate data available are deterministic. Assumptions have been made to construct probability distributions of costs and advance rates:

- The underlying distribution of cost variables is lognormal. The lognormal distribution is

estimated as explained in the viaduct case study (section 5.4.2).

- The underlying distribution of advance rate variables is approximated with the triangular distribution. The triangular distribution is estimated as explained in the viaduct case study (section 5.4.2).

In the following, the total cost and the total time for the construction of the tunnel are calculated, first deterministically and then probabilistically. In the probabilistic calculations, correlation type 1 (correlation between the costs of a repeated activity) is modeled with scenarios and analyzed in detail with a sensitivity analysis.

Deterministic Calculation

The total cost of the tunnel construction is given by the sum of the costs per meter. The total time is given by the sum of the times per meter (it is assumed construction begins at a portal and finishes at the other portal). The total cost and total time are €5,050,000 and 225 days, respectively (Table 5.11).

Table 5.11: Deterministic total cost and total time of tunnel construction. They will be compared with the the probabilistic analyses of total cost and total time.

Geological conditions	length [m]	cost [euro/m]	advance rate [m/day]	total cost [euro]	total time [days]
good	200	7,250	4	1,450,000	50
medium	250	11,000	2	2,750,000	125
poor	50	17,000	1	850,000	50
total	500	–	–	5,050,000	225

Probabilistic Calculation

The range of total cost and total time for the construction of the tunnel is estimated for four different scenarios. These are followed by a sensitivity analysis.

The probabilistic calculation is based on Monte Carlo simulations with 6,000-sample runs. The number of 6,000 samples per simulation has been selected because it bounds the total cost standard deviation within $\pm 1\%$ from simulation to simulation in 10 out of 10 simulations. The types of correlation described in section 5.4.1 for a tunnel are translated into scenarios for the construction of the 500m long tunnel:

- Scenario A is the reference scenario. Cost and time variables of each activity are independent.
- Scenario B represents a correlation between the costs of a repeated activity (correlation type 1 in Table 5.4). The costs of a repeated activity are assumed to be correlated due to the repetitiveness of the process. For example, the tunnel excavation in good geological conditions is repeated 200 times along the 200 meters. Thus, the activity is repeated 200 times and the costs of the 200 repetitions are correlated. A similar reasoning is applied to the excavation in medium and poor geological conditions. Three cases within scenario B are considered:

1. The correlation matrix, R_r , between costs of repeated activities is equal to the identity matrix, I . This case corresponds to scenario A.

$$R_r = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots \\ \dots & 0 & 1 & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}. \quad (5.54)$$

2. The cost of constructing one meter of tunnel (unit length) is correlated with the costs of constructing the preceding and following meters. The correlation decreases by 0.01 for every meter (unit length) of distance. Thus, the correlation matrix, R_r , between costs of the repeated activities is equal to the following matrix:

$$R_r = \begin{bmatrix} 1 & 0.99 & \dots & 0.01 & 0 & \dots & 0 \\ 0.99 & 1 & 0.99 & \dots & 0.01 & 0 & \dots \\ \dots & 0.99 & 1 & 0.99 & \dots & 0.01 & 0 \\ 0.01 & \dots & 0.99 & 1 & 0.99 & \dots & 0.01 \\ 0 & 0.01 & \dots & 0.99 & 1 & 0.99 & \dots \\ \dots & 0 & 0.01 & \dots & 0.99 & 1 & 0.99 \\ 0 & \dots & 0 & 0.01 & \dots & 0.99 & 1 \end{bmatrix}. \quad (5.55)$$

3. The cost of constructing one meter of tunnel (unit length) is fully correlated with the costs of constructing the preceding and following meters. Thus, the correlation matrix,

R_r , between costs of repeated activities is equal to a matrix of ones:

$$R_r = \begin{bmatrix} 1 & \dots & 1 \\ \dots & 1 & \dots \\ 1 & \dots & 1 \end{bmatrix}. \quad (5.56)$$

The mean of the total cost and the mean of the total time are equal to the sum of the costs means and the costs times, respectively. The costs means and the times means do not change from scenario to scenario. Thus, the mean of the total cost and the mean of the total time are not expected to vary in the three scenarios. The standard deviation of the total time is also not expected to vary because the time variables are assumed independent in the three scenarios. In scenario B, the standard deviation of the total cost is expected to increase from scenario B1 to scenario B3. These expectations are confirmed by the results detailed below.

The means and the standard deviations of total cost and total time are given in Table 5.12. The normalized differences, Δ , in standard deviation are also reported in Table 5.12. The normalized difference in standard deviations is calculated as

$$\Delta = \frac{s - s_r}{s_r}, \quad (5.57)$$

where s is the standard deviation of the scenario and s_r is the standard deviation of the reference scenario (scenario A).

The scatterplots, the frequency plots of cost and time in scenario A, scenarios B1, B2 and B3 are shown in Figure 5-22.

Table 5.12: Correlation type 1, scenarios A, B1, B2, and B3. Means and standard deviations of the total cost and total time, and normalized differences in total cost standard deviation for scenario A and scenarios B1 to B3. When modeling correlation type 1, the standard deviation increases dramatically: 757% in scenario B2 and 1,299% in scenario B3.

Scenario	cost			time	
	mean [euro]	st. dev. [euro]	Δ [%]	mean [days]	st. dev. [days]
A	5,492,800	35,600	–	232	1.2
B1	5,491,200	35,800	0.6	232	1.2
B2	5,492,500	305,100	757	233	1.3
B3	5,493,600	498,000	1,299	232	1.2

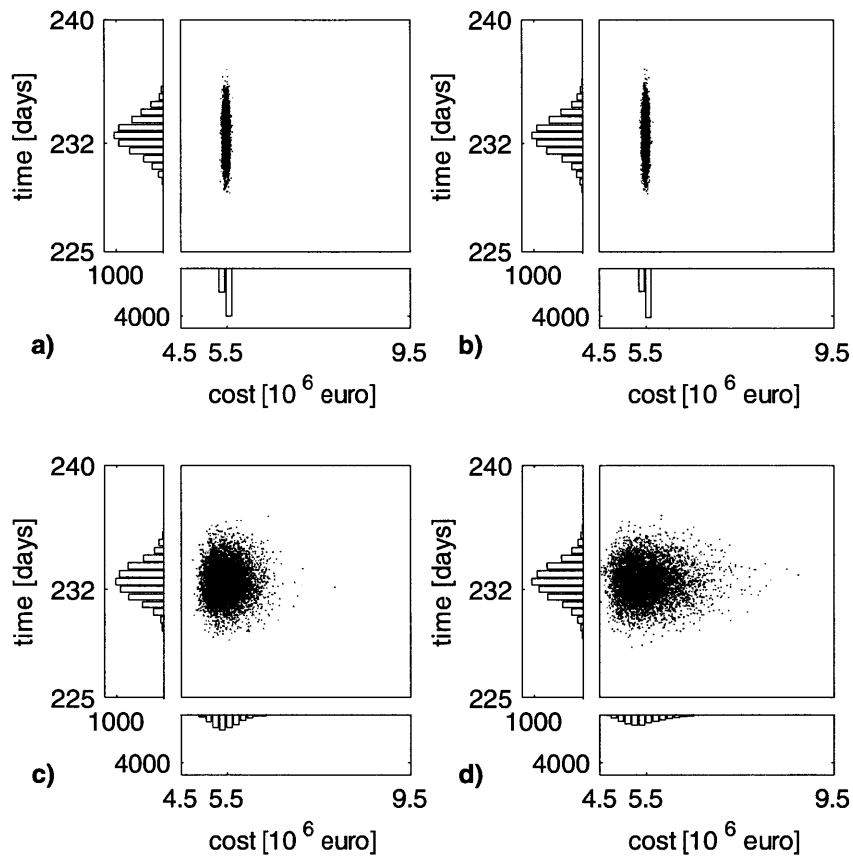


Figure 5-22: Correlation type 1, scenarios A, B1, B2, and B3. Scatterplots and frequency plots of the total cost and total time for a) scenario A, b) scenario B1, c) scenario B2, and d) scenario B3. When modeling correlation type 1 (scenarios B2 and B3), the spread of the total cost dramatically increases compared to the scenarios modeling independence (scenarios A and B1).

One needs to note that the results vary slightly from simulation to simulation due to the random generation. In fact, the number of 6,000 samples per simulation bounds the total cost standard deviation within $\pm 1\%$ in 10 out of 10 simulations. The following observations can be made:

- The mean of the total cost can be considered constant from scenario to scenario (see Table 5.12). This is confirmed by a constant center of gravity in the cost frequency plots (Figure 5-22).
- The standard deviation of the total cost increases dramatically from scenario B1 to scenario B3 (see Table 5.12). The increase in total cost standard deviation is clearly visible in the scattergrams and in the total cost frequency plots (Figure 5-22). Scenario B1 involves a correlation matrix, R_r , equal to the identity matrix, I , i.e. it models uncorrelated cost variables. Scenario B3 involves a correlation matrix, R_r , equal to a matrix of ones, i.e. it models fully correlated cost variables. Scenario B1 and scenario B3 can be considered the lower and upper end of the spectrum of possible scenarios. In the scattergram and the total cost frequency plot of scenario B1 the sample is concentrated around the mean, whereas in scenario B3 it is much more scattered (Figures 5-22b and 5-22d). The scattergram and the total cost frequency plot of scenario B3 are more scattered on the right side than on the left side because the input cost distributions are lognormal distributions. (Note that for a larger number than used here of input cost distributions, the total cost frequency plot would approach a normal distribution (Central Limit Theorem, (see e.g. Bertsekas & Tsitsiklis 2002)) so that the scattergram and the total cost frequency plot would be symmetric). The difference in total cost standard deviations between scenario B1 and scenario B3 is more than one order of magnitude, with a normalized difference in standard deviation of 1,300% (Table 5.12).
- Scenario B2 has a correlation matrix between the extreme cases of the identity matrix (scenario B1) and the matrix of ones (scenario B3). This is reflected in the scattergram and the total cost frequency plot where the samples of scenario B2 are more spread than the sample in scenario B1 and more concentrated than the sample in scenario B3 (Figure 5-22). This is confirmed by the magnitude of the normalized difference in standard deviation, which are lower than in scenario B3 (see Table 5.12). It is likely that the normalized difference in standard deviation depends on the tunnel length. This aspect is further investigated in "Scenario

B1 to B3 - Sensitivity Analysis".

- Scenario A and scenario B1 model the same case. This is reflected in means, standard deviations, scattergrams and frequency plots (see Table 5.12, and Figures 5-22a and 5-22b). The standard deviations are not exactly the same but can be considered equal because the normalized difference, Δ , of 0.6% is within the $\pm 1\%$ limit set by the choice of 6,000 samples per run.
- The means of total cost and total time are larger than the deterministic total cost and total time (see "Deterministic Calculation"). The means are larger than the deterministic results for the same reasons explained in section 5.4.2.

Four scenarios in the construction of a tunnel have been analyzed. It has been found that the standard deviation of the total cost can increase dramatically due to correlation type 1 (correlation between the costs of a repeated activity). The impact of correlation type 1 in scenarios B1 to B3 is further investigated in a sensitivity analysis.

Scenarios B1 to B3 - Sensitivity Analysis

The geological conditions of good, medium and poor are assumed for all scenarios in the proportions as in Table 5.12, i.e. 40% in good, 50% in medium and 10% in poor geological conditions. The costs per meter excavation, the advance rates and the correlation matrices are the same as in the scenarios B1 to B3.

14,000-sample simulations were run for tunnel lengths of 200m, 500m, 1,000m, 2,000m and 3,000m, respectively. In this case the number of 14,000 samples is the maximum number of samples allowed by the program for a length of 3,000m. 14,000 bounds the normalized difference in total cost standard deviation, Δ , within $\pm 30\%$ (scenario B2) and $\pm 125\%$ (scenario B3) in 10 out of 10 simulations. The standard deviations of total cost in the scenarios B1 to B3 for the different tunnel lengths are summarized in Table 5.13. Table 5.13 also shows the normalized differences in standard deviations, Δ . The normalized difference in standard deviations is calculated as

$$\Delta = \frac{s - s_r}{s_r}, \quad (5.58)$$

where s is the standard deviation of a scenario at a given length, and s_r is the standard deviation of the reference scenario (scenario B1) at the same length.

Table 5.13: Correlation type 1, scenarios B1, B2, and B3, sensitivity analysis. Total cost standard deviations and normalized differences in standard deviations for different number of units. With increasing number of units, the standard deviation increases dramatically: up to 3,243% in scenario B3.

length [m]	Total cost standard deviation [10^3 euro]					Δ [%]				
	200	500	1,000	2,000	3,000	200	500	1,000	2,000	3,000
scenario B1	22	35	51	71	86	–	–	–	–	–
scenario B2	160	296	449	670	826	614	738	787	849	857
scenario B3	192	488	977	1,936	2,880	759	1,279	1,829	2,642	3,234

One needs to note that the results vary slightly from simulation to simulation due to the random generation. The number of 14,000 samples per simulation bounds the normalized difference in total cost standard deviations, Δ , within $\pm 30\%$ (scenario B2) and $\pm 125\%$ (scenario B3) in 10 out of 10 simulations. The following observations can be made:

- The total cost standard deviation depends on both the tunnel length and the correlation matrix:
 - The total cost standard deviation increases with increasing tunnel length.
 - The total cost standard deviation increases from scenario B1 (the correlation matrix is equal to the identity matrix) to scenario B3 (the correlation matrix is equal to a matrix of ones).

This is confirmed by the results in Table 5.13 and Figure 5-23.

- The normalized difference in standard deviations, Δ , increases from scenario B2 to scenario B3. The normalized difference in standard deviations, Δ , also increases with tunnel length (see Table 5.13 and Figure 5-23). With increasing tunnel length, the increase rate of Δ decreases faster in scenario B2 than in scenario B3 (see Table 5.13 and Figure 5-23). This observation can be explained with the correlation matrix:
 - In scenario B3, for all tunnel lengths the correlation matrix is equal to a matrix of ones (all cost variables are fully correlated) so that with increasing size of the correlation matrix the normalized difference in standard deviations, Δ , increases.

- The correlation matrix of scenario B2 consists of zeros and a band diagonal with correlation coefficients larger than zero (see expression 5.55). In scenario B2, the correlation matrix increases in size with increasing tunnel length but the diagonal band maintains the same width. With increasing size, the number of independent cost variables and correlated cost variables increase. However, the number of correlated variables is limited by the width of the diagonal band. Thus, with increasing tunnel length the number of independent cost variables increases faster than the number of correlated cost variables so that the portion of correlation coefficients larger than zero in the matrix effectively decreases.

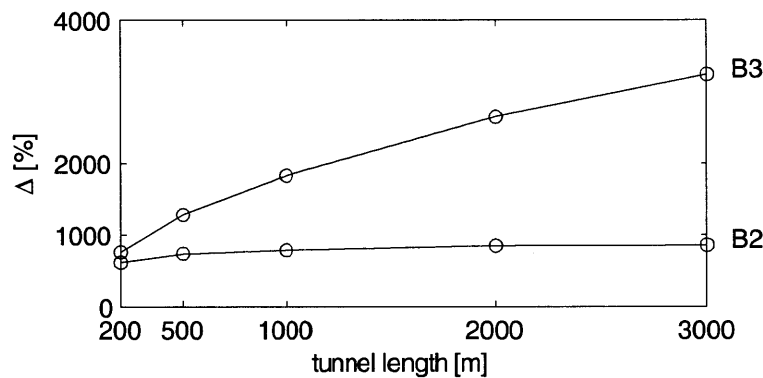


Figure 5-23: Correlation type 1, scenarios B2 and B3, sensitivity analysis. Normalized difference in total cost standard deviations, Δ , as a function of the tunnel length. Δ increases from scenario B2 to scenario B3. It also increases with tunnel length. However, with increasing tunnel length, the increase rate of Δ decreases faster in scenario B2 than in scenario B3.

The sensitivity analysis has shown that correlation type 1 (correlation between the costs of a repeated activity) causes the normalized difference in standard deviation to further increase with increasing tunnel length. The increase remains dramatic exceeding 3,200% in one of the scenarios for a 3,000 m long tunnel.

Summary

The tunnel case study has confirmed that the correlation between costs of activities causes the standard deviation of the total cost to increase (Newton 1992, Touran 1993). It has been observed

that the correlation model (independent, partially correlated, or fully correlated) is critically important, when activities are repeated in the construction of a structure. The longer the structure, i.e. the more repetitions of the activities, the more critical the impact of correlation on the total cost standard deviation. The impact of correlation type 1 on the total cost standard deviation is significantly larger than the impact of correlation type 2 (correlation between the costs of different activities; section 5.4.2).

5.5 Summary and Conclusions

In rail line construction, the standard deviation of the total costs can be underestimated if the costs of the activities are assumed to be independent when they are correlated. The Spearman correlation coefficient and the NORTA method have been identified as the most appropriate correlation measure and model, respectively, to model correlations in rail line construction.

Spearman correlations and NORTA have been applied to three types of correlation in construction (correlation between the costs of a repeated activity, correlation between the costs of different activities, and correlation between the cost and the time of an activity) to assess the impact of these correlations on the total cost distribution. The following has been found:

- Correlation between the costs of a repeated activity. This correlation type has a dramatic impact on the total cost distribution: in the case study, the spread of the total cost distribution exploded while the standard deviation increased by an order of magnitude. For an increasing number of correlated costs, the total cost standard deviation further increased significantly.
- Correlation between the costs of different activities. This correlation type has a limited impact on the total cost distribution: in the case study, the spread and the standard deviation of the total cost increased slightly. For an increasing number of correlated costs, the total cost standard deviation increased further, albeit marginally.
- Correlation between the cost and the time of an activity. This correlation type does not have any impact on the total cost standard deviation. However, it causes the correlation between the total cost and the total time to increase. In the case study, this did not reach the magnitude of the correlation between activity cost and time.

Due to limited or no impact on the total cost, the second and the third correlation types are not further pursued. Given the extremely large impact on the total cost distribution, the correlation between costs of a repeated activity has been further pursued and modeled as a source of uncertainty in the uncertainty model (chapter 4).

Chapter 6

Simulation Tool

A construction model (chapter 3) and an uncertainty model (chapters 4 and 5) have been developed to capture the uncertainties in the construction of rail lines. They need to be implemented in a simulation tool in order to apply them to construction projects of rail lines. The author reviewed the simulation tools currently available, extended the framework of the Decision Aids for Tunneling (DAT), and implemented the construction and uncertainty models in the DAT. This chapter presents the available simulation tools, details the working framework of the DAT, and explains how the construction and the uncertainty models are implemented in the DAT. The construction model of the four main structures of rail lines (tunnels, viaducts, cuts and embankments) is implemented with construction networks and activity networks. The uncertainty model representing the three sources of uncertainty (variability, cost correlations, and disruptive events) is implemented through the introduction in the DAT of the lognormal distribution and the NORTA correlation model, and through the representation of the occurrence of disruptive events with Markov processes. Hence, the construction model and the uncertainty model have been successfully implemented in the simulation tool DAT. They will be applied to the construction project of the new Portuguese high speed rail line in chapter 7.

The detailed structure of the chapter is as follows: first, the available simulation tools are described and compared against the requirements to be fulfilled (section 6.1). Second, the framework of the DAT is described (section 6.2) and extended to model viaducts, cuts and embankments (section 6.3). Then, the construction model and the uncertainty model are implemented (sections 6.4 and 6.5).

6.1 Simulation Tools to Model Construction Projects

In this section, a brief history of the development of construction simulation tools is given. After listing some generally desirable characteristics for a construction simulation tool, five specific requirements for the simulation tool used to develop the research work presented in this thesis are discussed. The available simulation tools are compared against the five specific requirements. One simulation tool will emerge as best suited since it complies with all but one requirement, which can be relatively easily implemented. Finally, it will be shown that the selected simulation tool also fulfills the majority of the generally desirable characteristics for a construction simulation tool.

AbouRizk (2010) recognizes three phases in the development of simulation tools to model construction projects. In a first phase in the 1970s, simulation tools consisted of elements to model work tasks and their logical relationship. An example of such tools is CYCLONE (Halpin & Woodhead 1976). In a second phase starting in the 1990s, simulation tools provided more flexible user interfaces and offered more simulation capabilities, e.g. the possibility to define and track resources. An example of a second generation tool is STROBOSCOPE (Martinez & Ioannou 1999). In the third phase during the last decade, tools were developed by integrating existing construction simulation tools with 3D visualization tools, such as AutoCAD. In parallel to the development described by AbouRizk (2010), other tools specifically created to model tunnel construction projects have been successfully implemented. The earliest example is the Tunnel Cost Model (Moavenzadeh et al. 1974) followed by the Decision Aids for Tunneling (Einstein et al. 1999). From their inception, the latter have been further developed to include additional capabilities (Einstein et al. 2011): the DAT can model geology and construction uncertainties; they can update the simulation with newly gained information; they have been applied to complex tunnel construction projects; and recently, they have been expanded to model linear/networked construction projects.

The construction simulation tools developed over the years have different features. These can be compared against generally desirable characteristics. Ideally, a simulation tool should have the following characteristics (Einstein et al. 2011):

- Manage project complexity
 - representation and optimization of construction cost, -time and resource usage
 - consideration of a variety of uncertainties

- simulation of various levels of detail
- simulation in real time
- User friendliness
 - no need of programming/coding by the user
 - user-friendly interface
 - input through tabular and/or graphical means
 - web access
 - compatibility with commercial programs such as AutoCAD, MS Excel, etc.
- Presentation of simulation results
 - types and variety of simulation output
 - visualization of simulation results
 - ease of transfer of simulation results
 - possibility to select the needed output
- Flexibility and expendability
 - structure organized in modules or similar
 - applicability to different types of civil structures.

In addition to these generally desirable characteristics, for the research work presented in this thesis the simulation tool must fulfill some specific requirements by the construction model (chapter 3) and the uncertainty model (chapters 4 and 5). These specific requirements are first listed and then compared against the capabilities of the simulation tools.

The construction model and the uncertainty model have the following requirements for the simulation tool:

- Construction model
 - The construction process depends on the geology. The simulation tool must be able to model this dependence relation and its impact on the cost and the time of construction.

- The construction process consists of sequential and parallel activities. These must be modeled by the simulation tool.
- Uncertainty model
 - The cost and time of activities are highly uncertain. This uncertainty is modeled with probability distributions of costs and times.
 - If costs are assumed independent when they are in fact correlated, the standard deviation of the total costs is underestimated. The simulation tool must be able to model correlations between costs.
 - Disruptive events can significantly increase the total cost and total time. They are events with large impact on cost and time and usually small probability of occurrence. Due to their large impact, the simulation tool must be able to model disruptive events.

Among the simulation tools available, there is not one that complies with the five specific requirements (Table 6.1). The simulation tools discussed in the following can be subdivided into two types: simulation tools used to represent construction processes (CYCLONE, STROBOSCOPE, SIMPROCESS and DAT) and simulation tools developed to model problems characterized by uncertainty (@RISK, @RISK for Project, Crystal Ball, and Pertmaster).

Developed for construction projects, the simulation tools CYCLONE (and its siblings INSIGHT, RESQUE, COOPS and CIPROS), STROBOSCOPE and SIMPROCESS model the construction process and calculate the cost and/or time of construction. CYCLONE (Halpin & Woodhead 1976) models construction processes including parallel activities, whose time is modeled probabilistically. STROBOSCOPE (Martinez & Ioannou 1999) can model more complex construction processes than CYCLONE, as well as the resources handling during construction. It is capable to model the uncertainty in cost, time and resources with probability distributions. SIMPROCESS is a commercially available system developed for business process simulation, which was modified to represent construction tasks and processes (Slaughter 1999). It has been applied to cyclic activities, whose cost and time are calculated. At different degrees, CYCLONE, STROBOSCOPE and SIMPROCESS comply with two of the five requirements, namely modeling sequential and parallel activities, and modeling the uncertainty of cost and time with probability distributions (Table 6.1).

Another simulation tool developed specifically for construction projects are the Decision Aids for Tunneling (DAT), developed at the Massachusetts Institute of Technology (Einstein 2004). As their name indicates, the DAT are a simulation tool to model the construction of tunnels; however, they are suitable for the modeling of any linear, networked project, such as transportation projects. They model the geology uncertainty and its impact on cost and time, sequential and parallel activities, the uncertainty of cost and time as well as the occurrence of disruptive events. They do not model correlations; hence, they satisfy only four out of five requirements (Table 6.1).

@RISK, Crystal Ball, @RISK for Project and Pertmaster are simulation tools that have not been developed specifically to model construction processes but more generally for a variety of problems that include uncertainty. @RISK by Palisade and Crystal Ball by Oracle are Excel add-ons that allow one to associate each cell in the spreadsheet with a probability distribution, to model the correlation between cells as well as the occurrence of disruptive events and their impact. Being add-ons based on a spreadsheet, they are not ideal to model sequence and dependence between activities. Superior in this last aspect are @RISK for Project by Palisade and Pertmaster by Primavera. In addition to including all the features of @RISK and Crystal Ball, they can model sequential and parallel activities in Gantt charts (bar charts illustrating the sequence in time of sequential and parallel activities). The length and the position of the horizontal bars in the charts indicate the duration and the start and end times, respectively. Each bar can be attributed a cost so that cost and time of the activities are calculated. @RISK for Project and Pertmaster cannot model geology; thus, similarly to DAT, they satisfy only four requirements (Table 6.1).

The DAT are preferred over @RISK for Project and Pertmaster for a strategic reason. The introduction of the geology and its impact on cost and time in @RISK for Project and Pertmaster is not feasible since these are packaged products that are bought on the market and cannot be modified for research purposes. On the other hand, introducing correlations in the DAT is a feasible task that has been successfully completed.

When comparing the DAT against the generally desirable characteristics for a construction simulation tool, one can see that the DAT feature the large majority of these characteristics. Regarding the management of project complexity, they can represent construction cost, -time and resource usage, and they can optimize them; they can model geology, cost, time and resource uncertainties; they can simulate projects at different levels of detail; and they simulate construction

Table 6.1: Available simulation tools compared to the requirements set by the construction and the uncertainty models. DAT, @Risk for Project and Pertmaster satisfy four out of five requirements: in their current form, the DAT cannot model cost correlations, while @Risk for Project and Pertmaster cannot model the uncertainties derived from geology. The DAT are expanded to model cost correlations as part of the research work presented in this thesis.

Requirements	simulation tools			
	CYCLONE	STROBOSCOPE	SIMPROCESS	DAT
construction model				
geology uncertainty				✓
sequential/parallel activities	✓	✓	✓	✓
uncertainty model				
cost/time prob. distributions	✓	✓	✓	✓
cost correlations				
disruptive events				✓
	@RISK	Crystal Ball	@RISK for Project	Pertmaster
construction model				
geology uncertainty				
sequential/parallel activities			✓	✓
uncertainty model				
cost/time prob. distributions	✓	✓	✓	✓
cost correlations	✓	✓	✓	✓
disruptive events	✓	✓	✓	✓

in real time. Regarding user friendliness, the user does not need to program or code; she/he can access the user-friendly interface with any web browser; she/he can input data in tables or graphically; and the DAT are in part compatible with MS Excel. Regarding simulation results, there are a variety of simulation outputs that can be visualized directly in graphs or downloaded for data processing and visualization; there is also the possibility to select only the needed output to speed up the simulation. Last, the DAT are very flexible thanks to their modular structure and have proved to be easily expandable, e.g. to modeling the construction of linear/networked projects.

In the preceding, requirements for the simulation tools and available simulation tools have been discussed. The DAT have emerged as best suited to implement the construction model and the

uncertainty model. They are described in the next section.

6.2 The Decision Aids for Tunneling (DAT)

The simulation tool used to implement the construction and the uncertainty models is the Decision Aids for Tunneling (DAT). The two major components of the DAT, the geology description and the construction simulation, are presented, and the calculation of the tunnel construction cost and time are described. Then, the uncertainty modeling based on Markov processes and probability distributions is introduced.

The DAT consist of two major components: the description of the geology, and the construction simulation (Einstein 2004). The description of the geology component has three inputs: areas, zones, and ground parameter sets (Figure 6-1). An area, the uppermost level of the organization for input of geology, consists of one or more zones. These are geologically homogeneous lengths of ground. A ground parameter set is, as the name indicates, a set of ground parameters. As explained in the following example, the states of the ground parameters determine the ground class. A ground parameter describes a characteristic of the ground, e.g. lithology, rock mass description (it depends on the presence and the orientation of discontinuities), overburden. It includes several ground parameter states, for example gneiss and granite are the states of the ground parameter lithology. The combination of ground parameter states determines the ground class. This can correspond to more than one combination of ground parameter states; for instance, in Figure 6-1 ground class GC2 corresponds to ground parameter states granite (ground parameter lithology), good (rock mass) and high (overburden), as well as to ground parameter states granite, good rock mass and medium overburden. The profiles of the ground parameters along a tunnel are combined into one ground class profile, which is the output of the description of the geology component in the DAT (Figure 6-1).

The construction simulation component simulates the construction process through the ground class profile (the output of the geology component) (Einstein 2004). It includes the tunnel network, the construction methods and the construction method definition:

- The tunnel network depicts the sequence of tunnels to be constructed. In the simplest case, it includes only one tunnel (Figure 6-2).

	0 km	5 km	10 km			
Area	area 1					
Zone	zone 1		zone 2			
Ground parameters						
Lithology	gneiss		granite			
Rock mass	poor	medium	very good	good		poor
Overburden	low	medium		high	low	
Ground class	GC5	GC4	GC1	GC2	GC3	GC5

Figure 6-1: Description of the geology component in the DAT. Areas, zones and ground parameters (input) determine the ground class profile (output). The ground parameters are described by their states. From the combination of ground parameter states, the ground class is determined. For example, in area 1, zone 1, the ground parameter states gneiss (lithology), poor (rock mass), and low (overburden) correspond to ground class GC5.

- A construction method is defined by one or more activities, which are connected into an activity network (Note: different from the tunnel network) (Figure 6-3). Each activity is characterized by cost, time and resource equations.
- The construction method definition describes the relation between ground classes (description of the geology component) and construction methods (construction simulation component): it defines which construction method is assigned to which ground class. A construction method can be assigned to more than one ground class, e.g. in Figure 6-2 construction method CM-C is assigned to ground classes GC1 and GC2.

The superposition of tunnel network and ground class profile produces the construction method profile (Figure 6-2). From this the total cost, the total time, and the resource amount for the construction of the tunnel network are calculated.

The DAT allow one to model uncertainties in the geology and in the construction process (Einstein 2004). In the description of the geology component, geologic profiles can be generated with five possible generation modes:

- Deterministic: The sequence and the lengths of the states of the ground parameters are input deterministically.

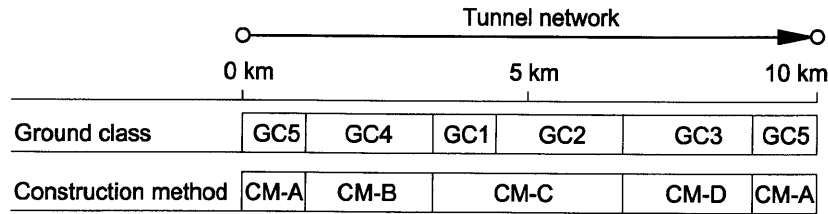


Figure 6-2: Construction simulation component in the DAT. The superposition of tunnel network and ground class profile produces the construction method profile. A construction method can be assigned to more than one ground class, for instance construction method CM-C is assigned to two ground classes, GC1 and GC2.

- Semi-deterministic: The sequence of the states of the ground parameters is input deterministically, while the lengths of the states are generated from a probability distribution.
- Markov: The sequence and the lengths of the states of the ground parameters are generated probabilistically. The sequence is generated with a Markov process based on transition probabilities, while the lengths are generated from an exponential distribution.
- Semi-fixed Markov: The sequence of the states of ground parameters are generated probabilistically with a Markov process based on transition probabilities. The lengths of the states of the ground parameters are generated from a triangular distribution.
- Fixed Markov: The sequence of the states of ground parameters are generated probabilistically with a Markov process based on transition probabilities. The lengths of the states of the ground parameters are equal to the mean length.

In the (fixed/semi-fixed) Markov generation modes, the length of ground parameter states and their transition probabilities are estimated. For instance, for the ground parameter rock mass, an estimation is made of the length of ground parameter states, such as poor, medium, good, and very good, as well as the probability that a rock mass quality is followed by the other rock mass qualities, e.g. the probability that poor rock mass is followed by medium, good, and very good rock mass. This information is used to simulate a possible profile for each ground parameter. The ground parameter profiles are then combined in the ground class profiles. A number of such ground class profiles (each being different) are simulated to represent the range of possible geologic conditions.

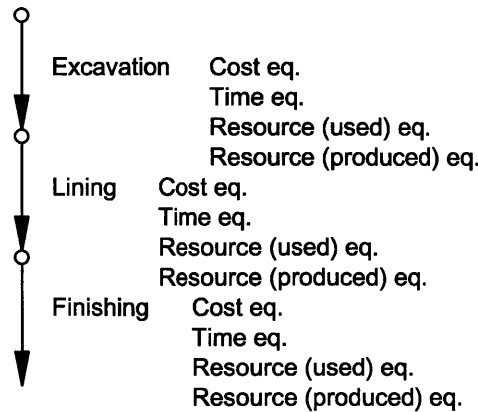


Figure 6-3: Activity network of a construction method. It consists of three activities, each characterized by cost, time and resource (used and produced) equations.

The result is a distribution of ground class profiles reflecting the uncertainty in the geology.

In the construction simulation component, the uncertainty of the variables used in cost and time equations (Figure 6-3), is modeled with probability distributions. For instance, the cost of excavating a unit length of tunnel can be modeled with a lognormal distribution, from which the construction cost of this tunnel unit length is randomly generated. A number of such construction costs per unit length (each being different) are generated to represent the range of possible construction costs per unit length. Similarly, construction time is generated. The result are distributions of total construction cost and time, which reflect the uncertainty in the construction process.

The construction simulation is based on the Monte Carlo procedure. First, one ground class profile from the distribution of ground class profiles is related to the corresponding construction methods. Then, the construction simulation proceeds through the construction method profile, producing a total cost and total time for the simulation, which is one point in the time-cost scattergram (Figure 6-4). This procedure is repeated by simulating other ground class profiles and the tunnel construction through them to produce the complete scattergram.

The description of the geology component, the construction simulation component and the uncertainty modeling in the DAT have been presented. Since the DAT were developed to simulate the construction of tunnels, their framework needs to be expanded in order to model other structures of rail lines (next section).

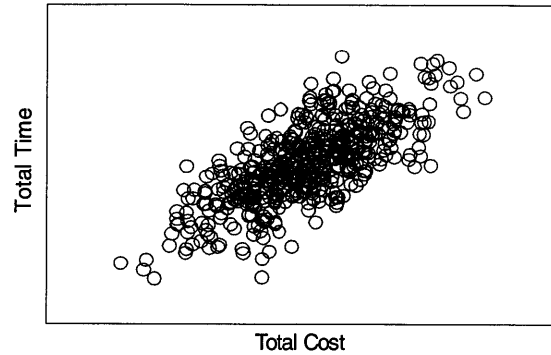


Figure 6-4: Time-cost scattergram. One point represents the total cost and total time for the construction of a tunnel. The simulation calculates this by proceeding through one construction method profile. The other points are generated by simulating other construction method profiles.

6.3 Extension of the DAT to Model Viaducts, Cuts and Embankments

The framework of the DAT has been extended to model the construction of rail line networks. This required changes in the concepts of four elements of the geology description component. In the following, the extended concepts of area, zone, ground parameter and ground class for viaducts, cuts and embankments are explained.

Since the construction of viaducts, cuts and embankments depends also on non-geologic conditions, the concepts of area, zone, ground parameter, and ground class are extended. The area becomes an alignment of the rail line, a zone corresponds to a structure: tunnel, viaduct, cut, or embankment (Figure 6-5). A ground parameter and a ground class are not exclusively geologic descriptors but become also structure descriptors. This is shown in an example with tunnels, viaducts, cuts, and embankments (Table 6.2 and Figure 6-5). In the case of tunnels, a ground parameter and its states describe the geology of a tunnel, as before. In the example in Table 6.2, the lithology and the faults of the tunnels are described. The lithology can be granite, sandstone, or shale, while there may or may not be faults. From the combination of the ground parameter states, the ground class is determined. There are four ground classes: GC1, GC2, GC3, and GC4. For tunnel 1 in Figure 6-5, the lithology is granite and there are no faults. The combination of the two states determines the ground class, which is GC1. In the case of tunnels, ground parameters and ground

classes thus describe the geology of the tunnel as before. Differently, in the case of viaducts, cuts and embankments, ground parameters and ground classes describe a specific characteristic of the structure.

Table 6.2: Example of ground parameters, ground parameter states, and ground classes of tunnels, viaducts, cuts and embankments. While for tunnels ground parameters, ground parameter states, and ground classes are geologic descriptors, for viaducts, cuts and embankments they are descriptors of a specific characteristic of the structure. In viaducts, they uniquely identify the particular viaduct; in cuts, they describe the percentage of blasting excavation; in embankments, they describe the need of improving prior to filling.

	Tunnel		viaduct	cut	embankment
ground parameter	lithology	faults	viaduct	blasting	improving
ground parameter state	granite	no	Moinhos	0%	no
	sandstone	yes	Aldeia	25%	yes
	shale		Ave	50%	
			Este	75%	
			Cambeses	100%	
ground class	GC1		Moinhos	GCB0	GCI
	GC2		Aldeia	GCB25	GCNI
	GC3		Ave	GCB50	
	GC4		Este	GCB75	
			Cambeses	GCB100	

In the case of viaducts, the ground parameter state and the ground class identify a viaduct uniquely. In the example in Table 6.2, the ground parameter is called "viaduct" and its states are the names of the viaducts in the alignment: Moinhos, Aldeia, Ave, Este, and Cambeses. Each ground parameter state corresponds to a ground class with the same name. In Figure 6-5, viaduct Moinhos has ground parameter state Moinhos, which determines ground class Moinhos. This ground parameter state and this ground class are uniquely applicable to viaduct Moinhos. Assigning a specific ground parameter state and a specific ground class to each viaduct is needed for the following reason. Each viaduct has a different construction method (as will be explained in section 6.4). In order to assign a different construction method to each viaduct, each viaduct must be assigned a unique ground parameter state and a unique ground class.

In the case of cuts, the state of the ground parameter describes the excavation means. In the example in Table 6.2, the ground parameter is called "blasting": its states describe the percentage

area	alignment 1					
zone	embank. 1	viaduct 1	embank. 2	cut 1	tunnel 1	cut 2
ground parameters						
lithology					granite	
faults					no	
viaduct	Moinhos					
blasting				25%		50%
improving	yes		no			
ground class	GCI	Moinhos	GCNI	GCB25	GC1	GCB50

Figure 6-5: Area, zones, ground parameters, and ground classes in the extended framework of the DAT. The area is alignment 1, and the zones are the structures, e.g. embankment 1, viaduct 1, cut 1, and tunnel 1. While for tunnels the ground classes are geology descriptors, for viaducts, cuts and embankments, ground classes describe a specific characteristic of the structure. For viaducts, the ground class identifies the viaduct: ground class "Moinhos" is the ground class for viaduct 1. For cuts, the ground class indicates the percentage of material excavated by blasting: cut 1 has ground class GCB25 (ground class blasting 25%) and cut 2 has ground class GCB50 (ground class blasting 50%). For embankments, the ground class indicates the need of improving the ground under the embankment: this is improved in embankment 1 (ground class CMI), while it is not improved in embankment 2 (ground class CMNI).

of material that is excavated with blasting, while the rest is excavated with mechanical means. The ground parameter state determines the ground class: ground parameter state 0% corresponds to ground class GCB0 (ground parameter blasting 0%), ground parameter state 25% corresponds to ground class GCB25, etc. In Figure 6-5, cut 1 is excavated by 25% blasting determining its ground class to be GCB25 (ground class blasting 25%). Similarly, cut 2 is excavated by 50% blasting determining its ground class to be GCB50 (ground class blasting 50%).

In the case of embankments, the ground parameter determines the need of improving the ground under the embankment prior to filling. In the example in Table 6.2, the ground parameter is called "improving". Its states describe whether there is or there is not the need of ground improvement. This corresponds to the ground classes GCI (ground class improving) and GCNI (ground class not improving), respectively. In Figure 6-5, embankment 1 has ground parameter state yes and ground

class GCI (ground class improving), while embankment 2 has ground parameter no and ground class GCNI (ground class not improving).

To summarize, in the case of tunnels, ground parameters and ground classes describe the geology of the tunnel, while for the other structures they describe some characteristics of the structure. In the case of viaducts, they describe and uniquely identify the viaduct. In the case of cuts, they describe the percentage of cut that is excavated with blasting. In the case of embankments, they describe whether the ground under the embankment needs to be improved (Table 6.3).

Table 6.3: Ground parameters, ground parameter states, and ground classes describe a specific characteristic of a structure. In tunnels, they describe the geology along a tunnel. In viaducts, they uniquely identify the particular viaduct. In cuts, they indicate the percentage of ground excavated with blasting. In embankments, they indicate whether the ground under an embankment needs to be improved prior to filling.

Structure	description
tunnel	geology
viaduct	viaduct identification
cut	percentage of blasting
embankment	need of improving

The concepts of area, zone, ground parameter and ground class in the DAT have been extended in order to model also viaducts, cuts and embankments. At this point, the construction model and the uncertainty model can be implemented in the DAT.

6.4 Implementation of the Construction Model in the DAT

In this section, the implementation of the construction model (chapter 3) in the DAT is presented. The construction model is implemented in the DAT with networks: construction networks and activity networks. First, these are explained and the difference clarified. Then, the construction network and the activity network of tunnels, viaducts, and of a sequence of cuts and embankments are presented and illustrated.

The construction network is the depiction of all structures in the construction project and the order in which the structures are built (Figure 6-6): structures can be built in sequence (structure 2 and structure 3) or in parallel (structure 1 and structure 2). In the construction network, each structure is represented by an arc that begins at e.g. point A (the structure's start point) and finishes at e.g. point B (the structure's end point). Alternatively, it can be represented by two

arcs: one beginning at point A and finishing at point B; the other beginning at point B and finishing at point A. In this case, the construction simulation stops at some point between point A and point B, when the two construction processes meet. For simplicity, the following explanations and examples are limited to the first case, where each structure is represented by one arc only.

The structures in the construction network are built with one or more methods. A construction method determines the activities to be performed. These are represented in the activity network (Figure 6-6). The activities of the activity network can be performed in sequence, e.g. activities a, b and c in the activity network of method I (Figure 6-6), or in parallel, e.g. activities a and d in the activity network of method III (Figure 6-6).

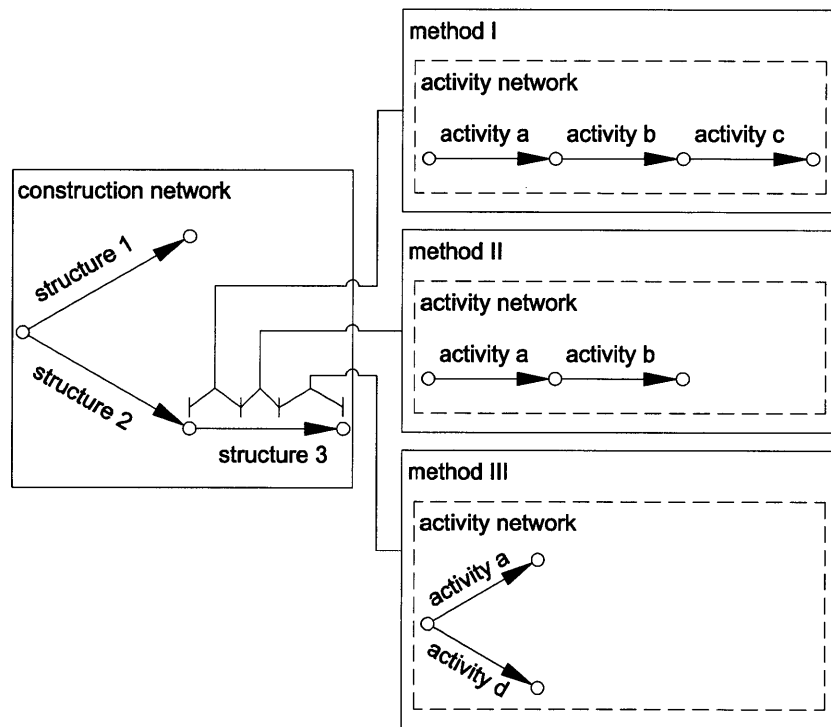


Figure 6-6: The construction network depicts the structures. These are built with one or more construction methods. The activity network of each construction method determines the order of the activities to be performed.

An activity network can be very simple or extremely complex. Depending on the level of detail, the construction of a structure can be modeled at the very general level (simplest activity network: one activity) or at the very detailed level (complex activity network: large number of

interconnected activities). For example, the construction of a tunnel can be modeled with an activity network comprising only one activity that models all the processes needed to construct one unit length of a tunnel. The activity network (in this case, the only activity) is repeated for every unit length of tunnel (Figure 6-7a). Conversely, the construction of a tunnel can be modeled with an activity network comprising several activities, one modeling each process needed to excavate one unit length of tunnel. For a tunnel excavated with blasting, the activity network can include five activities: a drilling activity, a loading activity, a detonating activity, a ventilating activity, and a mucking activity (Figure 6-7b). The activity network (in this case, including five activities) is repeated for every unit length of tunnel.

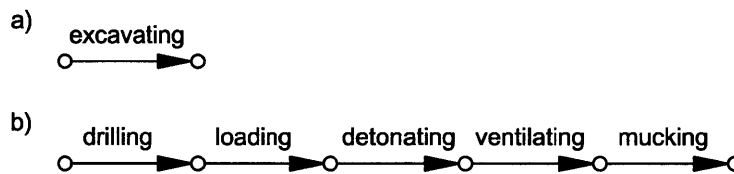


Figure 6-7: a) activity network modeling the construction of a unit length of tunnel at the general level, and b) activity network modeling the construction of a unit length of tunnel at the detailed level. The DAT can model the construction of a structure at different levels of detail.

In the research work presented in this thesis, tunnels, viaducts, cuts and embankments are modeled at different levels of detail. Tunnels are modeled with the simplest activity network (Table 6.4), one activity modeling all construction processes, while viaducts, cuts and embankments are modeled with more complex ones. The level of detail is determined by the data available: for tunnels, the construction cost and time are available per meter of tunnel; for viaducts, the construction cost and time of the elements (pier, deck section, etc.) are available; for cuts and embankments, the construction cost and time to clear a unit area, to excavate a unit volume or to fill a unit volume are available.

The networks to model the construction of viaducts and to model the construction of a sequence of cuts and embankments have been developed in Chapter 3 based on the construction processes to build viaducts, and cuts and embankments, respectively. Two types of networks modeling the construction of viaducts are detailed in section 3.2.2 and depicted in Figures 3-6 and 3-7. The network modeling the construction of a sequence of cuts and embankments is detailed in section

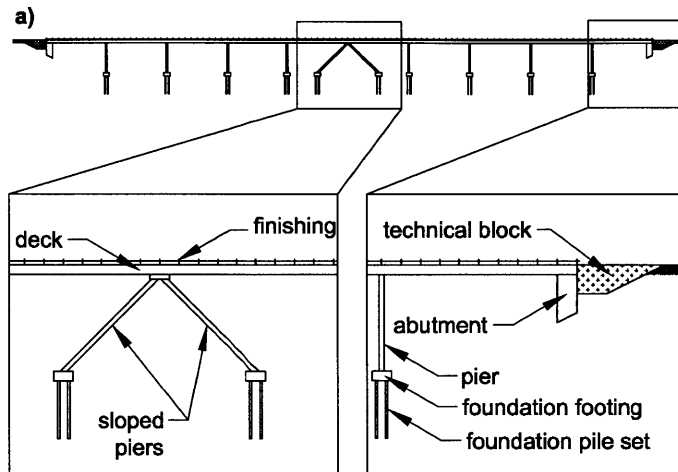
3.2.3 and depicted in Figure 3-11. These networks have been implemented in the DAT as described in the following paragraphs.

Among the structures modeled, viaducts have the most complex activity network. Figure 6-8 shows a) a 10-span viaduct and b) the activity network to model the construction of the viaduct. The activity network includes one activity for the construction of the following elements: abutments (a), technical blocks (tb), deck sections (ds), (sloped) piers (p, sp), foundation footings (ff), foundation pile sets (fp), and the finishing (fi). Additionally, three activities model the construction site setup time (pw), the deck construction setup time (pd), and the cost of the construction method (cm) to build the deck, e.g. the cost of a gantry. The activities can be either parallel or sequential. For instance, the deck sections are constructed sequentially, that is the activity to construct the first deck section is followed in sequence by the activity to construct the second deck section. At the same time, the second deck section is constructed in parallel with e.g. the third foundation pile set, or the third pier. For viaducts with a large number of spans, the activity network becomes proportionally longer. The longest viaduct modeled in the DAT so far has more than 100 spans.

Note that the activity network of a viaduct, thus the construction method, is structure-specific: since the number of spans changes from viaduct to viaduct, each viaduct needs a customized activity network, thus a customized construction method.

Cuts and embankments are modeled jointly as a sequence of the two types of structures since they share common processes, e.g. the area of a cut is cleared before excavating similarly to the area of an embankments, which is cleared before filling (section 3.2.3). Differently from tunnels and viaducts, the sequence of cuts and embankments cannot be modeled in an activity network in the DAT. If cuts and embankments were modeled separately (rather than jointly), two separate methods, each with a different activity network, would model the construction of cuts and embankments, respectively. The activity network would be repeated at every construction cycle along cuts and embankments, respectively. However, cuts and embankments are modeled jointly; thus, it is impossible to have a method with an activity network to model the construction of both cuts and embankments, since the construction process of cuts is different from the construction process of embankments. As an alternative, in the DAT a sequence of cuts and embankments is modeled in the construction network rather than in an activity network (Table 6.6).

Modeling the sequence of cuts and embankments with a construction network rather than with



- b)
- | | |
|-------------------------|-------------------------|
| pw: pre-work | fp: foundation pile set |
| a: abutment | ff: foundation footing |
| cm: construction method | p: pier |
| pd: pre-deck | sp: sloped pier |
| ds: deck section | tb: technical block |
| d: dummy activity | fi: finishing |

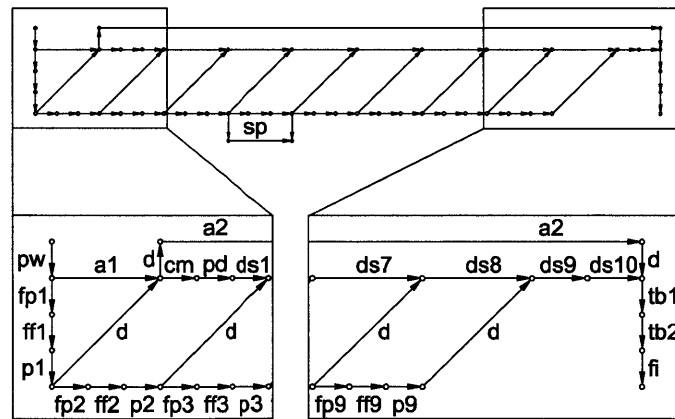
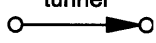



Figure 6-8: a) a viaduct consists of several elements, whose construction is modeled with b) the activity network. The more spans the viaduct has, the longer the activity network becomes.

an activity network requires some changes in the DAT input. In the construction network, a cut is modeled with four arcs (Table 6.6): a first arc to model the clearing of the cut (Cc), a second to model the excavation (E), a third to model the capping (Ac), and a fourth to model the sub-ballast (Sc). Similarly, the construction of an embankment is modeled with five arcs (Table 6.6): a first arc to model the clearing of the embankment (Ce), a second to model the improving (I), a third to model the filling (F), a fourth to model the capping (Ae), and a fifth to model the sub-ballast (Se). Since each arc is assigned a construction method, a cut is modeled with four construction methods (clearing, excavating, capping, and sub-ballast), and an embankment is modeled with five construction methods (clearing, improving, filling, capping, sub-ballast) (Table 6.6). These construction methods have simple activity networks (Table 6.6): for instance, the activity network of the method clearing consists of one activity: clearing.

To summarize, in the research work presented in this thesis tunnels, viaducts, cuts and embankments are modeled at different levels of detail. A tunnel has the simplest construction network (one arc) and the simplest activity network (one activity) (Table 6.4). A viaduct has a simple construction network (one arc) but a complex activity network that, depending on the viaduct length, can have hundreds of interconnected activities (Table 6.5). A sequence of cuts and embankments has a complex construction network, whose size increases with the number of cuts and embankments in the sequence. On the other hand, the construction methods modeling it have the simplest activity network (one activity) (Table 6.6).

Table 6.4: Construction network and activity network of a tunnel. Tunnels are modeled with the simplest construction network (one arc) and the simplest activity network (one activity).

Construction network	tunnel 
method	excavation
activity network	excavating 

As shown above, the construction model has been implemented in the DAT with construction networks and activity networks of tunnels, viaducts, cuts and embankments. In the following, the implementation of the uncertainty model is discussed.

Table 6.5: Construction network and activity network of a viaduct. Viaducts are modeled with the simplest construction network (one arc) but a complex activity network.

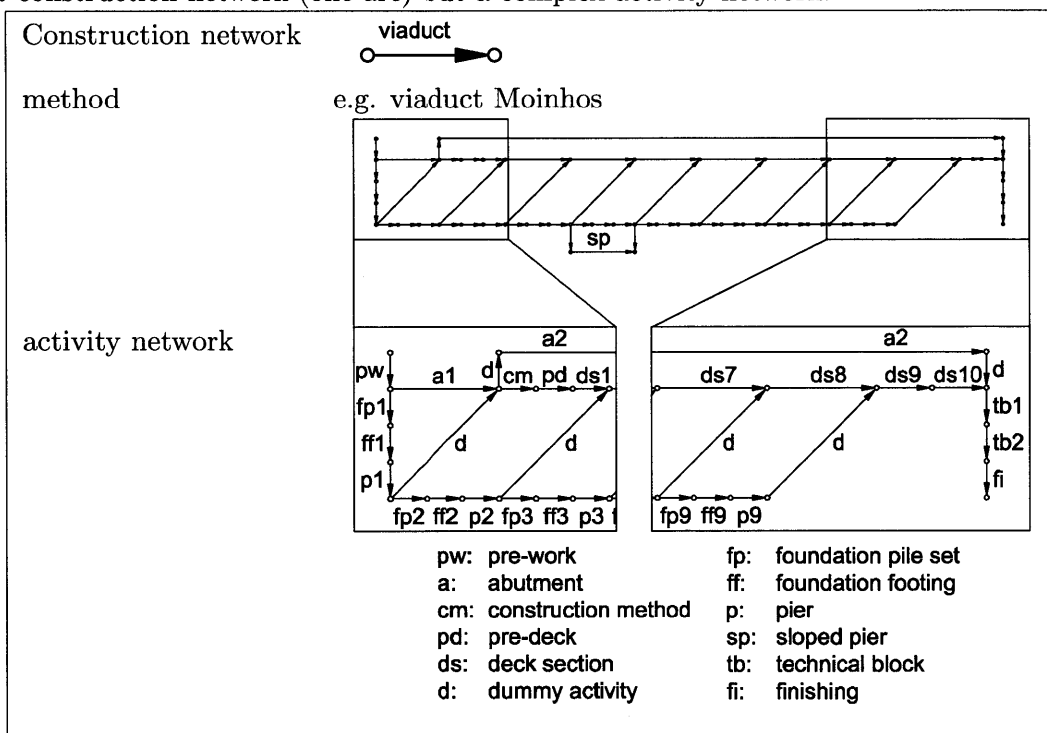
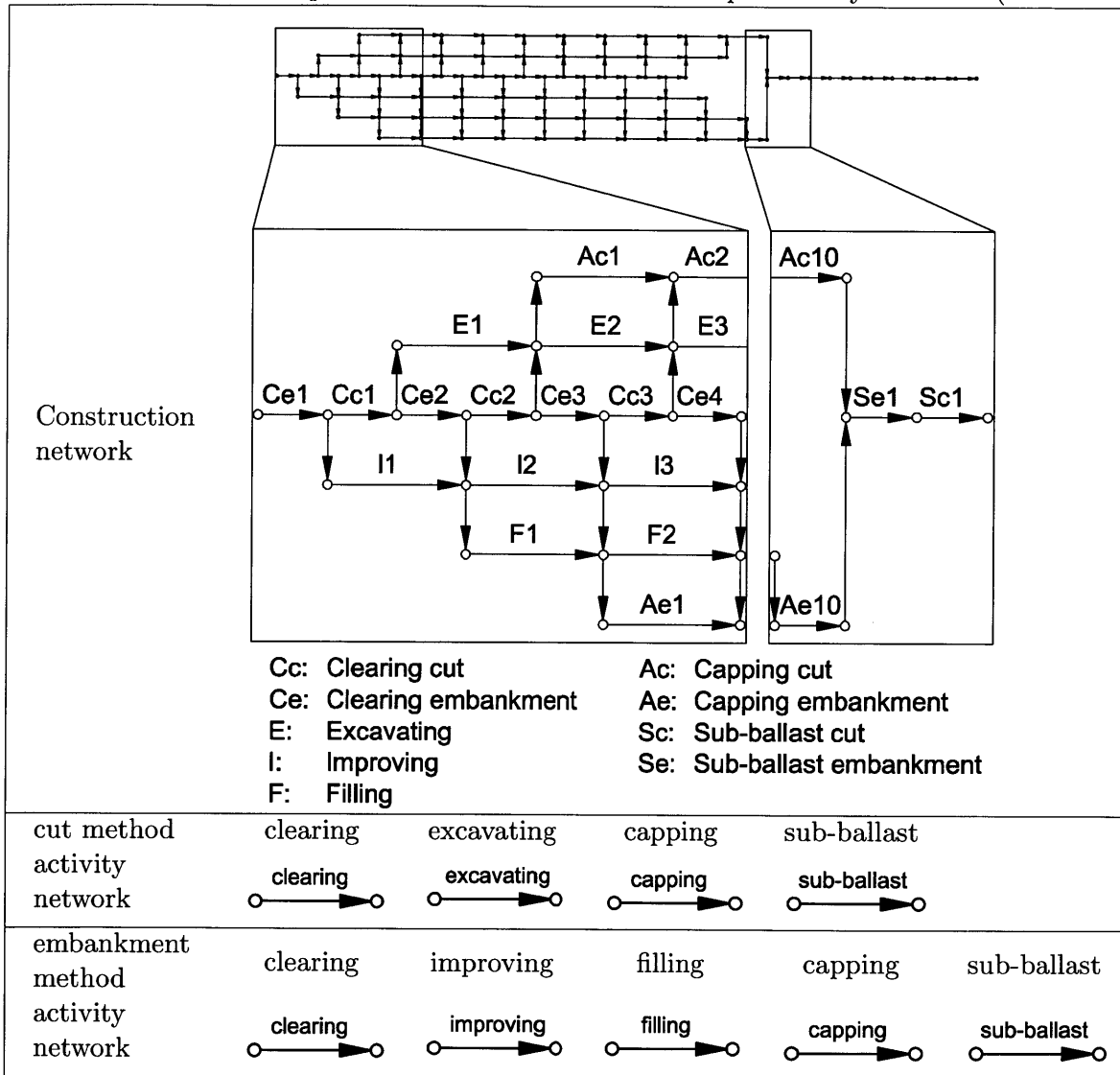


Table 6.6: Construction network and activity network of a sequence of cuts and embankments. This is modeled with a complex construction network but simple activity networks (one activity).



6.5 Implementation of the Uncertainty Model in the DAT

In this section, the implementation of the uncertainty model (chapters 4 and 5) in the DAT is presented. In the uncertainty model, the uncertainty is represented with three sources of uncertainty: variability, cost correlations and disruptive events. First, the probability distributions to model the variability and the impacts of the disruptive events are presented. Then, the correlation models in the DAT are explained. Last, it is shown how disruptive events and their probability of occurrence are modeled.

6.5.1 Probability Distributions

The distributions readily available in the DAT (uniform, triangular, triangular with spikes) and the newly introduced lognormal distribution are presented here. Particular attention will be given to the sample generation from a lognormal distribution in the DAT.

For the research work presented in this thesis, four inputs are required in form of probability distributions: the cost and time of activities, and the cost and time impacts of disruptive events (Table 6.7). While the cost of activities is modeled with the lognormal distribution, the other inputs are represented with the triangular distribution (section 4.2).

Table 6.7: Inputs required in form of probability distributions for the research presented in this thesis. The cost of activities is modeled with lognormal distributions, while the time of activities as well as the cost and time impacts of disruptive events are modeled with triangular distributions.

	Cost	time
activity	lognormal	triangular
disruptive event	triangular	triangular

The DAT can model the uniform distribution, the triangular distribution and the triangular distribution with spikes (Figure 6-9). These are used to model the thin tails of a distribution, by concentrating the tails into residual probabilities at the minimum and the maximum values of the distribution.

The lognormal distribution has been added to the set of probability distributions modeled in the DAT. A lognormal distribution (with minimum different from zero) is defined given three points. In the DAT input, these are the minimum, the mode and a known percentile of the distribution (Figure 6-10). From these three points, the DAT calculate the mean μ and the standard deviation

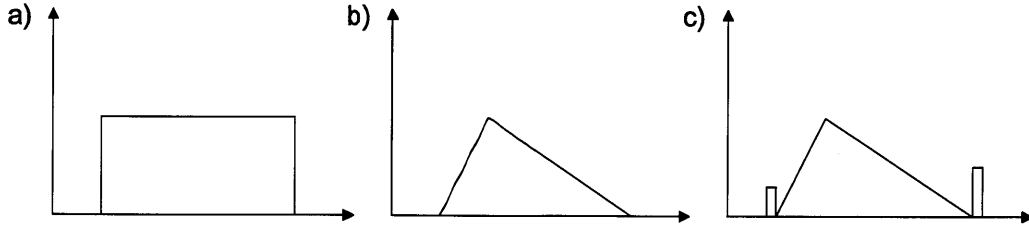


Figure 6-9: a) the uniform distribution, b) the triangular distribution, and c) the triangular distribution with spikes can be modeled in the DAT.

σ of the corresponding normal distribution, according to the following expressions

$$\begin{cases} mode - min = \exp(\mu - \sigma^2) \\ percent = 0.5 + 0.5 \operatorname{erf}\left(\frac{\ln(percent - min) - \mu}{\sigma\sqrt{2}}\right) \end{cases}, \quad (6.1)$$

where min , $mode$ are the minimum and the mode of the lognormal distribution. $percentile$ is a percentile of the lognormal distribution. It is not exceeded with a probability equal to $percent$, and it is exceeded with a probability equal to $(100\% - percent)$. For example, the 90th percentile is not exceeded with probability equal to 90%, and it is exceeded with probability $(100\% - 90\%) = 10\%$.

The DAT generate samples from a lognormal distribution in three steps. First, they generate samples from a normal distribution

$$X \sim normal(0, 1). \quad (6.2)$$

Second, they apply the transformation

$$Y = \exp(\sigma X + \mu) \quad (6.3)$$

to obtain samples from the lognormal distribution

$$Y \sim lognormal(\mu, \sigma). \quad (6.4)$$

Then, they apply the transformation

$$C = min + Y \quad (6.5)$$

to obtain samples from the lognormal distribution

$$C \sim (\text{min} + \text{lognormal}(\mu, \sigma)). \tag{6.6}$$

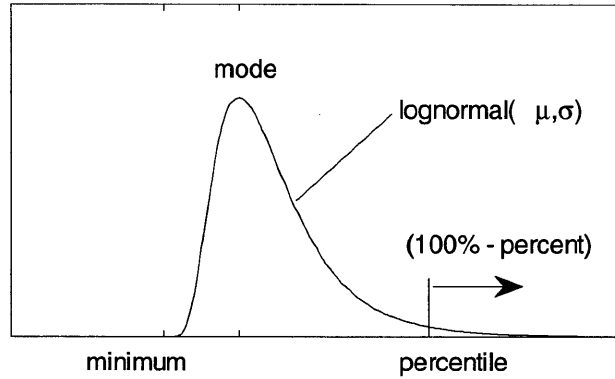


Figure 6-10: The DAT generate samples from a lognormal distribution (μ, σ) . μ and σ are the parameters of the corresponding normal distribution. Given the minimum, the mode and a percentile of the distribution, they are determined with expression 6.1.

Above, the distributions available in the DAT and the newly introduced lognormal distribution are presented. These distributions ensure the modeling of the variability and the impacts of the disruptive events. In the following, the modeling of correlations in the DAT is discussed.

6.5.2 Cost Correlations

In this section, the correlation that can be modeled in the DAT (correlation between different variables) and the newly introduced correlation (correlation between the costs of a repeated activity) are presented. Examples of this latter type of correlation in rail line construction are given and the input in the DAT is described.

In the DAT, correlations between different variables can be modeled, e.g. the correlation between the cost and the time of an activity. Two variables can be modeled as positively correlated, uncorrelated, or negatively correlated. For positively correlated variables, the DAT generate two values at the same percentile of the respective distributions. For uncorrelated variables, the DAT generate the values from their respective distributions separately. For negatively correlated variables, the DAT generate a value at a percentile Z from the distribution of the first variable and a

value at the percentile $(1 - Z)$ from the distribution of the second variable.

In order to implement the uncertainty model, the correlation between the costs of a repeated activity (correlation type 1 in chapter 5) is introduced in the DAT. Examples of correlations between the costs of a repeated activity are given for the four structures modeled in this thesis:

- tunnel: the cost of excavating one unit length of tunnel is correlated with the costs of excavating preceding and following unit lengths of tunnel.
- viaduct: the cost of constructing one deck section is correlated with the costs of constructing preceding and following deck sections.
- cut: the cost of blasting one unit volume of cut is correlated with the costs of blasting preceding and following unit volume in the cut excavation process.
- embankment: the cost of filling one unit volume of embankment is correlated with the costs of filling preceding and following unit volume in the embankment construction process.

The correlation between the costs of a repeated activity are modeled in the DAT with the NORTA method (section 5.3). This generates samples with the desired correlation matrix and the desired marginal distributions. The size of the correlation matrix is equal to the number of costs correlated and it is proportional to the size of the structure. For example, for a $1,000m$ long tunnel, there are 1,000 correlated costs (the unit length of the excavation is equal to $1m$). Thus, the matrix size is $1,000 \times 1,000$.

Due to the large size of the correlation matrix, the input of the correlation coefficients in the DAT has been simplified. Although correlation coefficients in the correlation matrix could be all different, according to the expert estimates (section 7.2.1) they have a specific pattern described here (expression 6.7):

- costs are correlated with correlation coefficient $\rho = 0.9$,
- with the exception of the first 50 costs, which are correlated with correlation coefficient $\rho = 0.5$.

This pattern is translated into the following correlation matrix:

(6.7)

where the outer shell of the matrix (50 top and bottom rows, and 50 left and right columns) has correlation coefficient equal to $\rho = 0.5$, while the core of the matrix has correlation coefficient equal to $\rho = 0.9$.

In order to accommodate the specific pattern shown in the numerical example above, the DAT take three correlation inputs (expression 6.8):

- Correlation coefficient 1 (ρ_1): the correlation coefficient in the outer shell of the correlation matrix.
- Number of rows and columns, N , with correlation coefficient 1 (ρ_1): the number of variables that are correlated with correlation coefficient 1 (ρ_1).
- Correlation coefficient 2 (ρ_2): the correlation coefficient in the core of the correlation matrix.

$$R = \begin{matrix} & \begin{matrix} 1 & \rho_1 & & & \rho_1 \end{matrix} \\ \begin{matrix} \rho_1 \\ \rho_1 \\ \rho_1 \\ \rho_1 \\ \rho_1 \end{matrix} & \begin{matrix} \rho_1 & \rho_1 & \rho_2 & \rho_2 & \rho_1 \\ \rho_1 & \rho_2 & \rho_2 & \rho_2 & \rho_1 \\ \rho_2 & \rho_2 & \rho_2 & \rho_2 & \rho_1 \\ \rho_2 & \rho_2 & \rho_2 & \rho_2 & \rho_1 \\ \rho_1 & \rho_1 & \rho_1 & \rho_1 & \rho_1 \end{matrix} \\ & \begin{matrix} \rho_1 & \rho_1 & \rho_1 & \rho_1 & 1 \end{matrix} \end{matrix}, \quad (6.8)$$

Although the input of the correlation matrix is simplified, the DAT can model any correlation matrix. However, for types of matrices different from the one in the example, the input in the DAT must be adapted.

In the preceding, the modeling of correlations in the DAT has been discussed, with special attention on the correlation between the costs of a repeated activity. The last aspect of the uncertainty model in the DAT that remains to be modelled is the occurrence of disruptive events (next section).

6.5.3 Disruptive Events

Disruptive events occur with a probability of occurrence and cause additional cost and time. In this section, the Markov processes used in the DAT to model the probability of occurrence are presented. It will be shown that depending on the type of occurrence, a disruptive event is modeled in different ways in the DAT.

Disruptive events are events that have a large impact on the construction process and occur with some probability of occurrence (usually small). Examples of disruptive events are a cave-in in a tunnel, differing soil conditions requiring the re-design of viaduct foundations, or the flooding of the earthwork area.

In the DAT, the probability of occurrence of disruptive events is modeled with a Markov process. The states of the Markov process are "no" (the disruptive event does not occur) and "yes" (the disruptive event occurs) (Figure 6-11 and Table 6.8). The transition probability from state "no" to state "yes" is equal to the probability of occurrence of the disruptive event, p , while the transition probability from state "yes" to state "no" is equal to one minus the probability of occurrence of the disruptive event, $(1 - p)$. Returning to the same state is also possible: given the current state is "yes", the probability of returning to state "yes" is equal to the probability of occurrence of the disruptive event, p ; given the current state is "no", the probability of returning to state "no" is equal to one minus the probability of occurrence of the disruptive event, $(1 - p)$.

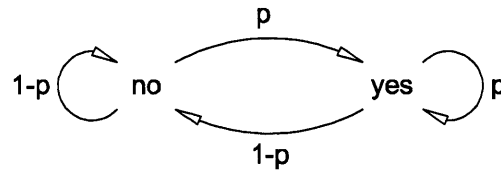


Figure 6-11: A Markov process models the probability of occurrence of a disruptive event in the construction of a structure. The state "no" indicates that the disruptive event does not occur, while the state "yes" indicates that the disruptive event occurs. The probability of transitioning to the "yes" state is equal to the probability of occurrence of the disruptive event, p . The probability of transitioning to the "no" state is equal to $(1 - p)$.

Table 6.8: The transition matrix of the Markov process determines the occurrence (state "yes") or no occurrence (state "no") of the disruptive event. The probability of transitioning to the state "yes" is equal to p , while the probability of transitioning to the state "no" is equal to $(1 - p)$.

		states	
		yes	no
states	yes	p	$(1 - p)$
	no	p	$(1 - p)$

Since the DAT cannot model a transition to the same state, the Markov process is modified into an equivalent Markov process that includes four states: "no1", "no2", "yes1", and "yes2" (Figure 6-12 and Table 6.9). The transition from the "no" state to itself is replaced with the transition from the "no1" state to the "no2" state, or vice versa. Similarly, the transition from the "yes" state to itself is replaced with the transition from the "yes1" state to the "yes2" state, or vice versa. In the modified Markov process, the probability of transitioning to "yes1" or "yes2" states is equal

to the probability of occurrence of the disruptive event, p , while the probability of transitioning to "no1" or "no2" states is equal to $(1 - p)$.

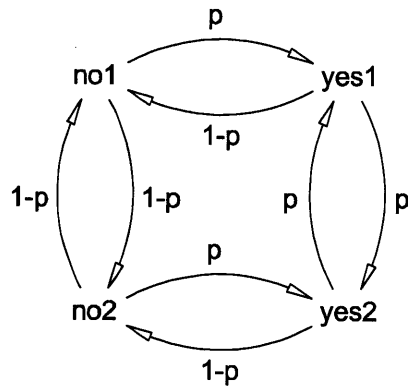


Figure 6-12: In the modified Markov process, the transition to the same state is replaced with the transition to an equivalent state. For instance, the transition from the "no" state to itself is replaced with the transition from the "no1" state to the "no2" state, or vice versa. Similarly, the transition from the "yes" state to itself is replaced with the transition from the "yes1" state to the "yes2" state, or vice versa.

Table 6.9: The transition matrix of the Markov process determines the occurrence (states "yes1" or "yes2") or no occurrence (states "no1" or "no2") of the disruptive event. The probability of transitioning to the states "yes1" and "yes2" is equal to p , while the probability of transitioning to the states "no1" and "no2" is equal to $(1 - p)$.

		states			
		yes1	yes2	no1	no2
states	yes1	0	p	$(1 - p)$	0
	yes2	p	0	0	$(1 - p)$
	no1	p	0	0	$(1 - p)$
	no2	0	p	$(1 - p)$	0

In the DAT, the impacts of a disruptive event are modeled with triangular probability distributions (section 6.5.1). If the Markov Process transitions to a state that means the disruptive event occurs ("yes1" or "yes2"), cost and time impacts are generated from the triangular distributions and summed to the total cost and the total time, respectively.

In the research work presented in this thesis, two disruptive events per type of structure are modeled (Table 6.10). In tunnels, the disruptive events are a sudden water inflow and a cave-in. In viaducts, they are differing site conditions that require a one-time re-design of the foundations

and a construction problem or accident with significant cost and delay. In earthwork (cuts and embankments), they are the flooding of the earthwork area and differing site conditions.

Table 6.10: Disruptive events modeled in the construction of tunnels, viaducts, cuts and embankments. Their probability of occurrence is modeled with the modified Markov process.

Structure	disruptive events	
tunnel	water inflow	cave-in
viaduct	differing site conditions	construction accident
cut & embankment	flooding	differing site conditions

The disruptive events can be grouped according to the type of occurrence (Table 6.11):

1. the disruptive event is modeled in every construction cycle: water inflow and cave-in in the tunnel may occur in every construction cycle.
2. the disruptive event is modeled once for the entire construction project: the problem or accident in viaduct construction, and the differing site conditions in viaduct construction and in earthwork construction may occur once per structure.
3. the disruptive event is modeled every day of construction: the flooding of the earthwork area may happen every day.

The three types of occurrences can be modeled for tunnels, viaducts, cuts and embankments. However, in this specific case the disruptive event in every construction cycle occurs in tunnels, the one-time disruptive event occurs in viaducts and in the sequence of cuts and embankments, and the daily disruptive event occurs in the sequence of cuts and embankments.

Table 6.11: The disruptive events water inflow and cave-in may occur in every construction cycle of a tunnel. The disruptive events construction accidents (in viaducts) and differing site conditions (in viaducts, and in cuts and embankments) may occur once for the construction of a structure. The disruptive event flooding (in cuts and embankments) may occur in each construction day of the cuts and embankments. Depending on the type of occurrence, the disruptive event is modeled differently in the DAT.

Occurrence	disruptive events	
in every construction cycle	water inflow	cave-in
once	construction accident	differing site conditions
daily for N days	flooding	

The occurrence of a disruptive event determines the way the event is modeled in the DAT:

1. at every construction cycle: at every construction cycle, the disruptive event may or may not occur, and if it occurs, it causes additional cost and time. The Markov process determines the state of each construction cycle. If the state of a construction cycle is "yes1" or "yes2", the disruptive event occurs. If the state is "no1" or "no2", the disruptive event does not occur.
2. once: the occurrence of the disruptive event is modeled with an ad-hoc solution. The ad-hoc solution consists in modeling the disruptive event with a fictitious structure (Table 6.12). The disruptive event may or may not occur, and if it occurs, it causes additional cost and time. Similarly, the fictitious structure may or may not be built, and if it is built, it requires additional cost and time. These are equal to the additional cost and time of the disruptive event. Since the disruptive event may occur only once, the fictitious structure is one-unit long. In this ad-hoc solution, the Markov process determines the state of the one-unit fictitious structure. If the state is "yes1" or "yes2", the disruptive event occurs and the fictitious structure is built. If the state is "no1" or "no2", the disruptive event does not occur and the fictitious structure is not built.

Table 6.12: A disruptive event that may occur once is modeled with a fictitious structure that is one-unit long. If the disruptive event occurs, the fictitious structure is built causing additional cost and time equal to the cost and time of the disruptive event.

Disruptive event	fictitious structure
occurrence: once	length: one unit
the disruptive event may occur with probability p	the fictitious structure may be built with probability p
if the disruptive event occurs, it causes additional cost C and time T	if the fictitious structure is built, it requires cost C and time T

3. daily for N days: the occurrence of the disruptive event is also modeled with an ad-hoc solution. The ad-hoc solution consists in modeling the disruptive event with a fictitious structure (Table 6.13). Any day of the N days, the disruptive event may or may not occur, and if it occurs, it causes additional cost and time. Similarly, any unit of the N -unit long fictitious structure may or may not be built, and if it is built, it require additional cost and time. These are equal to the additional cost and time of the disruptive event. In this ad-hoc solution, the Markov process determines the state of the units of the fictitious structure. If

the state of a unit is "yes1" or "yes2", the disruptive event occurs on the corresponding day and the unit of the fictitious structure is built. If the state is "no1" or "no2", the disruptive event does not occur in the corresponding day and the unit of the fictitious structure is not built.

Table 6.13: A disruptive event that may occur daily for N days is modeled with a fictitious structure that is N -unit long. If the disruptive event occurs on a day, the corresponding unit of the fictitious structure is built causing additional cost and time equal to the cost and time of the disruptive event.

Disruptive event	fictitious structure
occurrence: daily for N days	length: N units
the disruptive event may occur with probability p	a unit of the fictitious structure may be built with probability p
if the disruptive event occurs, it causes additional cost C and time T	if a unit of the fictitious structure is built, it requires cost C and time T

It has been shown how in the DAT the occurrence of disruptive events are modeled with Markov processes, and how disruptive events are represented depending on their type of occurrence. To summarize, the three sources of uncertainty (variability, cost correlations, and disruptive events) are implemented in the DAT with triangular and lognormal distributions, the NORTA model and Markov processes.

6.6 Conclusions

In the DAT, the construction of tunnels, viaducts, cuts and embankments is modeled with construction networks and activity networks. The three sources of uncertainty are represented in the DAT as follows: variability in cost and time is modeled with probability distributions, the correlation between the costs of a repeated activity is modeled with NORTA, and disruptive events are modelled with Markov processes (probability of occurrence) and probability distributions (cost and time impacts).

The construction model (chapter 3) and the uncertainty model (chapters 4 and 5) are successfully implemented in the simulation tool DAT. They are applied to the construction project of the new Portuguese high speed rail line in the next chapter (chapter 7).

Chapter 7

Application of the Construction and Uncertainty Models to the New Portuguese High Speed Rail Line

The construction model and the uncertainty model developed in this research can now be applied to a rail project using the simulation tool DAT (Decision Aids for Tunneling). The impact of the different sources of uncertainty on the construction cost and time need to be evaluated. The author collected data on the cost and time variability of construction activities, the correlations between costs, and the disruptive events to model the construction of four alignments of the new Portuguese high speed rail line. This chapter presents this application and demonstrates the effectiveness of the construction and uncertainty models. The deterministic total cost and total time by definition cannot reflect the range of possible outcomes. Including the variability of the activity cost and time procures uncertainty in total cost and total time, including cost correlations cause the total cost standard deviation to increase dramatically, and including disruptive events further increases the total cost and total time. If the variability, the cost correlations, and the disruptive events are disregarded, the ranges of the total cost and total time are underestimated.

The detailed structure of the chapter is as follows: first, the new Portuguese high speed rail line project is described (section 7.1). Second, the input data to calculate the deterministic total cost and total time and to simulate the distributions of the total cost and total time are presented (section 7.2). Then, the construction model and the uncertainty model are applied to the project

(sections 7.3 and 7.4). Finally, parallel and sequential construction of the structures of the rail line are compared from the perspective of the construction time (section 7.5).

7.1 Description of the Project

In this section, the project of the new Portuguese high speed rail line is introduced.

The new Portuguese high speed rail line consists of three main axes: Lisbon-Madrid, Lisbon-Porto, and Porto-Vigo (Figure 7-1). In this chapter, the construction of the Porto-Vigo alignment is modeled. Four possible alignments are considered:

- Alignment A and Alignment C (B - BA - A) between Porto and Braga (Figure 7-2).
- Alignment A and Alignment B between Porto and Vigo (Spanish border) (Figure 7-3)

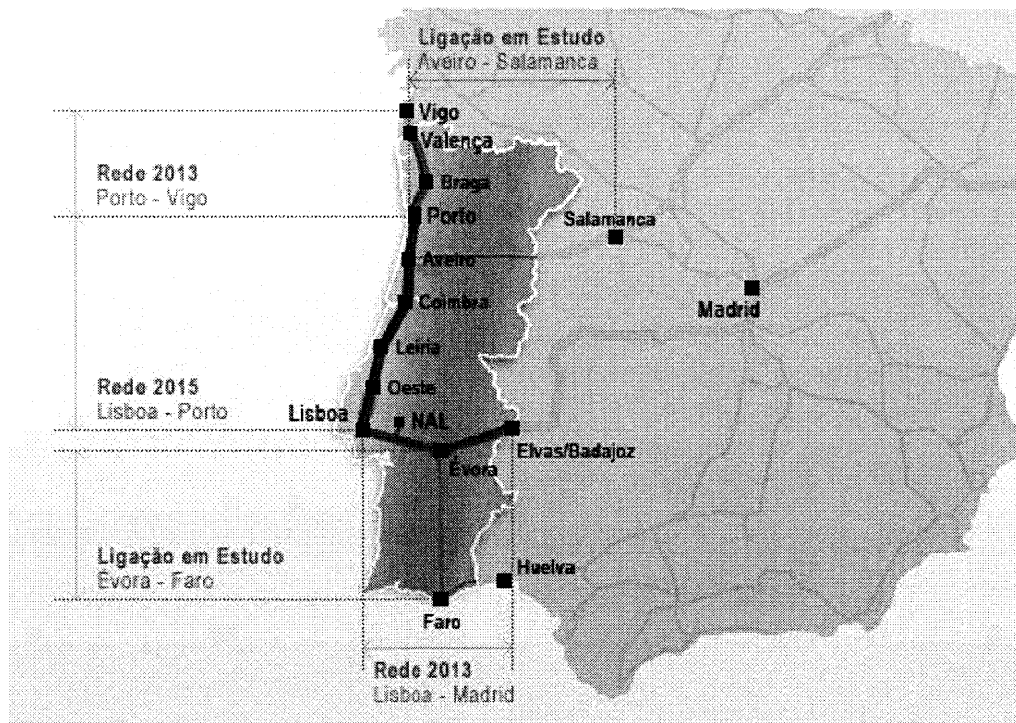


Figure 7-1: New Portuguese high speed rail. The three main axes (blue) are Lisbon-Madrid (Spanish border), Lisbon-Porto, and Porto-Vigo. Four alignments of the Porto-Vigo (Spanish border) axis are investigated in this chapter.

For reasons of clarity, the four alignments are assigned unique names that will be used throughout the chapter (Table 7.1).

Table 7.1: The four alignments of the Porto-Vigo axis are assigned unique names, used throughout the chapter.

Alignment name	description
A-a	Alignment A between Porto and Braga (Figure 7-2)
A-c	Alignment C between Porto and Braga (Figure 7-2)
B-a	Alignment A between Braga and Vigo/Spanish border (Figure 7-3)
B-b	Alignment B between Braga and Vigo/Spanish border (Figure 7-3)

To calculate the construction cost and time of the alignments presented above, cost and time input data are provided in the next section.

7.2 Input Data

In this section, the input data 1) to calculate the deterministic total cost and total time and 2) to model the uncertainty in total cost and total time in of the four main structures of rail lines (tunnels, viaducts, cuts and embankments) are provided. Specifically, the following data are presented:

- deterministic input costs and times to calculate the deterministic total cost and total time;
- cost and time distributions to model the variability,
- correlations between the costs of activities,
- probability of occurrence and cost and time impacts of disruptive events.

The sources of the input data vary depending on the structure and on the input data: they are historical sources and estimations of experts. For reference, they are summarized in Table 7.2.

In the following, the input data to model the construction cost and time of tunnels (section 7.2.1), of viaducts (section 7.2.2) and of cuts and embankments (section 7.2.3) are presented.

7.2.1 Tunnel Data

In this section, the input data to model the construction cost and time of tunnels is presented in the following order:

Table 7.2: The input data consists of historical data and expert estimations.

Input data		structure		
		tunnel	viaduct	cut & embankment
cost	deterministic input	(RAVE 2006d) (RAVE 2008d)	design engineer	(RAVE 2008b)
	prob. distr.	(RAVE 2006d) (RAVE 2008d)	(tHR 2007)	(tHR 2007)
time	deterministic input	(RAVE 2006d) (RAVE 2008d)	(RAVE 2006c) (RAVE 2008c)	earthwork expert
	prob. distr.	(RAVE 2006d) (RAVE 2008d)	viaduct expert	earthwork expert
cost correlations		tunnel expert	viaduct expert	earthwork/tunnel expert
disruptive events		tunnel expert	viaduct expert	earthwork expert

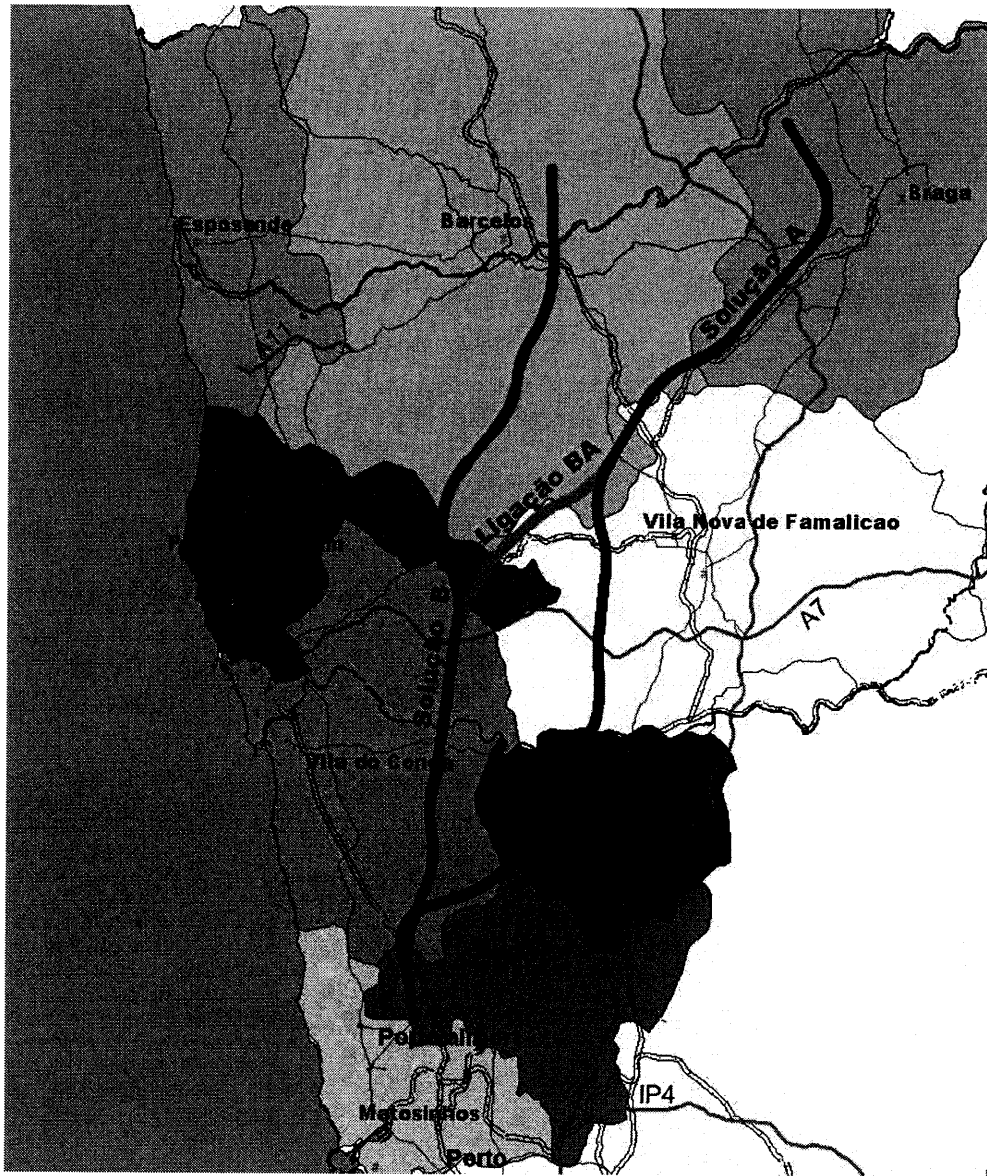


Figure 7-2: New Portuguese high speed rail line, Porto-Vigo axis, Porto-Braga alignments. Alignment A (blue) and alignment C (B (red) - BA (green) - A (blue)) are two of the four alignments studied in this chapter.

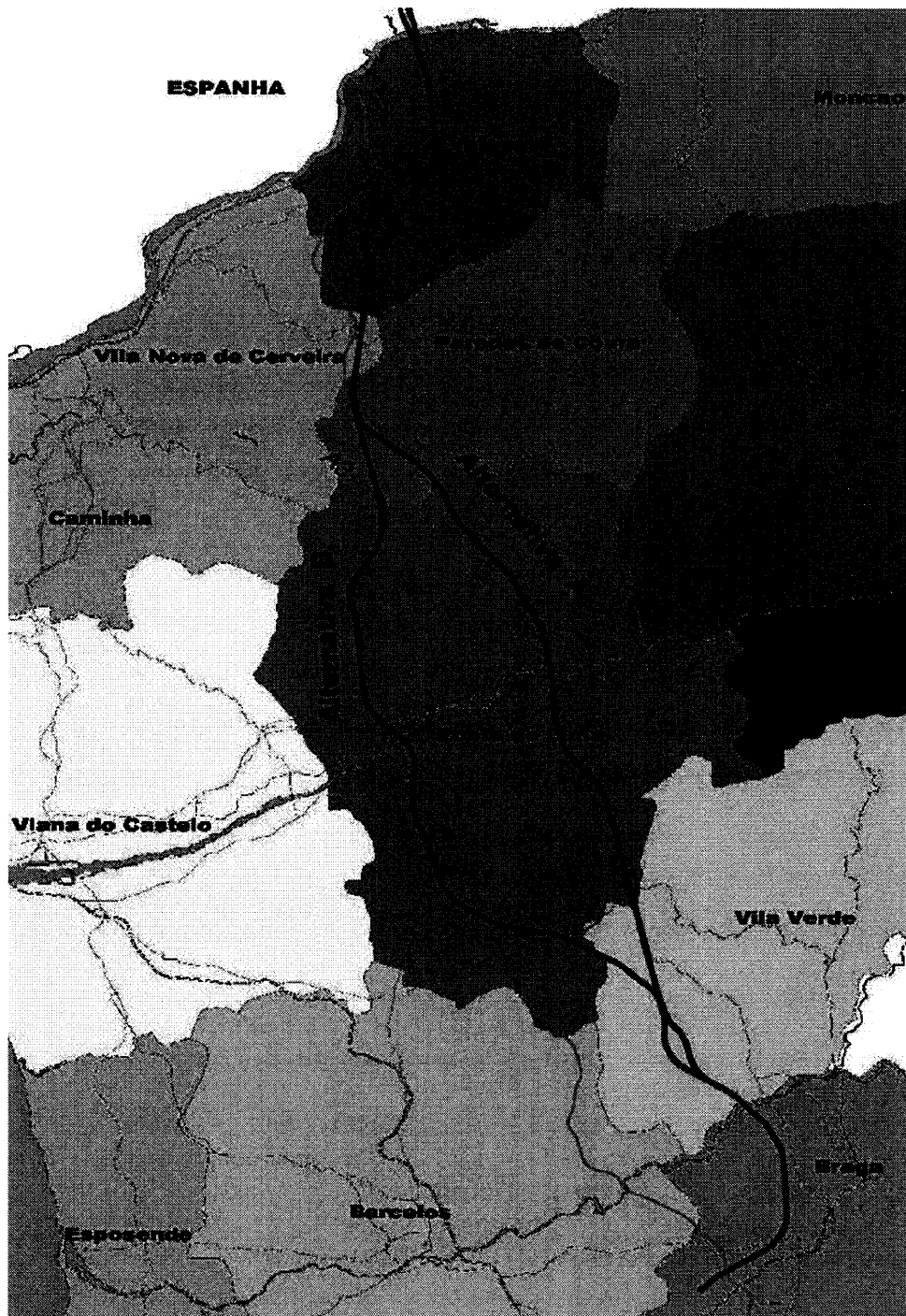


Figure 7-3: New Portuguese high speed rail line, Porto-Vigo axis, Braga-Vigo (Spanish border) alignments. Alignment A (red) and alignment B (blue) are two of the four alignments studied in this chapter.

- deterministic input cost and time to calculate the deterministic total cost and total time and the input cost and time distributions to model the variability,
- correlations between the cost variables,
- disruptive events, i.e. the probability of occurrence and the cost and time impacts.

Cost and Time Input for Tunnels

In this section, it is explained how the deterministic input cost and time and the cost and time input distributions to model the variability are obtained.

Costs per unit length and advance rates for the construction of tunnels are available from historical sources (Table 7.2). The cost per unit length and the advance rate depend on the construction conditions that can be good, medium, or poor (Table 7.3). The geology of the tunnel is described depending on the percentage of the tunnel length excavated in good, medium, and poor conditions. For instance, if the construction conditions are 40% good, 50% medium, and 10% poor, a tunnel is considered in good geology (Table 7.4). The mean cost per unit length and the mean advance rate for a tunnel geology are calculated: in good geology the mean cost per unit length is the lowest and the mean advance rate is the largest, while in poor geology the mean cost per unit length is the largest and the mean advance rate is the lowest (Table 7.4). Besides the mined tunnels in good/medium/poor geology, a few tunnels are built by cut & cover. The cost per unit length and the advance rate of tunnels built by cut & cover are considered to be independent of geology (Table 7.4).

Table 7.3: Costs per unit length, advance rates and times per unit length in tunnel construction depending on the construction conditions.

		Construction conditions		
		good	medium	poor
cost per unit length	[euro/m]	7,250	11,000	17,000
advance rate	[m/day]	4	2	1
time per unit length	[day/m]	0.25	0.5	1

Note that multiplying the mean advance rate with the tunnel length does not result in the mean construction time, T , of the tunnel since

$$T = \frac{l_1}{r_1} + \frac{l_2}{r_2} + \frac{l_3}{r_3} \neq \frac{\sum_{i=1}^3 l_i}{\bar{r}}, \quad (7.1)$$

Table 7.4: Mean costs per unit length, mean advance rates and mean times per unit length for tunnels in good/medium/poor geology, calculated based on the percentages of tunnel constructed in good/medium/poor construction conditions. The cost per unit length, the advance rate and the time per unit length for tunnels built by cut & cover are considered to be independent of geology.

Tunnel in	construction conditions			mean		
	good	medium	poor	cost per unit length [euro/m]	advance rate [m/day]	time per unit length [day/m]
	[%]					
good geology	40	50	10	10,100	2.7	0.45
medium geology	20	50	30	12,050	2.1	0.6
poor geology	10	20	70	14,826	1.5	0.825
cut & cover	–	–	–	10,600	2.67	0.375

where l_i is the length and r_i is the advance rate in good/medium/poor construction conditions, and \bar{r} is the mean advance rate. Alternatively, the mean construction time, T , of the tunnel can be calculated with the mean time per unit length since

$$T = t_1 \times l_1 + t_2 \times l_2 + t_3 \times l_3 = \bar{t} \times \sum_{i=1}^3 l_i, \quad (7.2)$$

where l_i is the length and t_i is the time per unit length in good/medium/poor construction conditions, and \bar{t} is the mean time per unit length. For simplicity, in the following the construction time is calculated with the times per unit length. These are presented with the costs per unit length and the advance rates in Tables 7.3 and 7.4.

The deterministic construction cost and time of a tunnel are calculated with the mean cost per unit length and the mean time per unit length. These are summarized for tunnels in good/medium/poor geology and for tunnels built by cut & cover in Table 7.5.

Table 7.5: Deterministic input costs and times per unit length of mined tunnels in good/medium/poor geology and of tunnels built in cut & cover.

Tunnel in	cost per unit length [euro/m]	time per unit length [day/m]
good geology	10,100	0.45
medium geology	12,050	0.6
poor geology	14,825	0.825
cut & cover	10,600	0.375

The distributions of the cost per unit length of a tunnel in good/medium/poor geology are

obtained by fitting lognormal distributions to the discrete distributions (Table 7.4). The fit is achieved by matching the mean and the standard deviation, i.e. the lognormal distribution has the same mean and standard deviation as the discrete distribution (Table 7.6). The probability density function and the cumulative distribution function of the lognormal distribution are plotted with the probability mass function and the cumulative distribution function, respectively, of the discrete distribution for tunnels in good, medium, and poor geology in Figure 7-4.

Table 7.6: Distributions of the costs per unit length for tunnels in good/medium/poor geology. The lognormal distributions are fitted to the discrete distributions by matching the means and the standard deviations.

Tunnel in good geology					
discrete distribution	40%	50%	10%	mean	st. dev.
cost per unit length [euro/m]	7,250	11,000	17,000	10,100	2,900
lognormal distribution	minimum	mode	98 th percentile		
cost per unit length [euro/m]	0	8,969	17,308	10,100	2,900
tunnel in medium geology					
discrete distribution	20%	50%	30%	mean	st. dev.
cost per unit length [euro/m]	7,250	11,000	17,000	12,050	3,537
lognormal distribution	minimum	mode	98 th percentile		
cost per unit length [euro/m]	0	10,645	20,867	12,050	3,537
tunnel in poor geology					
discrete distribution	10%	20%	70%	mean	st. dev.
cost per unit length [euro/m]	7,250	11,000	17,000	14,825	3,461
lognormal distribution	minimum	mode	98 th percentile		
cost per unit length [euro/m]	0	13,690	23,170	14,825	3,461

The lognormal distribution is fitted to the discrete distribution assuming the minimum of the lognormal distribution is equal to zero. Although the cost of excavating one meter of tunnel cannot be equal to zero, this assumption is reasonable since the first-percentile of the lognormal distribution for tunnels in good/medium/poor geology after the fitting is a relatively large cost per unit length (Table 7.7). For instance, in 99% of the cases the cost of excavating one meter of a tunnel in good/medium/poor geology is greater than 5,025/5,900/8,425 euro. These can be considered as the minimum costs per meter for tunnel construction in good/medium/poor geology.

The distributions of the time per unit length of a tunnel in good/medium/poor geology are obtained by fitting triangular distributions to the known discrete distributions. The fit is achieved by matching the mean and the standard deviation. This means that the triangular distribution has the same mean and standard deviation as the discrete distribution (Table 7.8). The probability

Table 7.7: 1-percentile of the cost per unit length distributions in good/medium/poor geology. Assuming the minimum of the lognormal distribution is equal to zero is reasonable since the 1-percentile is a relatively large cost per unit length.

Tunnel in	cost per unit length 1-percentile [euro/m]
good geology	5,025
medium geology	5,900
poor geology	8,425

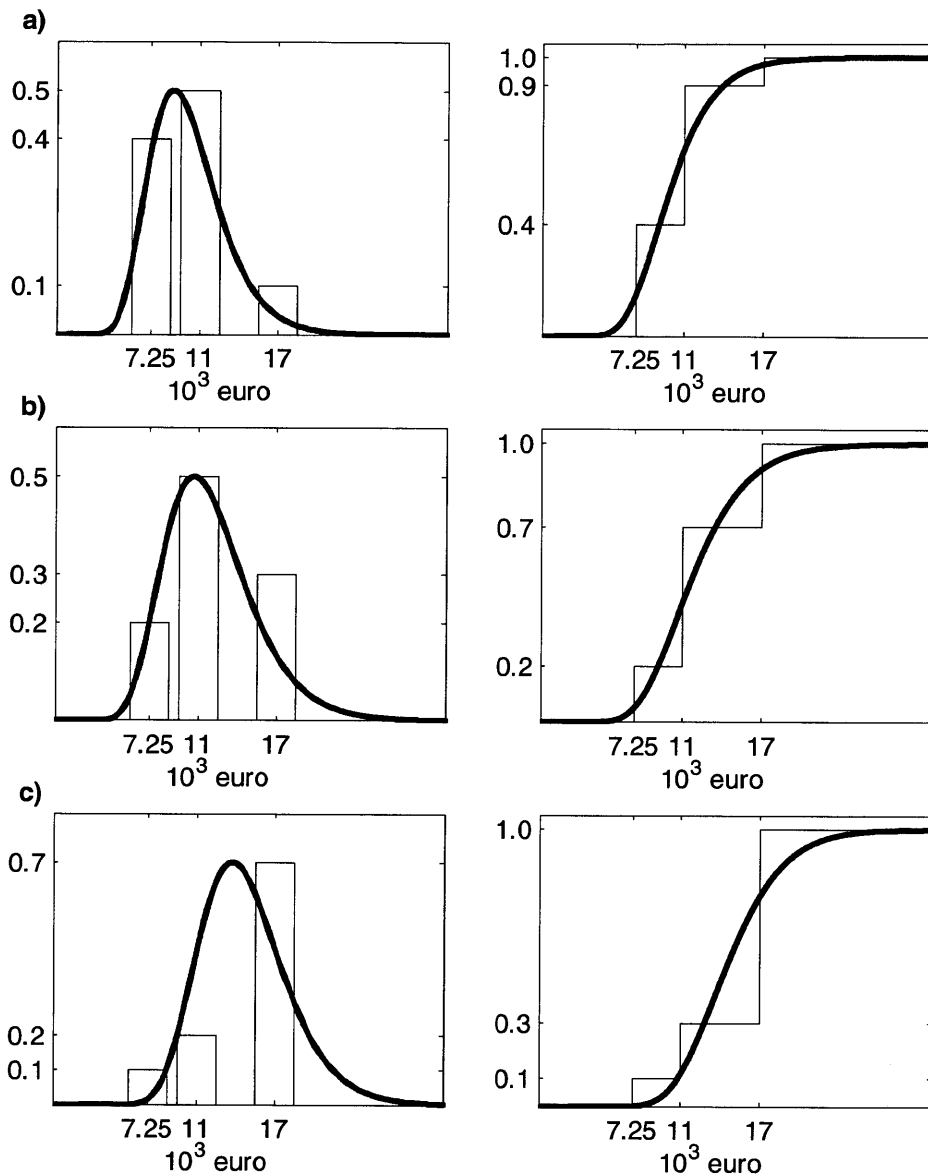


Figure 7-4: The probability density function of the lognormal distribution and the probability mass function of the discrete distribution are plotted (left) for the cost of tunnels in a) good, b) medium, and c) poor geology. The cumulative distribution functions of the lognormal distribution and of the discrete distribution are plotted (right) for the cost of tunnels in a) good, b) medium, and c) poor geology. Lognormal and discrete distributions have been fitted by matching the mean and the standard deviation. The lognormal distributions are the input for the uncertainty model.

density function and the cumulative distribution function of the triangular distribution are plotted with the probability mass function and the cumulative distribution function, respectively, of the discrete distribution for tunnels in good, medium, and poor geology in Figure 7-5.

Table 7.8: Distributions of the times per unit length for tunnels in good/medium/poor geology. The triangular distributions are fitted to the discrete distributions by matching the means, the standard deviations, and one point of the distributions.

Tunnel in good geology					
discrete distribution	40%	50%	10%	mean	st. dev.
time per unit length [day/m]	0.25	0.5	1	0.45	0.048
triangular distribution	minimum	mode	maximum		
time per unit length [day/m]	0.04	0.26	1.05	0.45	0.048
tunnel in medium geology					
discrete distribution	20%	50%	30%	mean	st. dev.
time per unit length [day/m]	0.25	0.5	1	0.6	0.078
triangular distribution	minimum	mode	maximum		
time per unit length [day/m]	0.01	0.45	1.35	0.6	0.078
tunnel in poor geology					
discrete distribution	40%	50%	10%	mean	st. dev.
time per unit length [day/m]	0.25	0.5	1	0.825	0.0076
triangular distribution	minimum	mode	maximum		
time per unit length [day/m]	0.08	1	1.39	0.825	0.0076

Since the triangular distribution requires three parameters to be determined, a point of the distribution needs to be matched. The cumulative distribution of the triangular distribution is matched at the middle of a discontinuity of the cumulative distribution of the discrete distribution (0.25 day/m, 0.5 day/m, 1 day/m) (Figure 7-5, right). For the case of the time per unit length in good and medium geology, the distributions could be matched at the first discontinuity (Figures

*The variability input data of tunnels are in form of discrete distributions. Continuous distributions (lognormal for input costs, triangular for input times) "equivalent" to the available discrete distributions are sought since the correlation model NORTA requires continuous input distributions and for consistency (input cost and time distributions of other structures are lognormal and triangular). "Equivalent" distributions are obtained by matching the mean and the standard deviation. Lognormal distributions are easily obtained since they are determined with two parameters (mean and standard deviation of the corresponding normal distribution). On the other hand, triangular distributions require three parameters (minimum, mode, and maximum) to be determined. Thus, they need a third equality. To obtain this, two alternatives have been considered. First, the CDF of the triangular distribution is matched at the middle of a discontinuity of the CDF of the discrete distribution. Second, the mode of the triangular distribution is set equal to the time of the discrete distribution with the largest probability. In both cases, there is a boundary condition: the minimum of the triangular distribution must be a positive number. Given this boundary condition, the first alternative is effective in matching the discrete and the triangular distributions in good and medium geology, while the second alternative is effective in matching the discrete and the triangular distributions in poor geology. Since neither alternative is effective in good, medium, and poor geology, a mixed approach is adopted: the input time distributions in good and medium geology are obtained with the first alternative, while the input time distribution in poor geology is obtained with the second alternative.

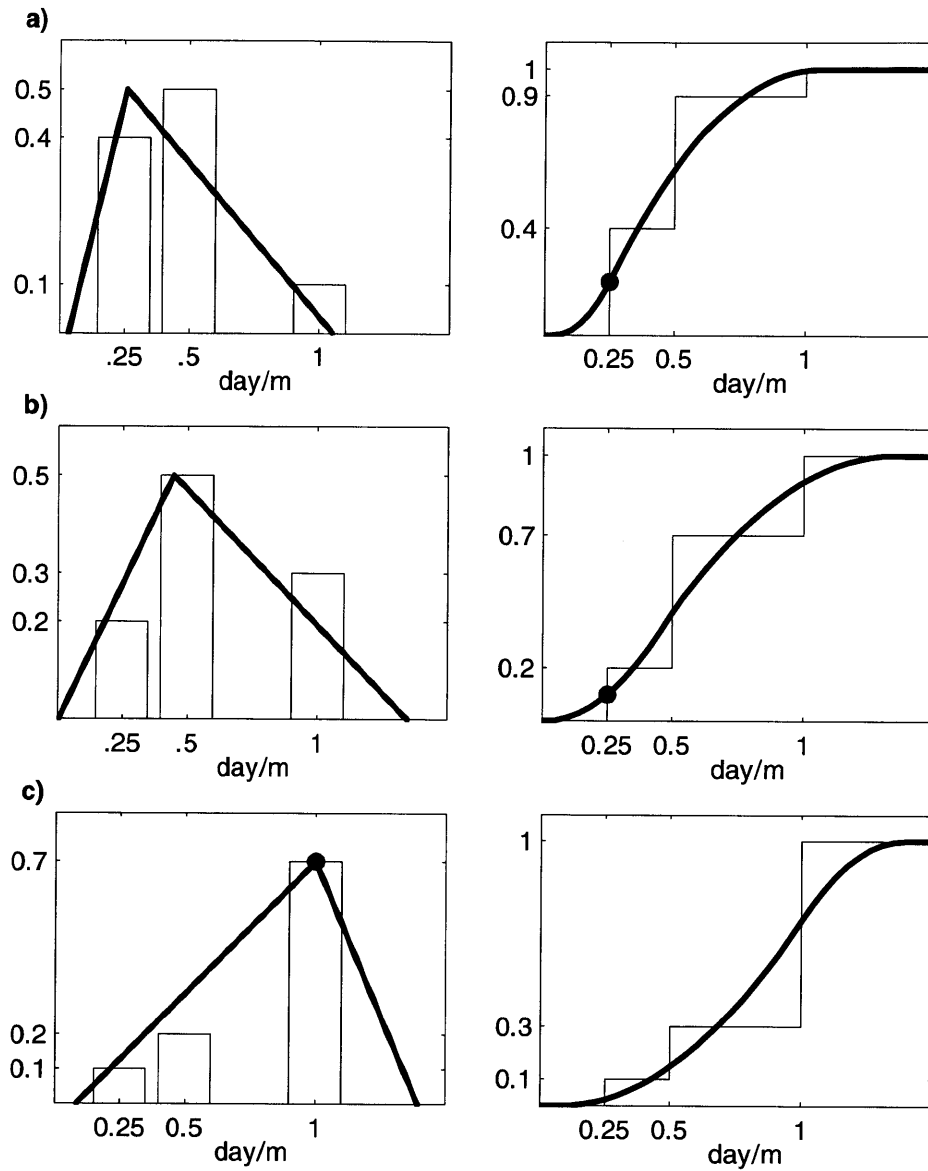


Figure 7-5: The probability density function of the triangular distribution and the probability mass function of the discrete distribution are plotted (left) for the time per unit length of tunnels in a) good, b) medium, and c) poor geology. The cumulative distribution functions of the triangular distribution and of the discrete distribution are plotted (right) for the time of tunnels in a) good, b) medium, and c) poor geology. Triangular and discrete distributions have been fitted by matching the mean, the standard deviation, and a point of the cumulative distribution in a) and b), and the mode of the probability mass/density function in c). The triangular distributions are the input for the uncertainty model.

7-5a and 7-5b, right), while for the case of the time per unit length in poor geology the distribution could not be matched at any of the three discontinuities. Alternatively, the distribution of the time per unit length in poor geology is matched at the mode (Figure 7-5c, left), assuming that the mode of the triangular distribution is equal to 1 day/m. This assumption is reasonable since the probability mass function of the time per unit length at 1 day/m is equal to 70% (Figure 7-5c, left).

The lognormal distributions of the cost per unit length and the triangular distributions of the time per unit length of a tunnel in good/medium/poor geology are the input for the uncertainty model (Table 7.9).

Table 7.9: Minimum, mode, and 98th percentile of the lognormal distributions of the cost per unit length and minimum, mode, and maximum of the triangular distributions of the time per unit length for a tunnel in good/medium/poor geology. They are the input for the uncertainty model.

Tunnel in	cost per unit length [euro/m]			time per unit length [day/m]		
	minimum	mode	98 th percentile	minimum	mode	maximum
good geology	0	8,969	17,308	0.04	0.26	1.05
medium geology	0	10,645	20,867	0.01	0.45	1.35
poor geology	0	13,690	23,170	0.08	1	1.39

The probability distributions of the cost per unit length and of the time per unit length of a tunnel built by cut & cover could not be obtained, since the data sources (Table 7.2) do not include discrete distributions for tunnels built by cut & cover. Thus, the deterministic cost per unit length and time per unit length for tunnels built by cut & cover are also the input for the uncertainty model.

Correlation Input for Tunnels

In this section, the correlations between the costs in tunnel construction are presented.

The data on the cost correlations stem from expert opinions (Table 7.2). The tunnel expert has more than 30 years of experience in tunneling. The estimations of correlations were obtained by means of a questionnaire (appendix A).

The construction of tunnels is modeled with one activity that represents all the processes to excavate one unit length of tunnel. The tunnel expert was asked to estimate the correlation between the costs of the repeated excavation activity. He indicated that the correlations follow a specific pattern:

Disruptive Event Input for Tunnels

The probability of occurrence and the cost and time impacts of disruptive events in tunnel construction are discussed here.

The data on disruptive events in tunnel construction stem from expert opinions (Table 7.2). Similarly to the correlation input, estimations were obtained by means of the questionnaire (appendix A).

The tunnel expert identified the two major disruptive events in tunnel construction in a tunnel cave-in and a sudden water inflow. He estimated the probability of occurrence and the probability distributions of cost and time impacts (Table 7.11). These are the input for the uncertainty model.

Table 7.11: Probability of occurrence and triangular distributions of the cost and time impacts of disruptive events in tunnel construction. They are the input for the uncertainty model.

Cave-in				
probability of occurrence				1/800m
		minimum	mode	maximum
cost	[10 ⁶ euro]	0.2	1	10
time	[day]	5	30	270
water inflow				
probability of occurrence				1/500m
		minimum	mode	maximum
cost	[10 ⁶ euro]	0.05	0.2	1
time	[day]	1	15	60

With the discussion of disruptive events, the presentation of the input data to model the uncertainty in tunnel construction is complete. Next, the input data to model the construction of viaducts are presented.

7.2.2 Viaduct Data

In this section, the input data to model the construction cost and time of viaducts is presented in the following order:

- deterministic input cost and input cost distributions to model the variability,
- deterministic input time and input time distributions to model the variability,
- disruptive events, i.e. the probability of occurrence and the cost and time impacts.

The correlations between the costs of a repeated activity are not modeled in viaduct construction. Despite the repetitiveness of the construction process, the costs of the activities to build the deck sections, the piers, the foundations, the abutments and the technical blocks are assumed independent for the following reasons (section 5.4): the deck section length varies from deck section to deck section; the pier height varies from pier to pier, the foundations vary depending on the geology, the two abutments and the two technical blocks are at the opposite ends of the viaduct. Thus, the costs of these activities are assumed independent.

Cost Input for Viaducts

In this section, the deterministic input cost and the input cost distributions to model the variability in viaduct construction are explained.

The deterministic input costs of viaduct activities are not available from historical sources (Table 7.2): although the total costs for the construction of the viaducts is known, the calculation of the total costs is not available from the sources. To reproduce the calculation of the viaduct costs, the design engineer suggested to calculate the viaduct costs as a function of the geometric parameters of the viaduct (Table 7.12) with cost expressions (Table 7.13) (Leite 2009). With these, the deterministic construction costs of viaducts are calculated.

Table 7.12: Geometric parameters of the viaducts. The construction costs of viaducts are calculated based on these geometric parameters.

L	viaduct length
l	span length
h	pier height
α	fraction of piers in the viaduct with pile foundation
d	depth of piles
w	viaduct width (14 m)
N	number of spans
S	number of sloped piers

The input probability distributions of the costs of the viaduct activities are not available. However, the minimum, the mode, and the maximum costs per unit area of a viaduct are available for several span lengths l and pier heights h from historical sources (tHR 2007). They are used to obtain the probability distribution of a cost factor that is then multiplied with the deterministic input costs of the viaducts. The ratios $\frac{\text{minimum}}{\text{mode}}$ and $\frac{\text{maximum}}{\text{mode}}$ of the costs per unit area are

Table 7.13: The deterministic input costs are calculated with the expressions in the table. The sources (Table 7.2) assume that activities pre-work (construction site setup) and pre-deck (deck construction setup) require time but do not cause additional cost. The activity construction method models the cost of employing staging or a gantry in the construction of the deck sections.

Activity	cost expressions [euro]	
abutment	250,000	$L \leq 1,000m$
	400,000	$L > 1,000m$
deck section	$1,100 + 60l$	$l \leq 35m$
	$1,000 + 75l$	$l > 35m$
pier	$(9,000 + 1,500h)(0.6 + 0.012l)$	$l \leq 35m, h \leq 15m$
	$(10,000 + 1,800h)(0.6 + 0.012l)$	$l \leq 35m, h > 15m$
	$(12,000 + 2,000h)(0.4 + 0.022l)$	$l > 35m$
foundation footing	$1,200l(0.4 + 0.03h)$	
foundation pile set	$\alpha(0.3 + 0.13l)(0.7 + 0.015h)600d$	
sloped pier	$\left\{ \begin{array}{l} (9,000 + 1,500h)(0.6 + 0.012l) \\ (10,000 + 1,800h)(0.6 + 0.012l) \\ (12,000 + 2,000h)(0.4 + 0.022l) \end{array} \right\}$ $+1,200l(0.4 + 0.03h) +$ $+\alpha(0.3 + 0.13l)(0.7 + 0.015h)600d +$ $+8(1/2h^3)$	$l \leq 35m, h \leq 15m$
		$l \leq 35m, h > 15m$ $l > 35m$
technical block	150,000	
finishing	450 [/m]	
pre-work	0	
pre-deck	0	
construction method	$4whL$	with staging
	$120,000 + 25,000N$	with gantry

calculated (Table 7.14) to determine the lognormal distribution of a cost factor: the minimum cost factor is equal to the average $\frac{\text{minimum}}{\text{mode}}$ ratio, the mode cost factor is equal to 1.0, and the 98th percentile is equal to the $\frac{\text{maximum}}{\text{mode}}$ ratio (compare Table 7.15 with Table 7.14). The probability distribution of the cost factor is displayed in Figure 7-6.

The probability distribution of the cost of a viaduct activity (distribution of input cost) is obtained by multiplying the distribution of the cost factor with the deterministic input cost of the viaduct activity (deterministic input cost):

$$(\text{distribution of input cost})_i = (\text{distribution of cost factor}) \times (\text{deterministic input cost})_i \quad (7.4)$$

The lognormal distribution of the viaduct cost factor is the input for the uncertainty model (Table 7.15 and Figure 7-6).

Table 7.14: Minimum-to-mode and maximum-to-mode ratios for different span lengths and pier heights. The average minimum-to-mode and maximum-to-mode ratios are used to obtain the lognormal distribution of the viaduct cost factor.

Span length [m]	pier height [m]	cost per unit area			ratios	
		minimum [euro/m ²]	mode [euro/m ²]	maximum [euro/m ²]	$\frac{\text{minimum}}{\text{mode}}$ [-]	$\frac{\text{maximum}}{\text{mode}}$ [-]
$l = 30$ m	$10 \text{ m} < h < 20 \text{ m}$	600	750	1,000	0.8	1.33
$l = 40$ m	$20 \text{ m} < h < 40 \text{ m}$	800	1,000	1,200	0.8	1.20
$l = 60$ m	$40 \text{ m} < h < 60 \text{ m}$	1,000	1,300	1,500	0.77	1.15
$l = 120$ m	$8 \text{ m} < h < 20 \text{ m}$	2,500	3,000	4,000	0.83	1.33
average					0.80	1.26

Table 7.15: The probability distribution of the cost factor is the input for the uncertainty model.

	Minimum	mode	98 th percentile
cost factor [-]	0.8	1.0	1.26

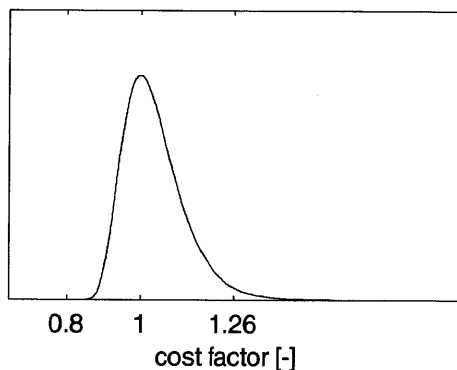


Figure 7-6: Probability distribution of the cost factor. It is multiplied with the deterministic input costs of viaducts to obtain the probability distributions of the costs of viaducts.

Time Input for Viaducts

In this section, the deterministic input times and the time distributions to model the variability in viaduct construction are presented.

The deterministic input times of viaduct activities are available from historical sources (Table 7.2). For the different activities in viaduct construction, the deterministic input times are summarized in Table 7.16. They are used to calculate the deterministic construction time of viaducts.

Table 7.16: Deterministic input times of viaduct activities. The sources (Table 7.2) assume a pier, a footing and a pile set are constructed in parallel with a deck section, and they assume their construction times equal to zero, since the deck section requires more time to be constructed. The activity construction method models the cost of employing staging or a gantry in the construction of the deck sections and does not require additional time.

Activity	time [days]
abutment	56
deck section	14
pier	0
foundation footing	0
foundation pile set	0
sloped pier	0
technical block	28
finishing	0.014/m
pre-work	42
pre-deck	with staging 28
	with gantry 56
construction method	0

The input probability distributions of the viaduct times stem from expert opinions (Table 7.2). The viaduct expert has more than 20 years of experience in her field. The estimations of the time distributions (Table 7.17) were obtained by means of a questionnaire (appendix A).

The probability distributions of the viaduct times are in disagreement with the deterministic input times (Table 7.17): the deterministic input time often lies outside of the probability distribution estimated by the expert (e.g. pre-work time) and, in general, the deterministic input time is different from the mode of the estimated probability distribution (Molenaar et al. (2010) showed that the deterministic input cost/time is equal to the mode of the cost/time distribution when probabilistically analyses on the project total cost and total time are conducted by transportation agencies). In order to obviate the disagreements between deterministic input time and distribution mode, the probability distribution estimated by the expert is moved along the time axis in order to

Table 7.17: Minimum, mode, and maximum viaduct times estimated by the expert, and deterministic input times. The input time distributions are determined by matching the mode of the estimated time distribution with the deterministic input time (Figure 7-7).

Activity	Time [days]			
	viaduct expert			(RAVE 2008b)
	minimum	mode	maximum	deterministic
pre-work	7	10	20	42
abutment	23	30	60	56
pre-deck	with staging	0	0	28
	with gantry	8	15	56
deck section	4	7	14	14
technical block	7	10	20	28
finishing	7	10	20	0.14 [/m]

match the mode of the distribution with the deterministic input time (Figure 7-7). The minimum, mode, and maximum of the time distribution are calculated with the expression

$$\begin{pmatrix} \text{minimum} \\ \text{mode} \\ \text{maximum} \end{pmatrix} = \frac{\text{time}_{deterministic}}{\text{mode}_{expert}} \times \begin{pmatrix} \text{minimum} \\ \text{mode} \\ \text{maximum} \end{pmatrix}_{expert} \quad (7.5)$$

For instance, the minimum, the mode and the maximum of the time distribution to build an abutment (Figure 7-7) are calculated

$$\begin{pmatrix} 43 \\ 56 \\ 112 \end{pmatrix} = \frac{56}{30} \times \begin{pmatrix} 23 \\ 30 \\ 60 \end{pmatrix} \quad (7.6)$$

In the case of the activity pre-deck (with staging), the minimum, mode, and maximum of the time distribution are calculated using the values estimated by the expert for the activity pre-deck (with gantry). The obtained time distributions of the viaduct times are the input for the uncertainty model (Table 7.18).

A similar approach as explained above will be used to obtain the time and cost input distributions for the construction of cuts and embankments (see below).

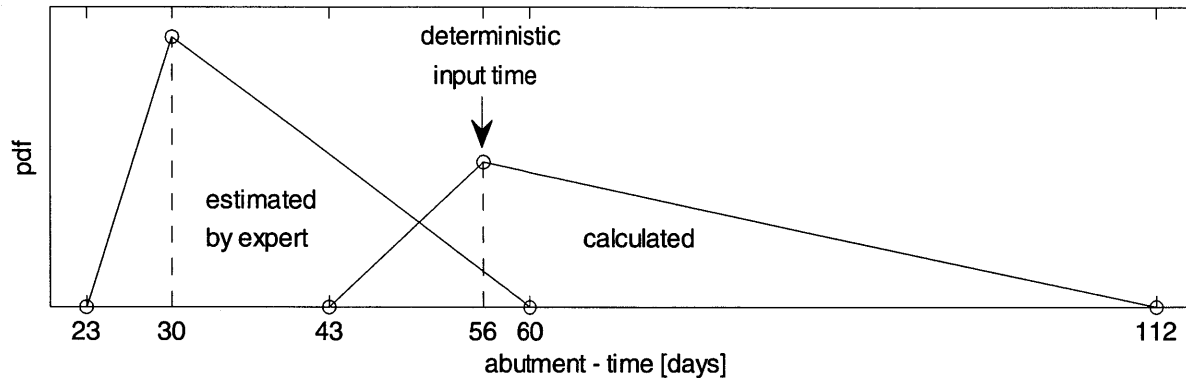


Figure 7-7: Deterministic input time, time distribution estimated by the expert and calculated time distribution for the construction of a viaduct. The calculated time distribution is obtained by moving the time distribution estimated by the expert along the time axis and matching the mode of the estimated time distribution with the deterministic input time. It is the input for the uncertainty model.

Table 7.18: Minimum, mode and maximum of the triangular distributions of the viaduct times. They are the input for the uncertainty model.

Activity	Time [days]			
	minimum	mode	maximum	
pre-work	29	42	84	
abutment	43	56	112	
pre-deck	with staging	19.4	28	56
	with gantry	29.9	56	112
deck section	8	14	28	
technical block	20	28	56	
finishing	0.010	0.014	0.028	

Disruptive Event Input for Viaducts

The probability of occurrence and the cost and time impacts of disruptive events in viaduct construction are discussed here.

The data on disruptive events in viaduct construction stem from expert opinions (Table 7.2). Estimations were obtained by means of the questionnaire (appendix A).

The viaduct expert identified the two major disruptive events in viaduct construction as 1) differing ground conditions requiring the redesign of the foundations, and 2) construction problems or accidents. She estimated the probability of occurrence and the probability distributions of cost and time impacts (Table 7.19). These are the input for the uncertainty model.

Table 7.19: Probability of occurrence and triangular distributions of the cost and time impacts of disruptive events in viaduct construction. They are the input for the uncertainty model.

Differing ground conditions				
probability of occurrence		5%		
		minimum	mode	maximum
cost	[% of cost]	0	20	50
time	[day]	2	60	240
construction accidents/problems				
probability of occurrence		1%		
		minimum	mode	maximum
cost	[% of cost]	1	15	30
time	[day]	7	60	365

With the discussion of the disruptive events the presentation of the input data to model the uncertainty in viaduct construction is complete. Next, the input data to model the construction of cuts and embankments are presented.

7.2.3 Cuts and Embankments Data

In this section, the input data to model the construction cost and time of cuts and embankments are presented in the following order:

- deterministic input cost and input cost distributions to model the variability,
- deterministic input time and input time distributions to model the variability,
- correlations between the cost variables,

- disruptive events, i.e. the probability of occurrence and the cost and time impacts.

Cost Input for Cuts and Embankments

In this section, the deterministic input cost and the input cost distributions to model the variability are explained.

The deterministic input costs per unit volume of cuts and embankments are available from historical sources (RAVE 2006b, RAVE 2008b). Since the two sources report slightly different costs, the data from the most recent source are used here (Table 7.2).

The deterministic input costs per unit volume for the construction of cuts and embankments are summarized in Table 7.20. They are used to calculate the deterministic construction cost of cuts and embankments.

Table 7.20: Deterministic input costs per unit volume of activities in the construction of cuts and embankments.

Activity	cost per unit volume [euro/m ³]
clearing	1.92
excavating mechanical	3.46
blasting	8.08
improving	4.85
filling	1.25
capping	6.75
sub-ballast	14.4

The input probability distributions of earthwork (cuts and embankments) costs are available from a different source (Table 7.2). They are summarized in Table 7.21.

The input probability distributions of the earthwork costs are in disagreement with the deterministic input costs (Table 7.21): the deterministic input cost often lies outside of the probability distribution and in general the deterministic input cost is different from the mode of the probability distribution (Molenaar et al. 2010). These disagreements are obviated as already explained for the time input of viaducts (page 261), i.e. by adapting the probability distribution to the deterministic input cost. The obtained probability distributions of the earthwork costs are the input for the uncertainty model (Table 7.22).

Table 7.21: Minimum, mode, and maximum costs per unit volume of earthwork (cuts and embankments) construction, and deterministic input costs. The input cost distributions are determined by matching the mode of the cost distribution with the deterministic input cost (page 261).

Activity	cost per unit volume [euro/m ³]				
	(tHR 2007)		(RAVE 2008b)		
	minimum	mode	maximum	deterministic	
clearing	1	2	3	1.92	
excavating	mechanical	2	2.5	5	3.46
	blasting	5	8	15	8.08
improving	10	12	15	4.85	
filling	3	4	7	1.25	
capping	10	15	20	6.75	
sub-ballast	15	20	25	14.4	

Table 7.22: Minimum, mode and 98th percentile of the lognormal distributions of the cost per unit volume for the construction of cuts and embankments. They are the input for the uncertainty model.

Activity	cost per unit volume [euro/m ³]			
	minimum	mode	98 th	percentile
clearing	0.96	1.92		2.88
excavating	mechanical	2.77	3.46	6.92
	blasting	5.05	8.08	15.15
improving	4.04	4.85		6.06
filling	0.94	1.25		2.19
capping	4.5	6.75		9
sub-ballast	10.8	14.4		18

Time Input for Cuts and Embankments

In this section, the deterministic input time and the input time distributions for earthwork (cuts and embankments) construction are presented.

The production rates of cuts and embankments are not available from historical sources (Table 7.2). They are obtained from the probability distributions of the production rates, as explained below.

The input probability distributions of the production rates of earthwork (cuts and embankments) construction stem from expert opinions (Table 7.2). The earthwork expert has more than 30 years of experience in his field. The estimations of the time distributions were obtained by means of a questionnaire (appendix A).

The earthwork expert estimated the minimum, mode and maximum of the triangular distribu-

tions of the production rates (Table 7.23).

Table 7.23: Minimum, mode, and maximum production rates in earthwork (cuts and embankments) construction, as estimated by the earthwork expert.

Activity	production rate				
	minimum	mode	maximum		
clearing	3	5	7	[acres/day]	
excavating	mechanical	3,000	4,200	5,800	[cy/day]
	blasting	1,000	1,300	2,000	[cy/day]
improving	same as excavating				
filling	same as excavating				
capping	5,000	5,500	6,000	[cy/day]	
sub-ballast	1,200	1,400	1,500	[tonnes/day]	

In order to obtain the input for the uncertainty model, the following adjustments were applied to the estimated production rates:

- The units of the production rates were changed from the imperial system to the metric system (m^2/day or m^3/day).
- The production rate of the sub-ballast activity was changed from a mass production rate to a volume production rate assuming a rock density of $2.7 t/m^3$.
- Since the percentage of mechanical and blasting excavation varies from cut to cut, the production rates of filling and capping were assumed equal to the production rate of mechanical excavation.
- The distribution of a production rate is expected to be skewed to the left, i.e. the mean production rate is expected to be smaller than the most probable production rate, in other words the mean construction time is expected to be larger than the most probable construction time (Note that the construction time is equal to the volume divided by the production rate). However, the distributions of the mechanical excavation and the blasting excavation production rates estimated by the expert are skewed to the right. This is corrected reducing the maximum production rates until the distributions become symmetric.
- Eight crews excavate the cuts: thus, the excavation rates are multiplied by eight. In Alignments A-a and A-c, the time to excavate the cuts is equal to 500 days and 620 days (RAVE 2006b). Given the excavation volumes (RAVE 2006b), and the production rates per crew

(mode production rates, Table 7.23), the number of crews needed is calculated: in Alignment A-a, 6 to 7 crews are needed; in Alignment A-c, 9 to 10 crews are needed. Following these calculations, eight crews to excavate the cuts are assumed.

The obtained probability distributions of the production rates are the input for the uncertainty model (Table 7.24).

Table 7.24: Minimum, mode, and maximum production rates in earthwork (cuts and embankments) construction. They are the input for the uncertainty model.

Activity		production rate			
		minimum	mode	maximum	
clearing		12,141	20,234	28,328	[m ² /day]
excavating	mechanical	18,349	25,689	35,475	[m ³ /day]
	blasting	6,116	7,951	12,233	[m ³ /day]
improving		18,349	25,689	35,475	[m ³ /day]
filling		18,349	25,689	35,475	[m ³ /day]
capping		3,823	4,205	4,587	[m ³ /day]
sub-ballast		403	470	504	[m ³ /day]

The modes of the distributions are the deterministic input production rates (Molenaar et al. 2010) (Table 7.25).

Table 7.25: Deterministic input production rates of activities in the construction of cuts and embankments.

Activity		production rate	
clearing		20,234	[m ² /day]
excavating	mechanical	25,689	[m ³ /day]
	blasting	7,951	[m ³ /day]
improving		25,689	[m ³ /day]
filling		25,689	[m ³ /day]
capping		4,205	[m ³ /day]
sub-ballast		470	[m ³ /day]

Correlation Input for Cuts and Embankments

In this section, the correlations between the costs in earthwork (cuts and embankments) construction are presented.

The data on the cost correlations stem from expert estimations (Table 7.2). these were obtained by means of a questionnaire (appendix A).

The construction of a cut is modeled with four activities (clearing, excavating, capping, and sub-ballast), while the construction of an embankment is modeled with five activities (clearing, improving, filling, capping, and sub-ballast). For the cost correlations, the focus is on the excavation and the filling activities, which represent the largest part of the earthwork construction costs. The earthwork expert was asked to estimate the correlation between the costs of the repeated excavation activity and between the costs of the repeated filling activity. Unfortunately, he did not feel confident to express opinions on correlations. Due to the similarity in the process of tunnel excavation and earthwork construction, the correlations between the costs of the repeated excavation activity and between the costs of the repeated filling activity are assumed to be equal to the correlations between the costs of the repeated excavation activity in tunnel construction (Table 7.26). They are the input for the uncertainty model.

Table 7.26: Correlations between the costs of excavating and filling activities in earthwork (cuts and embankments) construction. They are the input for the uncertainty model.

Correlation coefficient 1, ρ_1	0.5
number of rows and columns, N , with ρ_1	50
correlation coefficient 2, ρ_2	0.9

Disruptive Event for Cuts and Embankments

The probability of occurrence and the cost and time impacts of disruptive events in earthwork (cuts and embankments) construction are discussed here.

The data on disruptive events in earthwork (cuts and embankments) construction stem from expert estimations (Table 7.2). These were obtained by means of the questionnaire (appendix A).

The earthwork expert identified the two major disruptive events in viaduct construction as 1) a flooding of the construction area, and 2) differing site conditions (Table 7.27). He estimated the probability of occurrence and the cost and time impacts of the disruptive events. For the rare event "flooding" cost and time impacts are triangularly distributed, while for the rare event "differing site conditions" they are uniformly distributed. The probability of occurrence and the probability distributions of cost and time impacts of the disruptive events are the input for the uncertainty model.

The disruptive events in earthwork construction complete the set of input data to calculate the deterministic total cost and total time and to model the uncertainty in the construction of tunnels,

Table 7.27: Probability of occurrence and probability distributions of the cost and time impacts of disruptive events in earthwork (cuts and embankments) construction. They are the input for the uncertainty model.

Flooding				
probability of occurrence		1/year		
		minimum	mode	maximum
cost	[10 ⁶ euro]	0.15	0.7	2.0
time	[day]	1	18	35
differing site conditions				
probability of occurrence		50%		
		minimum	mode	maximum
cost	[% of cost]	5	–	30
time	[% of time]	15	–	100

viaducts, cuts and embankments. In the next sections, these input data are used to obtain the total cost and the total time of construction.

7.3 Application of the Construction Model: Deterministic Total Cost and Total Time

In this section, the construction model is applied to the construction of four alignments of the new Portuguese high speed rail line in the simulation tool DAT. The construction model represents a rail line with a sequence of four main types of structures: tunnels, viaducts, cuts and embankments. The start position, the end position, and the length of the structures in the four alignments are listed in appendix C, Tables C.1 to C.4. In the following, the ground parameters, the ground classes and the construction methods to model the construction in the DAT are presented for tunnels (section 7.3.1), viaducts (section 7.3.2), and cuts and embankments (section 7.3.3). The deterministic total cost and total time of the structures are calculated. They will be compared with the total cost and total time distributions when modeling the uncertainty (section 7.4).

7.3.1 Tunnels

In this section, the input for the DAT to model the construction of the tunnels is explained. Specifically, the ground parameters and their states, the ground classes and the construction methods are presented. The deterministic total cost and total time are calculated.

The tunnel ground parameter set consists of eight ground parameters GP1 to GP8 (Table 7.28): one related to the magnitude of the overburden, two related to rock type, and five related to fault orientations. The ground parameter related to the magnitude of the overburden (GP1) is called overburden; the two rock type related ground parameters (GP2 and GP3) are called granite and meta-sediment; the five fault orientation related ground parameters (GP4 to GP8) are called perpendicular, oblique, acute-angled, parallel, and transverse. While the ground parameter overburden has two states called $< 2D$ and $\geq 2D$, where D stands for the diameter of the tunnel, the other seven ground parameters have three states called "yes", "no" and "nk" (not known), respectively.

Table 7.28: Tunnel ground parameters GP1 to GP8 with their states. Ground classes are determined based on the combination of ground parameter states.

Ground parameters	state 1	state 2	state 3
GP1 overburden	$< 2D$	$\geq 2D$	–
GP2 granite	yes	no	nk
GP3 meta-sediment	yes	no	nk
GP4 perpendicular	yes	no	nk
GP5 oblique	yes	no	nk
GP6 acute-angled	yes	no	nk
GP7 parallel	yes	no	nk
GP8 transverse	yes	no	nk

The ground classes are determined by the combination of the ground parameters states (Table 7.29). Four ground classes are defined: good, medium, poor, and reduced overburden. The first three ground classes refer primarily to the geological conditions of the ground, whereas the last ground class refers primarily to the magnitude of the overburden. The ground parameter states and the ground classes of all tunnels in the four alignments are listed in appendix C, section C.2.1.

There are four tunnel construction methods: cut & cover, NATM (New Austrian Tunneling Method) fast, NATM medium, NATM slow. These are directly related to the ground classes (Table 7.30): cut & cover is used with ground class reduced overburden, NATM fast with ground class good, NATM medium with medium, and NATM slow with poor. A tunnel construction method is characterized by an activity network and activity cost and time equations. The tunnel activity network consists of one activity modeling all operations of excavating and supporting a unit length of tunnel (chapter 3). The cost and time equations of the activities (3.3 and 3.4), as defined in

Table 7.29: Definition of the four tunnel ground classes: good, medium, poor and reduced overburden. Different combinations of ground parameters states can correspond to the same ground class. For instance, the ground class medium is defined with three combinations of ground parameter states, which are listed in the second, third and fourth row of the table. An asterisk means that all states of the ground parameter fit in the ground class definition.

Ground classes	ground parameters							
	GP1	GP2	GP3	GP4	GP5	GP6	GP7	GP8
good	$\geq 2D$	yes	no	no	no	no	no	no
medium	$\geq 2D$	no	yes	yes	no	no	no	no
medium	$\geq 2D$	no	yes	no	yes	no	no	no
medium	$\geq 2D$	yes	*	*	*	*	no	*
poor	$\geq 2D$	no	yes	*	*	*	no	*
poor	$\geq 2D$	*	*	*	*	*	yes	*
poor	$\geq 2D$	*	*	nk	nk	nk	nk	nk
reduced overburden	$< 2D$	*	*	*	*	*	*	*

chapter 3, are presented here again for reason of clarity:

$$\text{cost} = (\text{cost per unit length}) \times \text{length}, \quad (7.7)$$

$$\text{time} = (\text{time per unit length}) \times \text{length}. \quad (7.8)$$

The cost per unit length and the time per unit length (Table 7.31), and the tunnel lengths determine the construction cost and time of the tunnels. These are calculated for all tunnels in the four alignments in Table 7.32. A more detailed table including the construction method of all tunnels in the four alignments can be found in appendix C, section C.2.2.

Table 7.30: Relationship between tunnel ground classes and tunnel construction methods.

Ground class	construction method
reduced overburden	cut & cover
good	NATM fast
medium	NATM medium
poor	NATM slow

The deterministic total cost and total time (assuming sequential construction of the tunnels) for the construction of all the tunnels in the four alignments (Table 7.33) will be compared with the results of the uncertainty model (section 7.4).

Table 7.31: Cost per unit length and time per unit length of the four tunnel construction methods. They are used to calculate the construction cost and time of the tunnels along the four alignments.

Construction method	cost per unit length [euro/m]	time per unit length [day/m]
cut & cover	10,600	0.375
NATM good	10,100	0.45
NATM medium	12,050	0.6
NATM poor	14,825	0.825

Table 7.32: Deterministic construction cost and time of the tunnels in the four alignments.

Alignment	tunnel	cost [10 ⁶ euro]	time [days]	alignment	tunnel	cost [10 ⁶ euro]	time [days]	
A-a	T1	10.6	374	A-c	T1	10.5	371	
	T2	21.6	1,200		T2	20.6	1,027	
	T3	8.2	409		T3	5.1	225	
	T4	5.1	225		T4	39.3	2,186	
	T5	39.3	2,186		T5	44.1	2,194	
	T6	44.1	2,194					
B-a	T1	1.6	56	B-b	T1	1.6	56	
	T2	27.0	673		T2	27.0	673	
	T3	22.3	620		T3	5.2	259	
	T4	2.8	97		T4	4.3	237	
	T5	44.8	1,247		T5	7.2	400	
	T6	2.1	75		T6	21.2	529	
	T7	15.8	394		T7	62.5	1,738	
	T8	2.1	75		T8	2.1	75	
	T9	67.5	1,879		T9	44.0	1,223	
	T10	2.1	75		T10	1.6	56	
	T11	33.3	928					
	T12	1.6	56					

7.3.2 Viaducts

In this section, the input for the DAT to model the construction of viaducts is explained. Specifically, the ground parameter states, the ground classes and the construction methods are presented. The deterministic total cost and total time are calculated.

The ground parameter state and the ground class identify a viaduct uniquely, and a different construction method is assigned to each viaduct (section 6.3). This is modeled by assigning the same name to the viaduct, the zone, the ground parameter state, the ground class and the construction method (Table 7.34).

Table 7.33: Deterministic total cost and total time for the construction of all the tunnels in the four alignments.

Alignment	total cost [10 ⁶ euro]	total time [days]
A-a	128.7	6,589
A-c	119.5	6,003
B-a	223.1	6,174
B-b	176.6	5,246

Table 7.34: Zones, ground parameter states, ground classes, and construction methods of the five viaducts in alignment A-a. A different construction method is assigned to each viaduct in the four alignments.

Viaduct	zone	ground parameter state	ground class	construction method
Moinhos	Moinhos	Moinhos	Moinhos	Moinhos
Aldeia	Aldeia	Aldeia	Aldeia	Aldeia
Ave	Ave	Ave	Ave	Ave
Este	Este	Este	Este	Este
Cambeses	Cambeses	Cambeses	Cambeses	Cambeses

Each viaduct is modeled by a unique construction method and a unique activity network. The activity networks of all viaducts can be found in appendix C, section C.3.1. The cost and time equations of the activities (3.5 and 3.6), as defined in chapter 3, are presented here again for reason of clarity:

$$\text{cost} = (\text{element cost}), \quad (7.9)$$

$$\text{time} = (\text{element time}). \quad (7.10)$$

The element cost and the element time are functions of the geometric parameters (Tables 7.12, 7.13 and 7.16). The construction cost and time of the viaducts in the four alignments are summarized in Table 7.35. Their detailed calculation can be found in appendix C, section C.3.2.

The deterministic total cost and total time (assuming sequential construction of the viaducts) for the construction of all the viaducts in the four alignments (Table 7.36) will be compared with the results of the uncertainty model (section 7.4).

Table 7.35: Deterministic construction cost and time of the viaducts in the four alignments. They are calculated with the cost and time expressions in Table 7.13 and 7.16. Cambeses is the same viaduct in alignments A-a and A-c, as are Mau and Coura in alignments B-a and B-b. On the other hand, Ave and Este in alignments A-a and A-c and Cavado, Neiva, Lima and Boriz in alignments B-a and B-b are different viaducts.

Align.	viaduct	cost [10 ⁶ euro]	time [days]	align.	viaduct	cost [10 ⁶ euro]	time [days]
A-a	V1 Moinhos	6.12	655	A-c	V1 Laje	1.25	225
	V2 Aldeia	5.38	626		V2 Ave	14.30	447
	V3 Ave	66.39	978		V3 Este	2.22	326
	V4 Este	33.53	1,004		V4 Cambeses	4.65	328
	V5 Cambeses	4.65	328				
B-a	V1 Mau	2.77	397	B-b	V1 Mau	2.77	397
	V2 Cavado	28.7	903		V2 Cavado	25.49	801
	V3 Neiva	1.49	226		V3 Leiras	19.78	965
	V4 Outeiro	2.22	269		V4 Neiva	9.73	558
	V5 Lima	27.31	611		V5 Pombarinhos	1.52	240
	V6 Labruja	12.40	490		V6 Lima	34.19	1,118
	V7 Coura	3.06	284		V7 Moinho Velho	1.43	226
	V8 Reguengo	2.90	326		V8 Cabraçao	11.58	618
	V9 Boriz	12.19	574		V9 Bouca	2.91	298
			V10 Formigoso		0.81	197	
			V11 Coura		3.06	284	
			V12 Bogalheiro		9.39	532	
			V13 Boriz		17.07	650	

Table 7.36: Deterministic total cost and total time for the construction of all the viaducts in the four alignments.

Alignment	total cost [10 ⁶ euro]	total time [days]
A-a	116.1	3,590
A-c	22.4	1,325
B-a	93.0	4,080
B-b	139.7	6,882

7.3.3 Cuts and Embankments

In this section, the input for the DAT to model the construction of the cuts and embankments is explained. Specifically, the ground parameters, the ground classes and the construction methods are presented. The deterministic total cost and total time are calculated.

The ground parameter set of the cuts consists of a ground parameter (blasting) with four states (5%, 25%, 50%, 75%) that quantify the cut volume excavated with blasting, while the rest of the volume is excavated with mechanical means. The ground parameter set of the embankments consists of a ground parameter (improving) with two states (yes, no) that determine whether the embankment requires the ground under the embankment to be replaced before filling the embankment (Table 7.37). Each ground parameter state corresponds to one ground class so that cuts and embankments are characterized by a total of six ground classes (Table 7.37).

Table 7.37: Ground parameters and their states, and ground classes modeling the construction of cuts and embankments. The ground parameter blasting determines the percentage of cut excavated with blasting, while the ground parameter improving determines the need of improving the ground under the embankment. Six ground parameter states correspond to six ground classes.

Structure	ground parameter	state	ground class
cut	blasting	5%	1
		25%	2
		50%	3
		75%	4
embankment	improving	yes	A
		no	B

Cuts and embankments are modeled with six construction methods: clearing, excavating, capping, sub-ballast, improving, and filling (section 6.3). Each construction method is characterized by an activity network consisting of one activity, with the exception of excavating. This consists of two activities: one modeling the mechanical excavation and one modeling the blasting excavation. The cost and time equations of the activities (equations 3.5 and 3.6), as defined in chapter 3, are presented here again for reason of clarity:

$$\text{cost} = (\text{cost per unit volume}) \times \text{volume}, \quad (7.11)$$

$$\text{time} = \frac{\text{volume}}{\text{production rate}}. \quad (7.12)$$

The costs per unit volume, the production rates (Table 7.38), and the volumes of the cuts and embankments determine the construction cost and time. The volumes of all cuts and embankments in the four alignments are calculated in appendix C, section C.4.1. The total volumes (clearing, mechanical excavating, blasting excavation, improving, filling, capping, sub-ballast) of the four alignments are summarized in Table 7.39. The calculation of the construction cost and time of the cuts and embankments is detailed in appendix C, section C.4.2.

Table 7.38: Deterministic input costs per unit volume and production rates to calculate the deterministic total cost and total time of cuts and embankments construction.

Construction method	activity	unit cost [euro/m ³]	production rate
clearing	clearing	1.92	20,234 [m ² /day]
excavating	mechanical excavation	3.46	25,689 [m ³ /day]
	blasting excavation	8.08	7,951 [m ³ /day]
improving	improving	4.85	25,689 [m ³ /day]
filling	filling	1.25	25,689 [m ³ /day]
capping	capping	6.75	4,205 [m ³ /day]
sub-ballast	sub-ballast	14.40	470 [m ³ /day]

Table 7.39: Total volumes (clearing, mechanical excavating, blasting, improving, filling, capping, sub-ballast) of the four alignments. With the costs per unit volume and the production rates (Table 7.38), they determine the construction cost and time of the cuts and embankments in the alignments.

alignment	Volume [10 ³ m ³]			
	A-a	A-c	B-a	B-b
clearing	872	1,009	929	951
mechanical excavation	5,479	5,701	7,880	8,876
blasting excavation	4,285	3,019	1,803	2,775
improving	364	776	293	309
filling	2,753	4,370	5,899	5,578
capping	214	275	353	364
sub-ballast	107	137	177	182

The deterministic total cost and total time for the construction of all cuts and embankments in the four alignments (Table 7.40) will be compared with the results of the uncertainty model (section 7.4).

In the preceding, the input to the DAT to model the construction of cuts and embankments has been presented and the deterministic total cost and total time have been calculated.

Table 7.40: Deterministic total cost and total time for the construction of all cuts and embankments in the four alignments.

Alignment	total cost [10 ⁶ euro]	total time [days]
A-a	63.4	984
A-c	63.5	981
B-a	57.3	918
B-b	68.5	1,090

7.4 Application of the Uncertainty Model: Total Cost and Total Time Distributions

The uncertainty model is applied to the construction of the rail line, in particular to the construction of Alignment A-a. With the uncertainty model, the sources of uncertainty are modeled and their impact compared. The deterministic total cost and total time are calculated, then the construction is simulated considering the uncertainties associated with "variability", "cost correlations", and "disruptive events". This section presents the impacts of the three sources of uncertainty and their comparison. With the variability, the ranges of the total cost and total time distributions increase. With the cost correlations, the range of the total cost distribution grows significantly. With the disruptive events, the ranges of the total cost and total time distributions increase further. It is crucially important to model the sources of uncertainty in order to capture the ranges of the possible total cost and total time.

In the following, the impacts on the total cost and total time distributions of the sources of uncertainty variability (section 7.4.1), cost correlations (section 7.4.2), and disruptive events (section 7.4.3) are discussed.

7.4.1 Variability

In this section, the deterministic total cost and total time are compared with the total cost and total time distributions when modeling variability. First, the deterministic input cost and time (used to calculate the deterministic total cost and total time) and the input cost and time distributions (used to simulate the total cost and total time distributions) are presented. Then, the impact of variability on the total cost and total time is discussed.

The construction of alignment A-a is simulated considering one source of uncertainty, the variability in the construction process; the simulation results are then compared with the deterministic total cost and total time. The deterministic input cost and time, already presented in section 7.2, are summarized in Tables 7.41 (tunnels), 7.42 (viaducts), and 7.43 (cuts and embankments). The input cost and time distributions, already presented in section 7.2, are summarized in Figures 7-8 (tunnels), 7-9 and 7-10 (viaducts), and 7-11 and 7-12 (cuts and embankments).

The deterministic total cost and total time and the scattergrams of the total cost and total

Table 7.41: Deterministic input cost and time per unit length for the construction of tunnels in good/medium/poor geology and of tunnels built by cut & cover. Note that the deterministic input costs per unit length and the deterministic input times per unit length are equal to the mean input costs per unit length and the mean input times per unit length in Figure 7-8.

Tunnel in	cost per unit length [euro/m]	time per unit length [day/m]
good geology	10,100	0.45
medium geology	12,050	0.6
poor geology	14,825	0.825
cut & cover	10,600	0.375

Table 7.42: Expressions to calculate the deterministic input cost and time of viaducts. Note that the deterministic input times are smaller than the mean input times in Figure 7-10. Since the deterministic input cost factor is equal to 1, this is smaller than the mean cost factor (Figure 7-9).

Activity		cost expressions [euro]	time expressions [days]
abutment	$L \leq 1,000m$	250,000	56
	$L > 1,000m$	400,000	
deck section	$l \leq 35m$	$1,100 + 60l$	14
	$l > 35m$	$1,000 + 75l$	
pier	$l \leq 35m, h \leq 15m$	$(9,000 + 1,500h)(0.6 + 0.012l)$	0
	$l \leq 35m, h > 15m$	$(10,000 + 1,800h)(0.6 + 0.012l)$	
	$l > 35m$	$(12,000 + 2,000h)(0.4 + 0.022l)$	
foundation footing		$1,200l(0.4 + 0.03h)$	0
foundation pile set		$\alpha(0.3 + 0.13l)(0.7 + 0.015h)600d$	0
sloped pier	$l \leq 35m, h \leq 15m$	$\left\{ \begin{array}{l} (9,000 + 1,500h)(0.6 + 0.012l) \\ (10,000 + 1,800h)(0.6 + 0.012l) \\ (12,000 + 2,000h)(0.4 + 0.022l) \end{array} \right\}$	0
	$l \leq 35m, h > 15m$		
	$l > 35m$		
technical block		$+1,200l(0.4 + 0.03h) +$ $+\alpha(0.3 + 0.13l)(0.7 + 0.015h)600d +$ $+8(1/2h^3)$	
technical block		150,000	28
finishing		450 [/m]	.014[/m]
pre-work		0	42
pre-deck		0	28
			56
construction method	with staging with gantry	$4whL$ $120,000 + 25,000N$	0

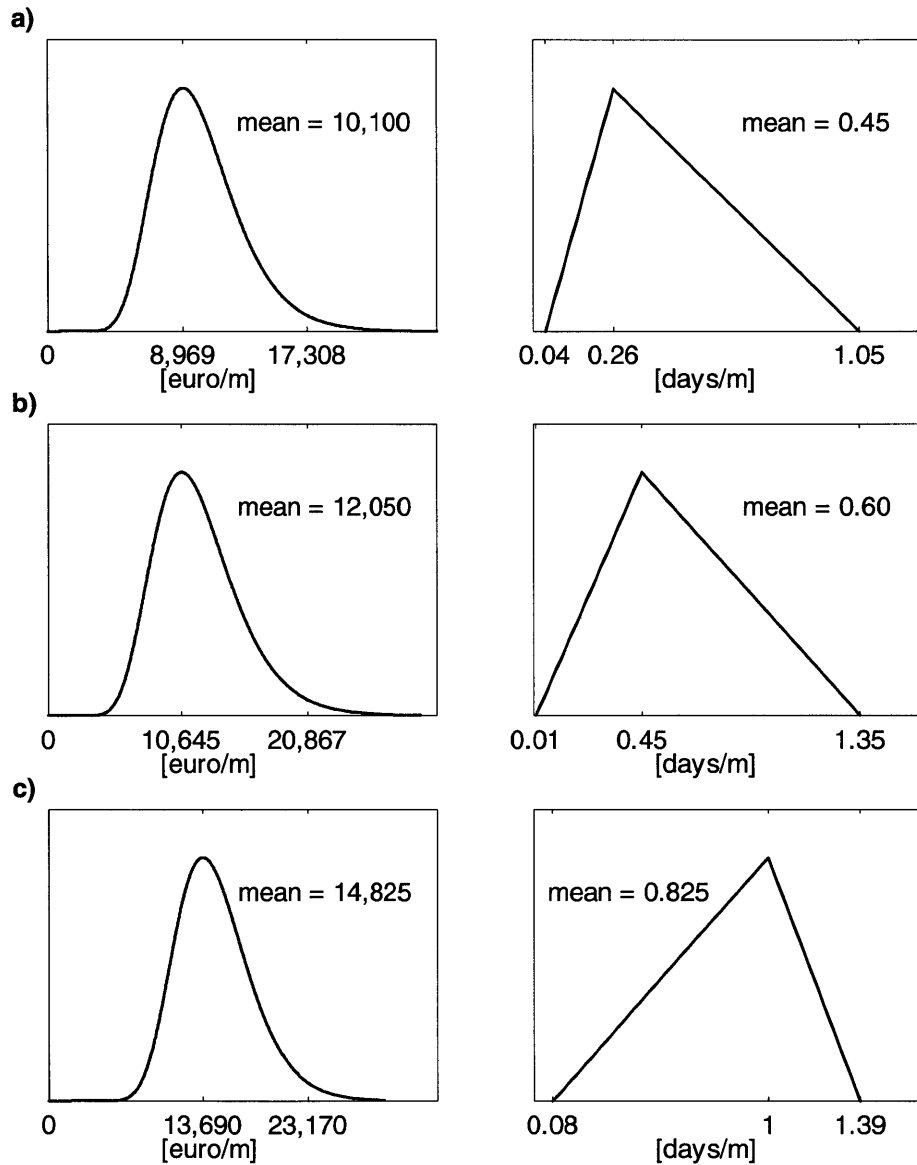


Figure 7-8: Probability distributions of the input cost per unit length (left) and the time per unit length (right) of tunnels in a) good, b) medium, and c) poor geology. Note that the mean input cost per unit length and the mean input time per unit length are equal to the deterministic input cost per unit length and the deterministic input time per unit length (Table 7.41).

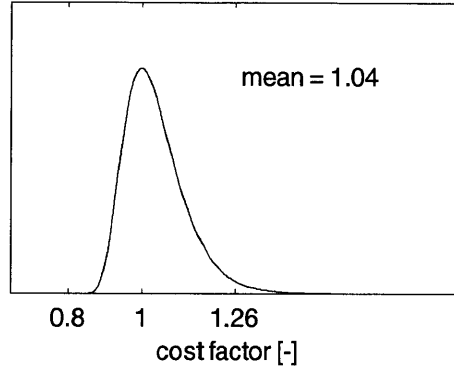


Figure 7-9: Probability distribution of the input cost factor of viaducts. Note that the mean input cost factor (1.04) is larger than the deterministic input cost factor of 1 (Table 7.42).

Table 7.43: Deterministic input costs per unit volume and deterministic input production rates for a sequence of cuts and embankments. Note that the deterministic input costs per unit volume are smaller than the mean input costs per unit volume in Figure 7-11, while the deterministic input production rates are larger than or equal to the mean input production rates in Figure 7-12.

Activity		cost per unit volume [euro/m ³]	production rate
clearing		1.92	20,234 [m ² /day]
excavating	mechanical	3.46	25,689 [m ³ /day]
	blasting	8.08	7,951 [m ³ /day]
improving		4.85	25,689 [m ³ /day]
filling		1.25	25,689 [m ³ /day]
capping		6.75	4,205 [m ³ /day]
sub-ballast		14.4	470 [m ³ /day]

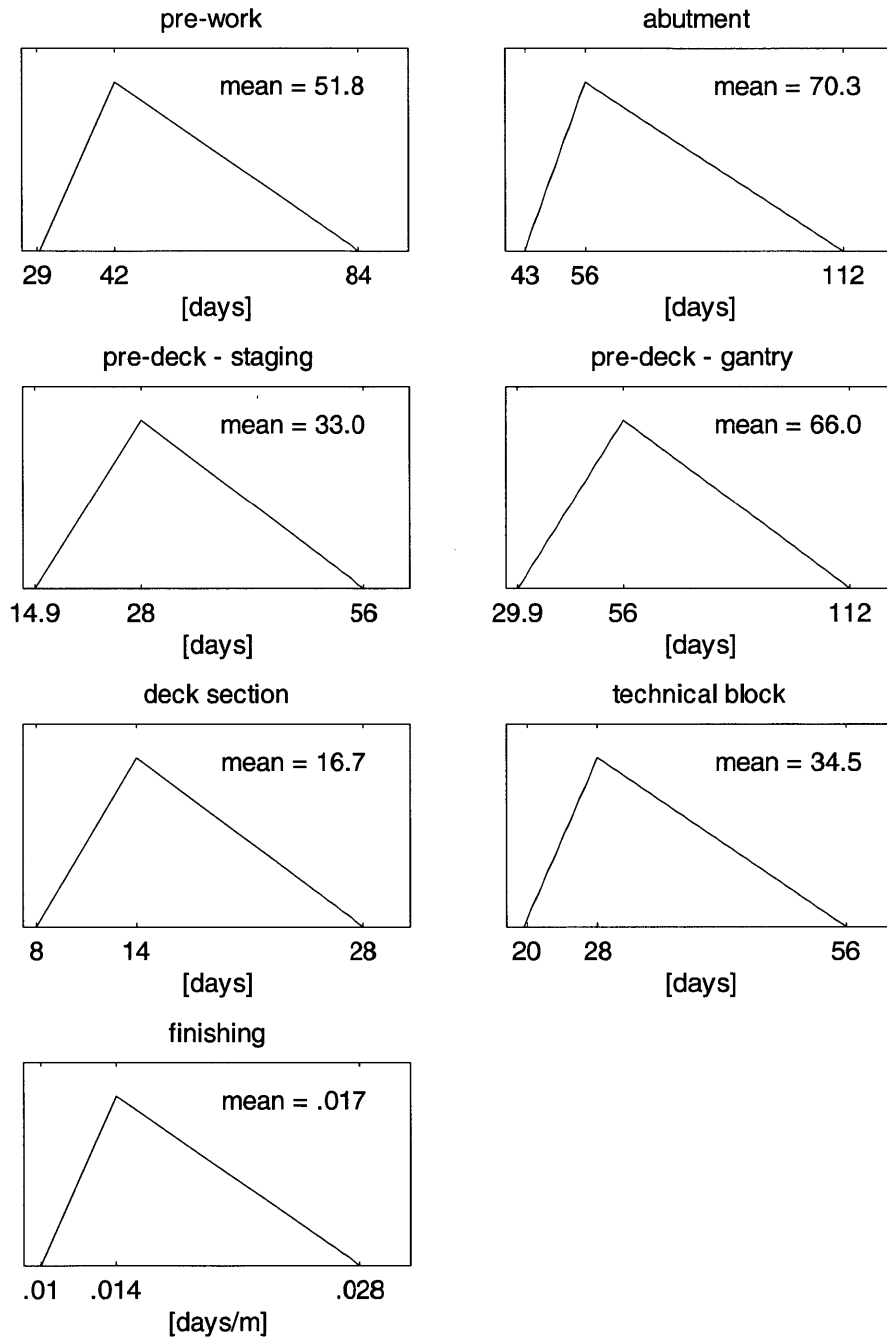


Figure 7-10: Probability distributions of input times of viaducts. Note that the mean input time is larger than the deterministic input time (Table 7.42).

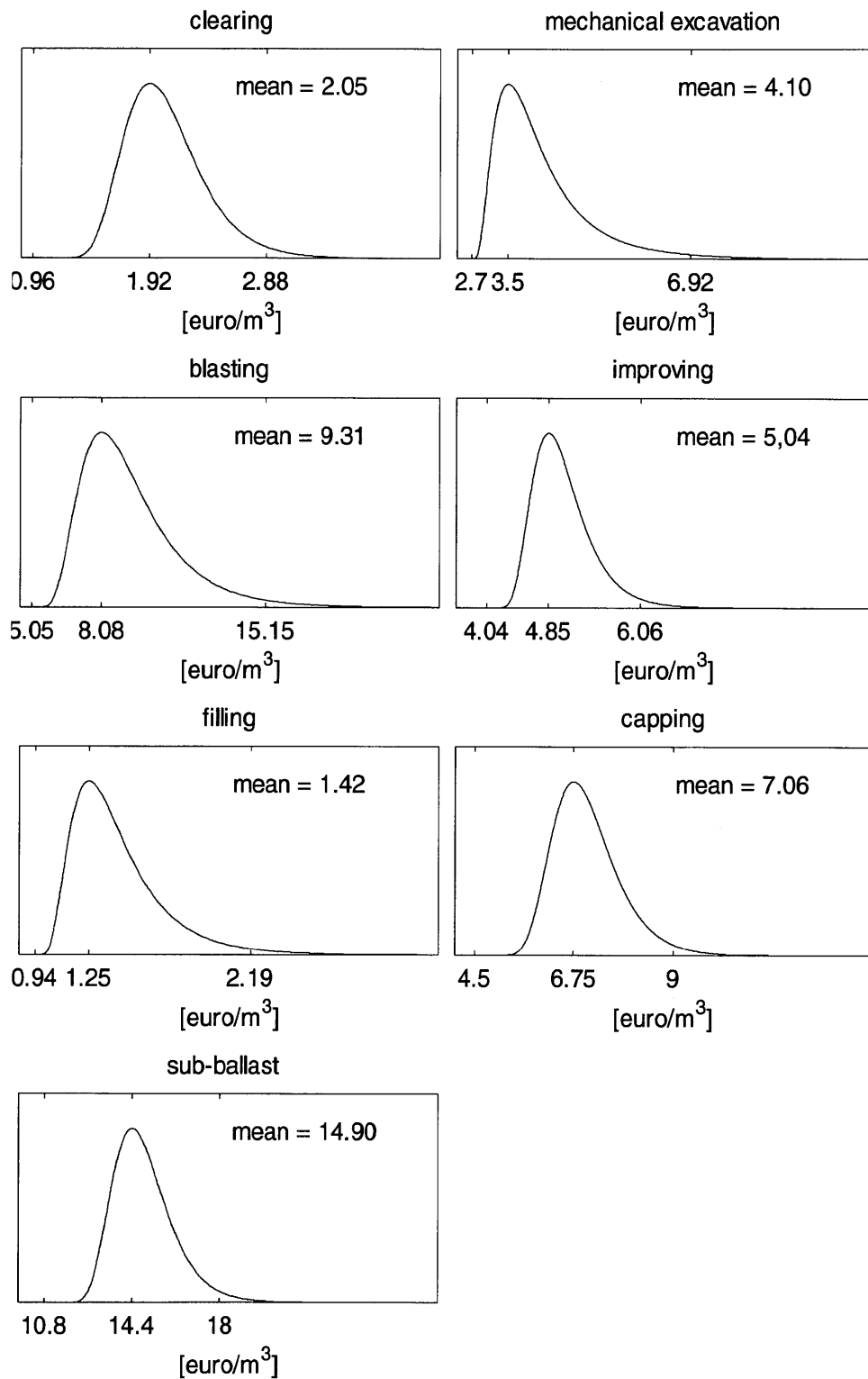


Figure 7-11: Probability distributions of the input costs per unit volume of cuts and embankments. Note that the mean input cost per unit volume is larger than the deterministic input cost per unit volume (Table 7.43).

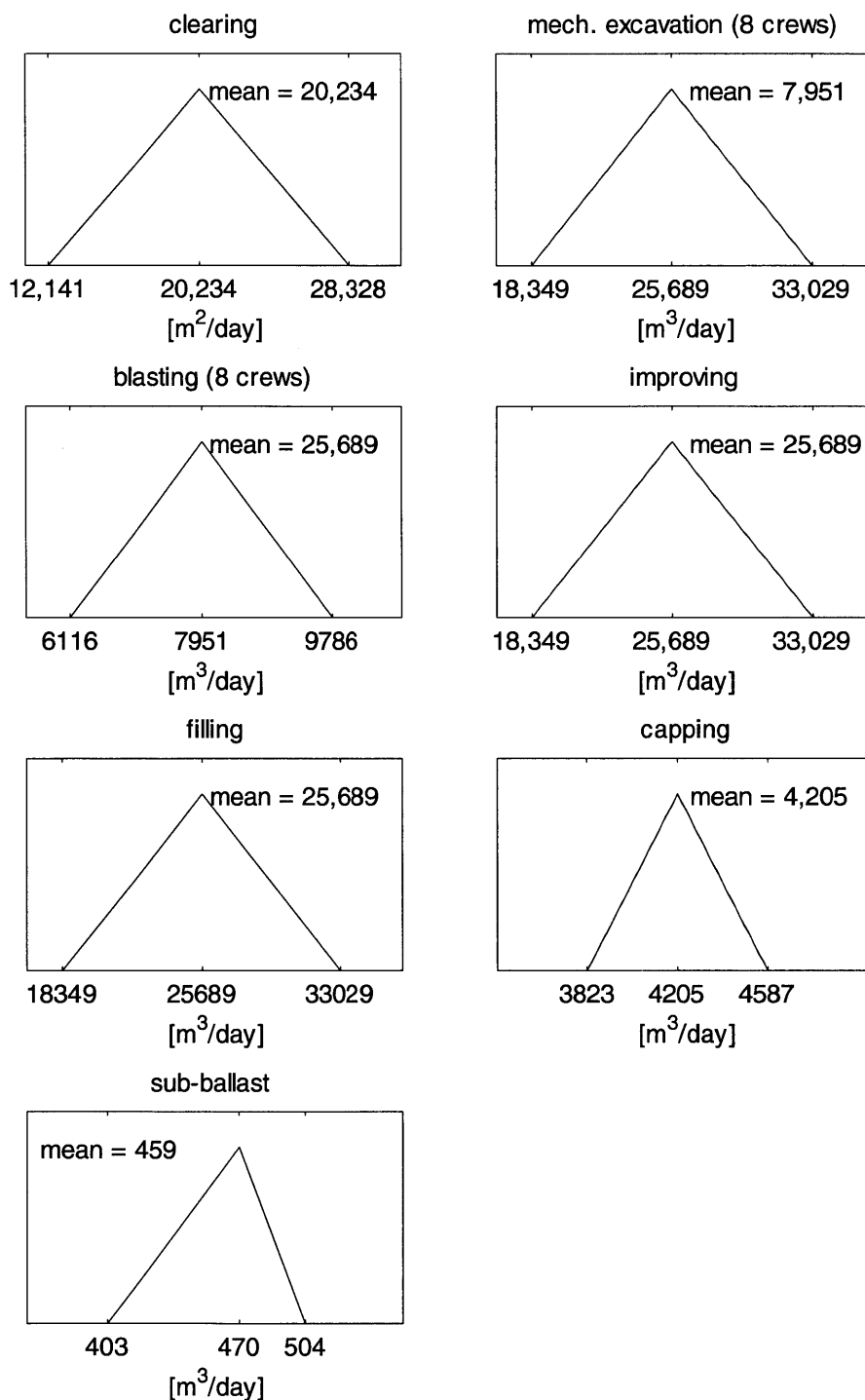


Figure 7-12: Probability distributions of the input production rates of cuts and embankments. Note that the mean input production rate is smaller than or equal to the deterministic input production rate (Table 7.43).

time distributions of the construction of tunnels, viaducts, cuts and embankments, and of all the structures are plotted in Figure 7-13. The deterministic total cost and total time, the means, the 90th percentiles, and the standard deviations of the total cost and total time distributions are presented in Table 7.44.

From the analysis of the simulation results and the comparison with the deterministic total cost and total time, the following has to be noted:

- While the deterministic total cost and total time are represented by a point (one total cost and one total time), the results from the simulation of the variability are a scattergram of points, one for each simulation run (Figure 7-13). The scattergram is the clear manifestation that the uncertainty in total cost and total time is caused by variability.
- For tunnels, the deterministic total cost and total time are located in the center of the scattergram, while for viaducts and cuts and embankments the deterministic total cost and total time lie outside the scattergram (Figure 7-13).
 - For tunnels (Figure 7-13a), the deterministic total cost and total time are equal to the mean total cost and the mean total time (Table 7.44), since the deterministic input cost and the deterministic input time per unit length are equal to the mean input cost and the mean input time per unit length (compare Table 7.41 with Figure 7-8).
 - For viaducts and cuts and embankments (Figures 7-13b and 7-13c), the deterministic total cost and total time are significantly smaller than the mean total cost and the mean total time (Table 7.44) for the following reasons:
 1. The deterministic total cost and total time are the sum of the deterministic input cost and time, which are equal to the modes of the input cost and time distributions. For viaducts, compare Table 7.42 with Figures 7-9 and 7-10. For cuts and embankments, compare Table 7.43 with Figures 7-11 and 7-12.
 2. The mean total cost and the mean total time are equal to the sum of the mean input costs and times, respectively (see e.g. Bertsekas & Tsitsiklis (2002)).
 3. Input cost distributions are skewed to the right, i.e. the mode input cost is smaller than the mean input cost (for viaducts: Figures 7-9; for cuts and embankments:

Figure 7-11). Thus, the sum of the mode input costs is smaller than the sum of the mean input costs.

4. Input time distributions are skewed to the right, i.e. the mode input time is smaller than the mean input time (for viaducts: Figures 7-9). Thus, the sum of the mode input times is smaller than the sum of the mean input times.
 5. Input production rate distributions (for cuts and embankments: Figure 7-12) are skewed to the left, i.e. the mode input production rate is larger than the mean input production rate. However, since the construction time is equal to the volume divided by the production rate, the sum of the construction times calculated with the mode input production rates is smaller than the sum of the construction times calculated with the mean input production rates.
- For tunnels, viaducts, cuts and embankments, and for all structures, the 90th percentiles of the total cost and the total time distributions are larger than the the mean total cost and the mean total time (Table 7.44). However, this difference is not particularly large in comparison with the differences caused by the other sources of uncertainty (sections 7.4.2 and 7.4.3).
 - When modeling the construction of all structures (tunnels, viaducts, cuts and embankments) (Figure 7-13d), there are three important observations on the total cost and total time:
 - The deterministic total cost is significantly smaller than the mean total cost (Table 7.44) for the following reasons:
 1. The deterministic/mean total cost of all structures is equal to the sum of the deterministic/mean total costs of tunnels, viaducts, cuts and embankments.
 2. The deterministic total costs of viaducts and of cuts and embankments are smaller than the respective mean total costs.
 - The deterministic total time is equal to the mean total time (Table 7.44) for the following reasons:
 1. The deterministic/mean total time of all structures is equal to the sum of the deterministic/mean times of tunnels, viaducts, cuts and embankments on the critical path. In the alignment, tunnels are constructed in sequence, viaducts are constructed in sequence, and cuts and embankments are constructed in sequence while

the tunnel construction is in parallel with the viaduct construction and the cuts and embankments construction. Since the tunnels require the longest time to be constructed, they determine the critical path. Thus, the deterministic/mean total time of all structures is equal to the deterministic/mean total time of the tunnels.

2. The deterministic total time of the tunnels is equal to the mean total time of the tunnels.

Table 7.44: Alignment A-a. Deterministic total cost and total time and means, 90th percentiles, and standard deviations of the total cost and total time distributions for the construction of the tunnels, the viaducts, the cuts and embankments, and of all the structures. The variability in the construction process is responsible for the range of possible total cost and total time.

Alignment A-a	tunnels		viaducts	
	no uncertainty	variability	no uncertainty	variability
total cost [10 ⁶ euro]				
deterministic	128.7	–	116.1	–
mean	–	128.8	–	120.7
90 th percentile	–	129.2	–	121.4
st. dev.	–	0.3	–	0.6
total time [days]				
deterministic	6,589	–	3,590	–
mean	–	6,588	–	4,329
90 th percentile	–	6,621	–	4,432
st. dev.	–	26	–	81
	cuts & embankments		all structures	
	no uncertainty	variability	no uncertainty	variability
total cost [10 ⁶ euro]				
deterministic	63.4	–	308.2	–
mean	–	73.0	–	322.4
90 th percentile	–	73.1	–	323.2
st. dev.	–	0.1	–	0.7
total time [days]				
deterministic	984	–	6,589	–
mean	–	998	–	6,588
90 th percentile	–	999	–	6,621
st. dev.	–	1	–	26

From the findings described above, the following conclusions can be drawn:

- The cost and the time of a project are uncertain. This uncertainty cannot be expressed with a single number estimate (deterministic total cost and total time) but rather with a range of

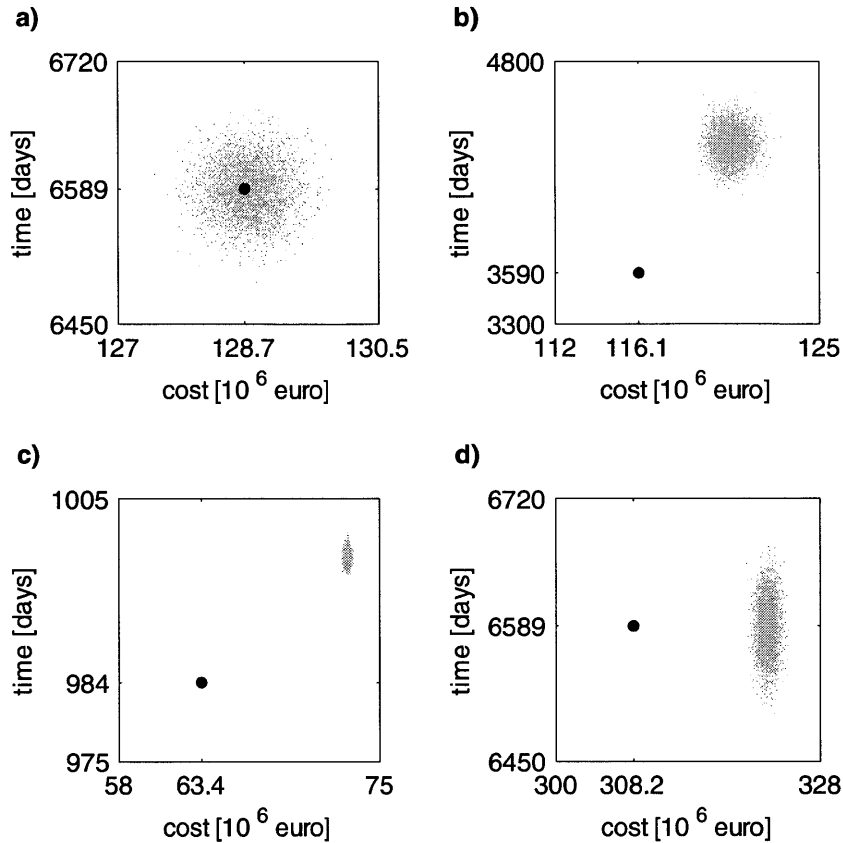


Figure 7-13: Alignment A-a, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. The scattergrams make evident the uncertainty in the construction cost and time caused by variability of input parameters. The deterministic total cost and total time (black dot) of the tunnels (a) is located in the center of the scattergram since the deterministic input cost and time are equal to the mean input cost and time. The deterministic total cost and total time of the viaducts (b) and of the cuts and embankments (c) lie below the respective scattergrams since the deterministic input cost and time are equal to the mode input cost and time. The deterministic total cost of all structures (d) is smaller than the mean total cost since it is equal to the sum of the deterministic total costs of tunnels, viaducts, cuts and embankments, while the deterministic total time of all structures is equal to the mean total time of the tunnels since these determine the critical path.

possible outcomes.

- The practice of calculating the deterministic total cost and total time with the mode (most probable) input cost and time (Molenaar et al. 2010) is problematic: the deterministic total cost and total time calculated with the mode input cost and time are smaller than the range of possible outcomes. The deterministic total cost and total time should be calculated with the mean input cost and time.
- The range of the total cost can be reduced by decreasing the variability of the input costs, while the range of the total time can be reduced by decreasing the variability of the input times on the critical path.

The findings and the conclusions are obtained with the analysis of one of the four alignments (Alignment A-a) of the project. However, similar conclusions are reached in the other three alignments (Appendix D, section D.1).

7.4.2 Cost Correlations

In this section, the impact on total cost and total time of cost correlations is analyzed. First the input data are presented, then the impact is discussed.

The construction of alignment A-a is simulated first considering one source of uncertainty, the variability in the construction process, and then considering two sources of uncertainty, the variability and the correlations between the costs of the activities. The results from the two simulations are compared. The input to model the variability was summarized in the previous section. Cost correlations, already presented in section 7.2, are summarized in Table 7.45.

Table 7.45: Correlations between the costs of repeated activities.

	Tunnels	viaducts	cuts & embankments
correlation coefficient 1, ρ_1	0.5	–	0.5
number of rows and columns, N , with ρ_1	50	–	50
correlation coefficient 2, ρ_2	0.9	–	0.9

The scattergrams of the simulations of the construction of tunnels, viaducts, cuts and embankments, and of all the structures are plotted in Figure 7-14. The means, the 90th percentiles, the standard deviations of the simulation results are presented in Table 7.46.

From the analysis of the simulation results, the following has to be noted:

- The range of the total cost increases dramatically from the simulation modeling variability to the simulation modeling variability and cost correlations (with the exception of the viaducts; see below). This is confirmed by the following observations:
 - The increase in total cost is clearly visible in the scattergrams (Figure 7-14): on the cost axis the gray clouds are more spread than the black clouds.
 - The standard deviation of the total cost increases by more than one order of magnitude from the simulation modeling variability to the simulation modeling variability and cost correlations (Table 7.46).
 - The 90th percentiles of the total cost are significantly larger than the the mean total costs in the simulation modeling variability and cost correlations (Table 7.46).
- For viaducts, the total cost and the total time distributions for the simulation modeling variability are the same as for the simulation modeling variability and cost correlations since no cost correlations are modeled (section 7.2.2).
- The means of the total cost remain constant (Table 7.46): correlations impact the standard deviation of the sum of correlated costs, while they do not affect the mean.
- The ranges, the means, the 90th percentiles and the standard deviations of the total time remain constant (Table 7.46), which is confirmed by the constant spread of the scattergrams on the time axis (Figure 7-14). In fact, times are assumed independent, i.e. the total time distribution of the simulation modeling variability is the same as the total time distribution of the simulation modeling variability and cost correlations.

From the findings described above, the following conclusions can be drawn:

- Due to cost correlations, the range of the total cost increases dramatically. If cost correlations are disregarded, the range of the possible total cost is underestimated.
- Cost correlations feature a threat (uncertainty with negative outcome) and an opportunity (uncertainty with positive outcome) aspect. In fact, the cost correlations cause the range of the total cost to increase on both sides of the mean: there is the threat of a total cost significantly larger than the mean total cost as well as the opportunity of a total cost significantly smaller than the mean total cost.

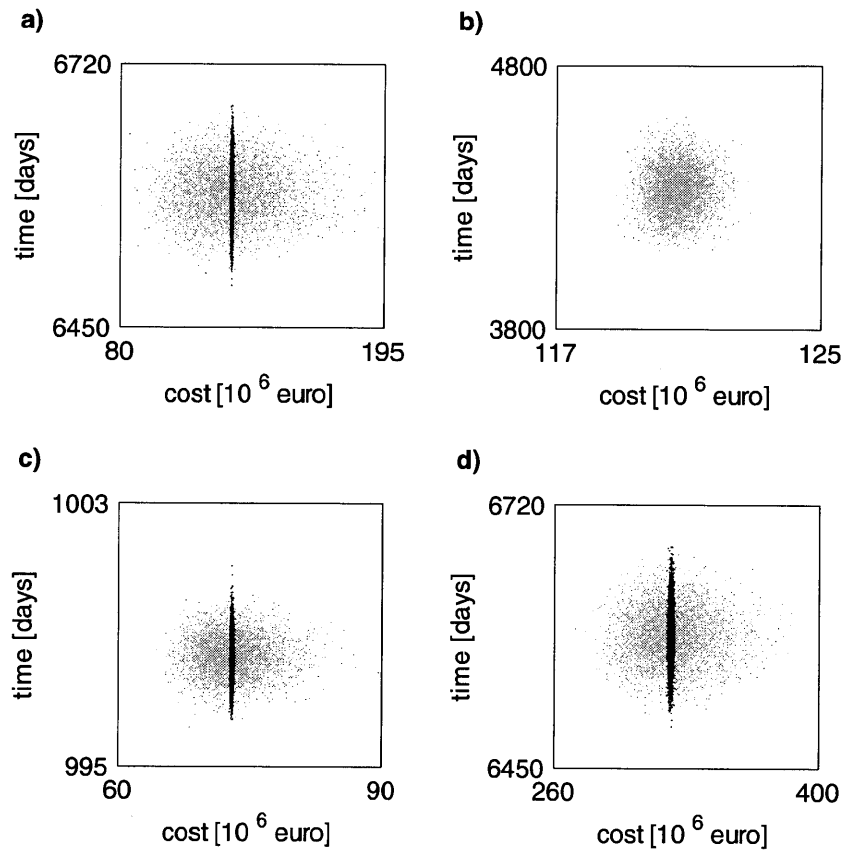


Figure 7-14: Alignment A-a, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. Due to cost correlations, the range of the total cost increases dramatically.

- The total cost can be reduced in two ways:
 1. by eliminating the cost correlations, or by reducing their magnitude,
 2. by exploiting the opportunity aspect of cost correlations, i.e. by achieving a total cost smaller than the mean total cost.

The findings and the conclusions are obtained with the analysis of one of the four alignments (Alignment A-a) of the project. However, similar conclusions are reached in the other three alignments (Appendix D, section D.2).

In the preceding, the impact on total cost of cost correlations has been discussed. In the next

Table 7.46: Alignment A-a, uncertainty sources: variability versus variability and cost correlations. Means, 90th percentiles, and standard deviations of the total cost and total time for the construction of the tunnels, the viaducts, the cuts and embankments, and of all the structures. Due to cost correlations, the standard deviations of the total cost increase by an order of magnitude (in viaduct construction, no correlation is modeled).

Alignment A-a	tunnels		viaducts	
	variability	variability and cost correlations	variability	variability and cost correlations
total cost [10 ⁶ euro]				
mean	128.8	128.8	120.7	120.7
90 th percentile	129.2	149.4	121.4	121.4
st. dev.	0.3	15.8	0.6	0.6
total time [days]				
mean	6,588	6,587	4,329	4,329
90 th percentile	6,621	6,620	4,432	4,432
st. dev.	26	26	81	81
	cuts & embankments		all structures	
	variability	variability and cost correlations	variability	variability and cost correlations
total cost [10 ⁶ euro]				
mean	73.0	73.0	322.4	322.5
90 th percentile	73.1	76.9	323.2	343.6
st. dev.	0.1	3.0	0.7	16.1
total time [days]				
mean	998	998	6,588	6,587
90 th percentile	999	999	6,621	6,620
st. dev.	1	1	26	26

section, the impact of disruptive events is analyzed.

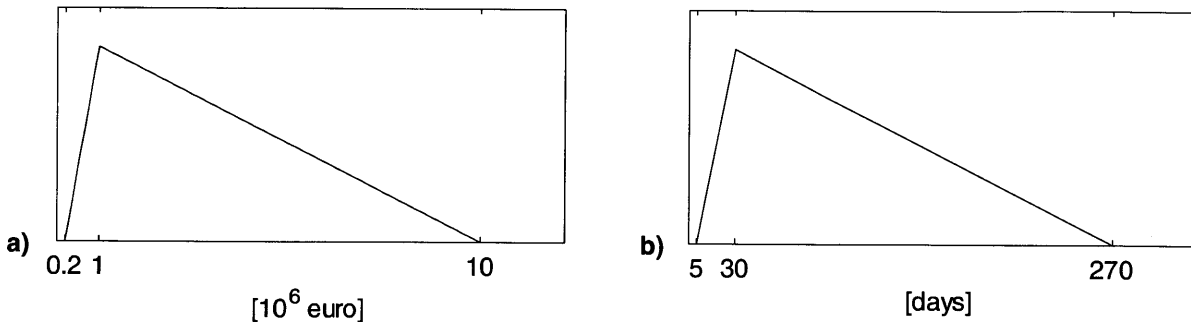
7.4.3 Disruptive Events

In this section, the impact on total cost and total time of disruptive events is analyzed. First the input data are presented, then the impact is discussed.

The construction of alignment A-a has been simulated first by considering two sources of uncertainty, the variability in the construction process and the correlations between the costs of the activities, and then considering three sources of uncertainty, the variability, the cost correlations and the disruptive events. The results from the two simulations are compared. The input data to model the variability and the cost correlations were summarized in sections 7.4.1 and 7.4.2, respectively. The input to model the disruptive events, already presented in section 7.2, is summarized

in Figures 7-15 (tunnels), 7-16 (viaducts), and 7-17 (cuts and embankments).

Cave-in $p = 1/800m$



Water inflow $p = 1/500m$

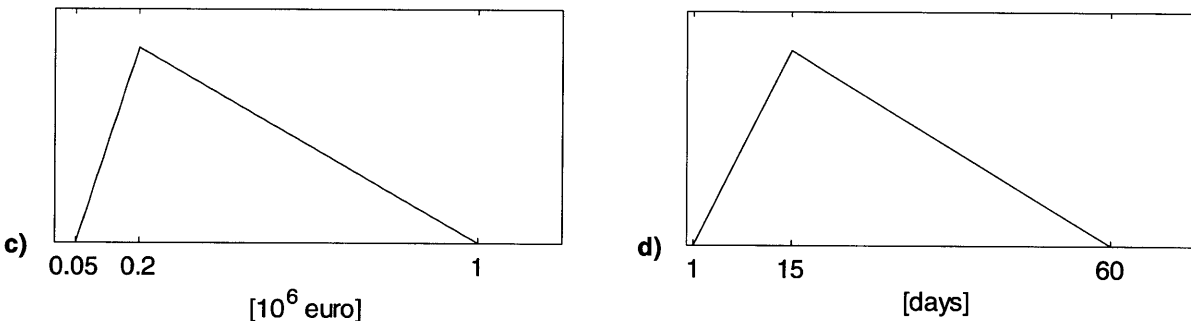


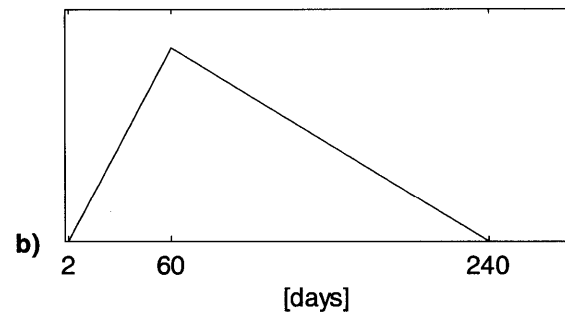
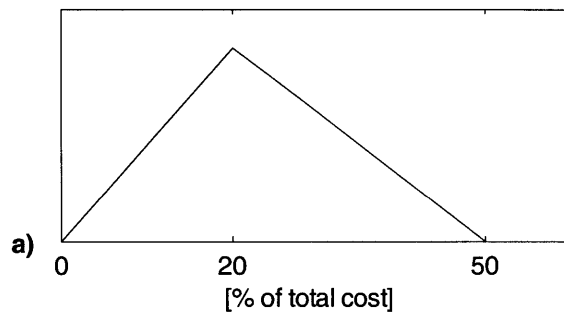
Figure 7-15: Disruptive events in tunnel construction. Cave-in: probability of occurrence and probability distributions of a) the cost impact and b) the time impact. Water inflow: probability of occurrence and probability distributions of c) the cost impact and d) the time impact.

The scattergrams of the simulations of the construction of tunnels, viaducts, cuts and embankments, and of all the structures are plotted in Figure 7-18. The means, the 90th percentiles, the standard deviations of the simulation results are presented in Table 7.47.

From the analysis of the simulation results, the following was observed:

- The ranges of the total cost and total time explode from the simulation modeling variability and cost correlations to the simulation modeling variability, cost correlations and disruptive events. This is confirmed by the following observations:
 - The increase in total cost and total time is clearly visible in the scattergrams (Figure 7-18): the gray clouds are more spread than the black clouds and represent larger total cost and/or total time.

Differing ground conditions $p = 5\%$



Construction accident/problem $p = 1\%$

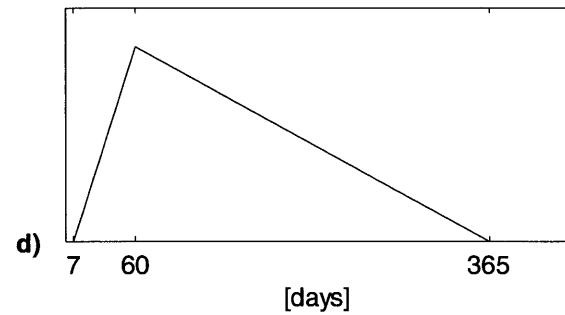
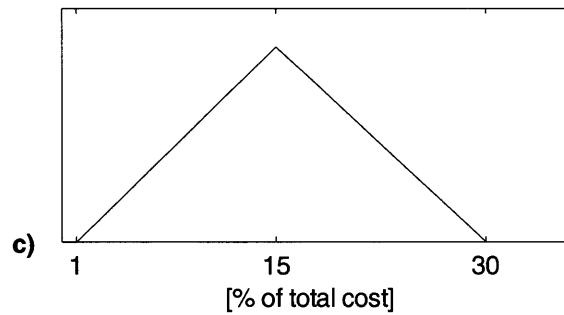
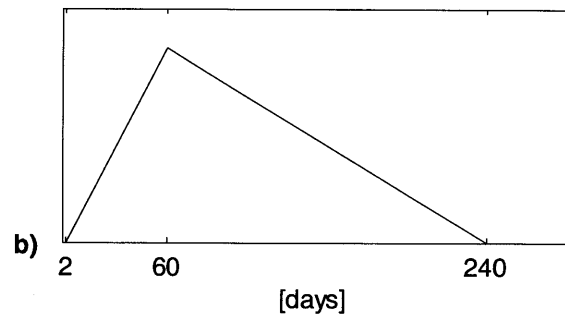
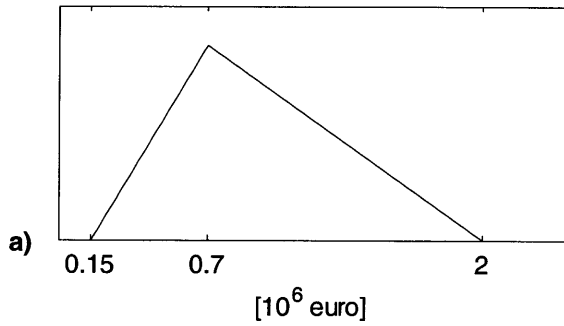


Figure 7-16: Disruptive events in viaduct construction. Differing ground conditions: probability of occurrence and probability distributions of a) the cost impact and b) the time impact. Construction accident/problem: probability of occurrence and probability distributions of c) the cost impact and d) the time impact.

Flooding $p = 1/\text{year}$



Differing site conditions $p = 50\%$

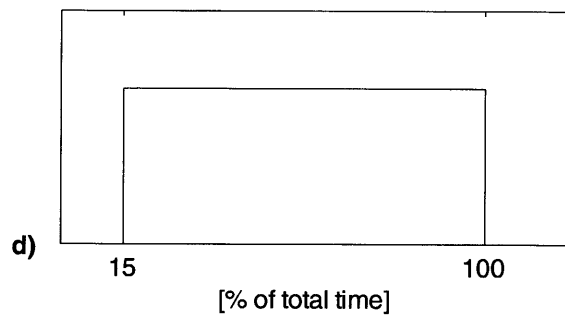
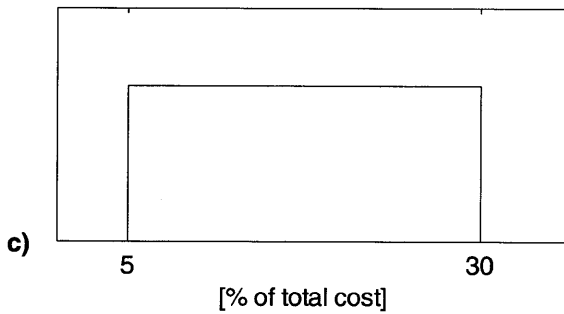


Figure 7-17: Disruptive events in earthwork (cuts and embankments) construction. Flooding: probability of occurrence and probability distributions of a) the cost impact and b) the time impact. Differing site conditions: probability of occurrence and probability distributions of c) the cost impact and d) the time impact.

- The means, the 90th percentiles, and the standard deviations of the total cost and total time increase significantly (Table 7.47).
- The ranges of the total cost and the total time increase with different magnitudes and patterns (Figure 7-18) depending on the structures analyzed. Specifically:
 - For tunnels (Figure 7-18a), two clouds are observed: the black cloud is the simulation modeling variability and cost correlations, while the gray cloud is the simulation modeling variability, cost correlations and disruptive events. Two separate clouds indicate that:
 1. In all simulation runs modeling disruptive events, one or more disruptive events occur during the construction of the tunnels;
 2. The time impact can be truly disruptive on the tunnel construction process: in fact, the mean time impact of one disruptive event, the cave-in, is just above 100 days (Figure 7-15b), which is five times larger than the difference of 20 days between the total time mean and the 90th percentile in the simulation modeling variability and cost correlations (Table 7.47).
 3. The cost impact is also very large, since it can reach 10 million euro for the disruptive event cave-in (Figure 7-15a). However, this does not appear as disruptive since it is a third of the difference of 30 million euro between the mean total cost and the 90th percentile in the simulation modeling variability and cost correlations.
 - For viaducts (Figure 7-18b), the black cloud (simulation modeling variability and cost correlations) and the gray cloud (simulation modeling variability, cost correlations and disruptive events) are superposed for the following reasons:
 1. Disruptive events do not occur in every simulation run: in the construction of a viaduct, the probability of occurrence of the two disruptive events (differing ground conditions and a construction accident or problem) is 5% and 1%, respectively (Figure 7-16).
 2. The largest time impact of a disruptive event (construction accident or problem) is 365 days (Figure 7-16d), which is an order of magnitude smaller than the minimum total time of the simulation modeling variability and cost correlations (Figure 7-18b).

3. On the other hand, the cost impact of the disruptive events appears truly disruptive (Figure 7-18b). In fact, it can reach 30% and 50% of the total cost (Figures 7-16a and 7-16c).
- For cuts and embankments (Figure 7-18c), three clouds are visible: the black cloud of the simulation modeling variability and cost correlations, a lower gray cloud of the first disruptive event, and an upper gray cloud of the second disruptive event. The three clouds show:
 1. The truly disruptive nature of the events and the different impact of these.
 2. The disruptive event "flooding" is responsible for the lower gray cloud (7-18c): its impact is limited to 2 million euro and 240 days (Figures 7-17a and 7-17b) compared to a mean total cost of 73 million euro and a mean total time of approximately 1,000 days in the simulation modeling variability and cost correlations (Table 7.47). Thus, due to flooding the total cost and the total time increase (lower gray cloud in Figure 7-18c) but not as much as for the disruptive event differing site conditions (upper gray cloud).
 3. The disruptive event "differing site conditions" is responsible for the upper gray cloud (7-18c). Its time impact can reach 30% of the total time while its cost impact can reach 100% of the cost time (Figures 7-17c and 7-17d).
 4. The different shapes of the upper and lower gray clouds are due to the different input distributions of the cost and time impacts of the disruptive events. While these are triangular for the disruptive event "flooding" (Figures 7-17a and 7-17b), they are uniform for the disruptive events "differing site conditions" (Figures 7-17c and 7-17d).
 5. If both disruptive events occur in a simulation run, the impact of the differing site conditions is larger than the impact of the flooding, i.e. the data point appears in the upper gray cloud.
 - For all structures (Figure 7-18d), similar patterns as for the tunnels (Figure 7-18a) are observed for the following reasons:
 1. Tunnels are constructed in parallel with the viaducts and the sequence of cuts and

embankments. Since the tunnel construction time is the longest, the tunnels determine the construction time of all structures.

2. The total cost for all structures is equal to the sum of all costs. However, the tunnels are the main cost driver (Table 7.47).

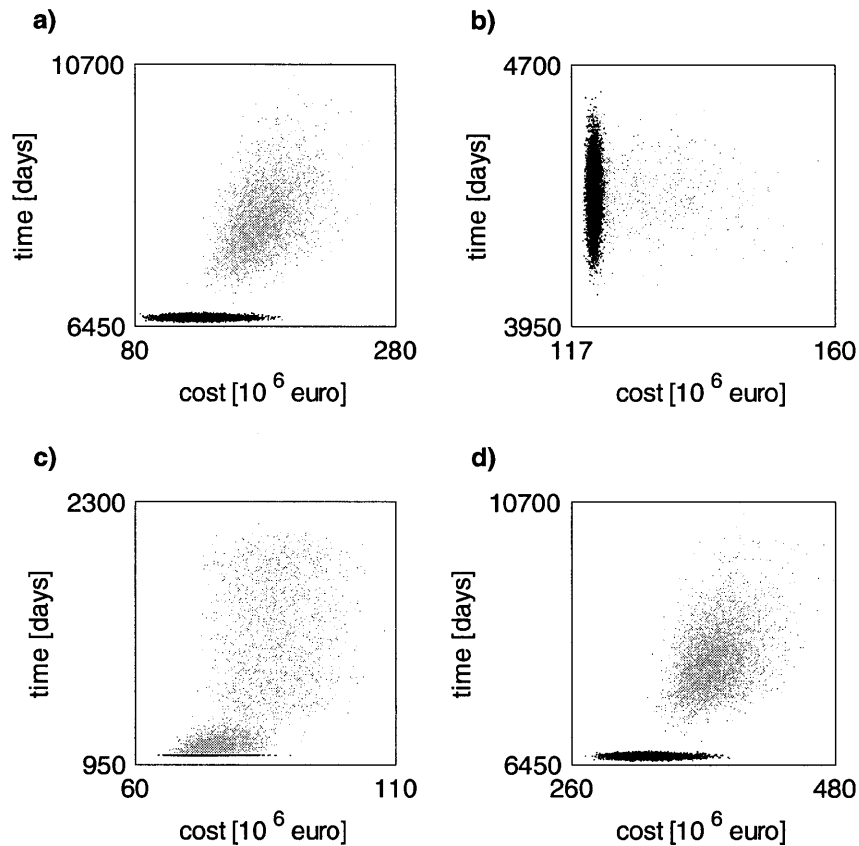


Figure 7-18: Alignment A-a, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. Due to the disruptive events, the ranges of total cost and total time increase significantly.

From the findings described above, the following conclusions can be drawn:

- Due to the disruptive events, the ranges of total cost and total time increase significantly. If disruptive events are not considered, the ranges of the possible total cost and total time are underestimated.

Table 7.47: Alignment A-a, uncertainty sources: variability and cost correlations versus variability, cost correlations and disruptive events. Means, 90th percentiles, and standard deviations of the total cost and total time for the construction of the tunnels, the viaducts, the cuts and embankments, and of all the structures. Due to the disruptive events, the means, the 90th percentiles and the standard deviations of total cost and total time increase significantly.

Alignment A-a	tunnels		viaducts	
	variability, cost correlations	variability, cost correlations, disruptive events	variability, cost correlations	variability, cost correlations, disruptive events
total cost [10 ⁶ euro]				
mean	128.8	177.6	120.7	122.2
90 th percentile	149.4	203.0	121.4	124.5
st. dev.	15.8	19.0	0.6	4.4
total time [days]				
mean	6,587	8,180	4,329	4,330
90 th percentile	6,620	8,857	4,432	4,433
st. dev.	26	506	81	81
	cuts & embankments		all structures	
	variability, cost correlations	variability, cost correlations, disruptive events	variability, cost correlations	variability, cost correlations, disruptive events
total cost [10 ⁶ euro]				
mean	73.0	81.3	322.5	381.0
90 th percentile	76.9	92.2	343.6	408.4
st. dev.	3.0	7.5	16.1	20.7
total time [days]				
mean	998	1,350	6,587	8,180
90 th percentile	999	1,913	6,620	8,857
st. dev.	1	351	26	506

- Although all disruptive events presented here are threats (uncertainty with negative outcome), disruptive events can also be opportunities, e.g. the delivery of more resistant disks for the tunnel boring machine (section 4.1). However, the tunnel, viaduct and earthwork (cuts and embankments) experts focused on disruptive events with negative outcomes.
- The total cost and the total time can be reduced by minimizing the impact of disruptive events with negative outcome and maximizing the impact of disruptive events with positive outcome.

The findings and the conclusions are obtained with the analysis of one of the four alignments (Alignment a-a) of the project. However, similar conclusions are reached in the other three align-

ments (Appendix D, section D.3).

In this section, the impact on total cost and total time of the last source of uncertainty, the disruptive events, has been discussed.

7.5 Parallel Versus Sequential Construction

In this section the construction of all the structures is modeled in parallel differently from the rest of the chapter. The simulation results are compared with previous results.

Throughout this chapter, tunnels were constructed in sequence, viaducts were constructed in sequence, and cuts and embankments were constructed in sequence while the tunnel construction was in parallel with the viaduct construction and the cuts and embankments construction. As a consequence, the total construction time is relatively large: for example, in the simulation modeling the three sources of uncertainty, the construction of all structures in Alignment A-a the mean total time is larger than 8,000 days, the equivalent of 22 years (Table 7.47).

Another reasonable assumption is: all structures are constructed in parallel, that is all tunnels, all viaducts and the sequence of cuts and embankments are constructed in parallel. If the total cost and the total time to construct the structures in sequence (Figure 7-19, left) and in parallel (Figure 7-19, right) are compared, one finds:

- The total costs remain the same, since the total cost is the sum of all the costs, which by first approximation are equal for the construction in sequence and in parallel.
- The total times decrease significantly due to the shortened critical path:
 - Tunnels (Figure 7-19a): For the construction in sequence the total time is the time needed to construct all the tunnels one after the other, whereas for the construction in parallel the total time is equal to the longest tunnel construction.
 - Viaducts (Figure 7-19b): Similarly to the tunnels, for the construction in sequence the total time is the time needed to construct all the viaducts one after the other, whereas for the construction in parallel the total time is equal to the longest viaduct construction.
 - Sequence of cuts and embankments (Figure 7-19c): The total time does not change since the construction in sequence is the same as the construction in parallel. In fact, cuts

- and embankments are constructed as a sequence of cuts and embankments (chapter 3). Thus, the sequence of cuts and embankments is the only structure to be constructed.
- All the structures (Figure 7-19d): For the construction in sequence the total time of all structures is equal to the total time of the tunnels since this is larger than the total time of the viaducts and of the sequence of cuts and embankments. For the construction in parallel, the total time is equal to the longest construction time among all the structures. In Alignment A-a, a tunnel requires the longest construction time. Thus, for the construction in parallel the total time of all the structures (Figure 7-19d, right) is equal to the total time of the tunnels (Figure 7-19a, right).
 - Interestingly, the pattern of the cost-time scattergrams described in section 7.4.3 (modeling variability, cost correlations and disruptive events) do not vary substantially between the simulation of the construction in sequence (Figure 7-19, left) and the simulation of the construction in parallel (Figure 7-19, right).

In the preceding, the construction in parallel of all the structures has been modeled and compared with the construction in sequence.

7.6 Value of the Construction and Uncertainty Models: Capturing the Impact of Uncertainty on Total Cost and Total Time

In this section, the deterministic total cost and total time and the 90th percentiles of the total cost and total time distributions of alignment A-a are compared to show that the impacts of the sources of uncertainty on the total cost and total time are significant.

The 90th percentile of the total cost and total time distributions is the measure compared with the deterministic total cost and total time for three reasons. First, the 90th percentile is located in the upper tail of the distribution so that it covers a large part of the probability distribution. Second, the 90th percentile implies that there is the chance of one cost overrun or one project delay in ten projects, which is assumed to be acceptable. Third, the 90th percentile has been used successfully by the Washington State Department of Transportation to evaluate projects and to set the project budget (Reilly et al. 2004).

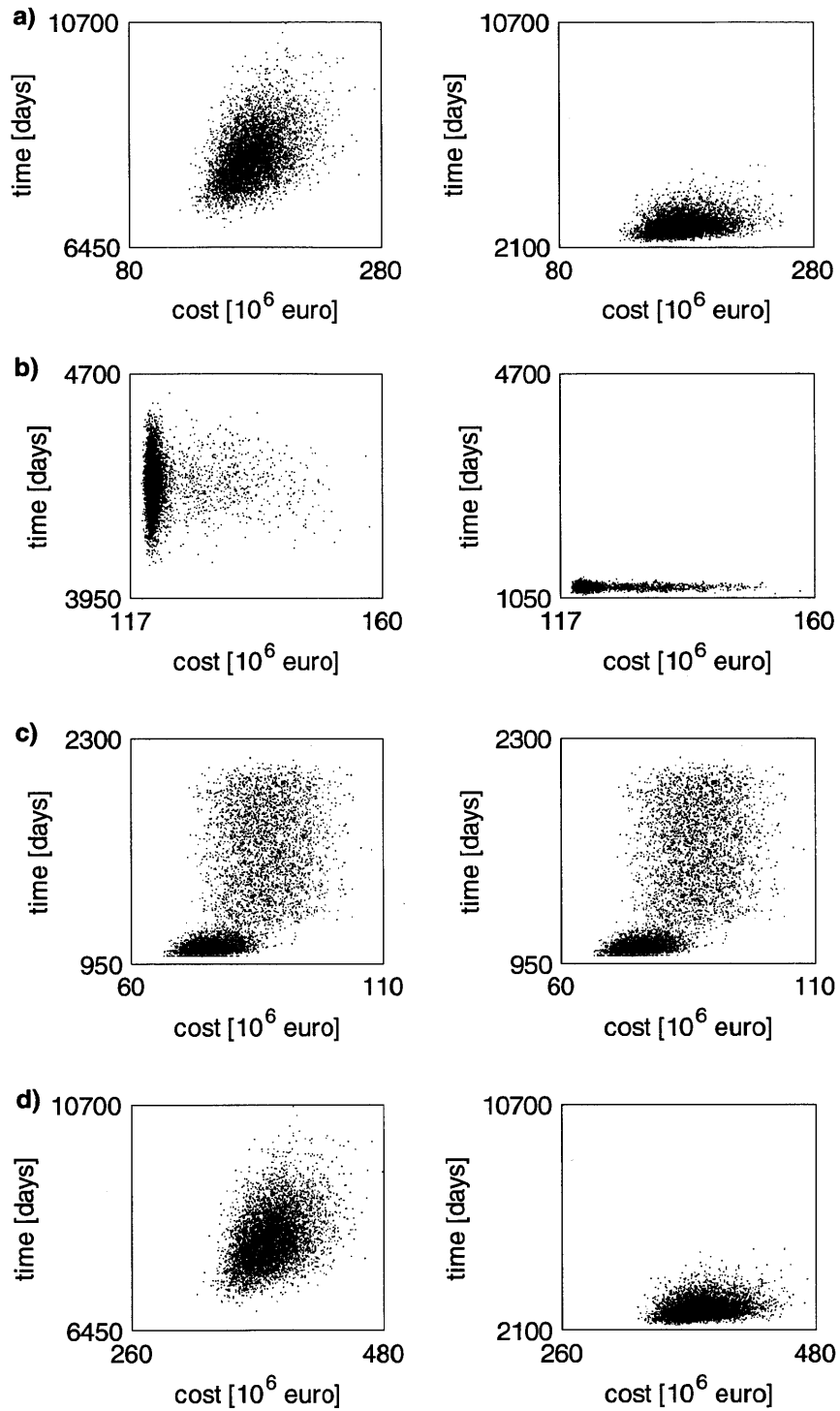


Figure 7-19: Alignment A-a: comparison of total cost and total time for the construction in sequence (left) and in parallel (right) of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures. While the total costs remain the same, the total times decrease significantly due to the shortened critical paths.

The increase in total cost and total time due to the three sources of uncertainty (variability, cost correlations, and disruptive events) in the construction of the structures of the rail line (tunnels, viaducts, cuts and embankments, and all structures) is quantified by comparing the 90th percentiles of the total cost and total time distributions with the deterministic total cost and total time. The increase is quantified with the following expression:

$$increase = \frac{P_{90} - D}{D}, \quad (7.13)$$

where P_{90} is the 90th percentile of the total cost/time distribution and D is the deterministic total cost/time. In order to read Table 7.48 and Figures 7-20 and 7-21 (below) correctly, note that the increases [%] and the bar lengths represent the increase from the deterministic total cost/time. Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events (Figures 7-20 and 7-21).

The deterministic total cost and total time, the 90th percentiles of the total cost and total time distributions, and the increases in total cost and total time are summarized in Table 7.48. The increases in total cost and total time are displayed in Figures 7-20 and 7-21 for the different structures and the sources of uncertainty.

Clearly, modeling all sources of uncertainty (variability, cost correlations, and disruptive events) causes the largest increase from the deterministic total cost and total time to the 90th percentiles of the total cost and total time distributions. For the construction of all structures, the increase of the total cost and total time from the deterministic estimate to the respective 90th percentile is 30% – 35% (Figure 7-20 and Table 7.48). The largest increases in total cost are observed for the construction of the tunnels and the construction of the cuts and embankments: for tunnels the 90th percentile of the total cost distributions is almost 60% larger than the deterministic total cost, while for cuts and embankments it is over 45% larger than the deterministic total cost (Figure 7-20 and Table 7.48). The largest increase in total time occurs in the construction of the cuts and embankments, where the total time increases almost 100%. This large increase is related to the two clouds representing the total cost and total time distributions of tunnels and of cuts and embankments presented in Figure 7-18.

Further insight in the results is gained by displaying the data differently: plotting the increase of the total cost and total time for the structures as a function of the source of uncertainty (Figure 7-

Table 7.48: Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) in alignment A-a depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 58% in tunnel construction, while the largest total time increase is 94% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations.

Total cost	tunnels		viaducts		cuts and embankments		all structures	
	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]
deterministic	128.7	–	116.1	–	63.4	–	308.2	–
90 th percentile variability	129.2	0.4	121.4	4.6	73.1	15.3	323.2	4.9
variability and cost correlations	149.4	16.1	121.4	4.6	76.9	21.3	343.6	11.5
variability, cost correlations and disruptive events	203.0	57.7	124.5	7.2	92.2	45.5	408.4	32.5
total time	tunnels		viaducts		cuts and embankments		all structures	
	[days]	[%]	[days]	[%]	[days]	[%]	[days]	[%]
deterministic	6,589	–	3590	–	984	–	6,589	–
90 th percentile variability	6,621	0.5	4,432	23.5	999	1.5	6,621	0.5
variability and cost correlations	6,620	0.5	4,432	23.5	999	1.5	6,620	0.5
variability, cost correlations and disruptive events	8,857	34.4	4,433	23.5	1,913	94.4	8,857	34.4

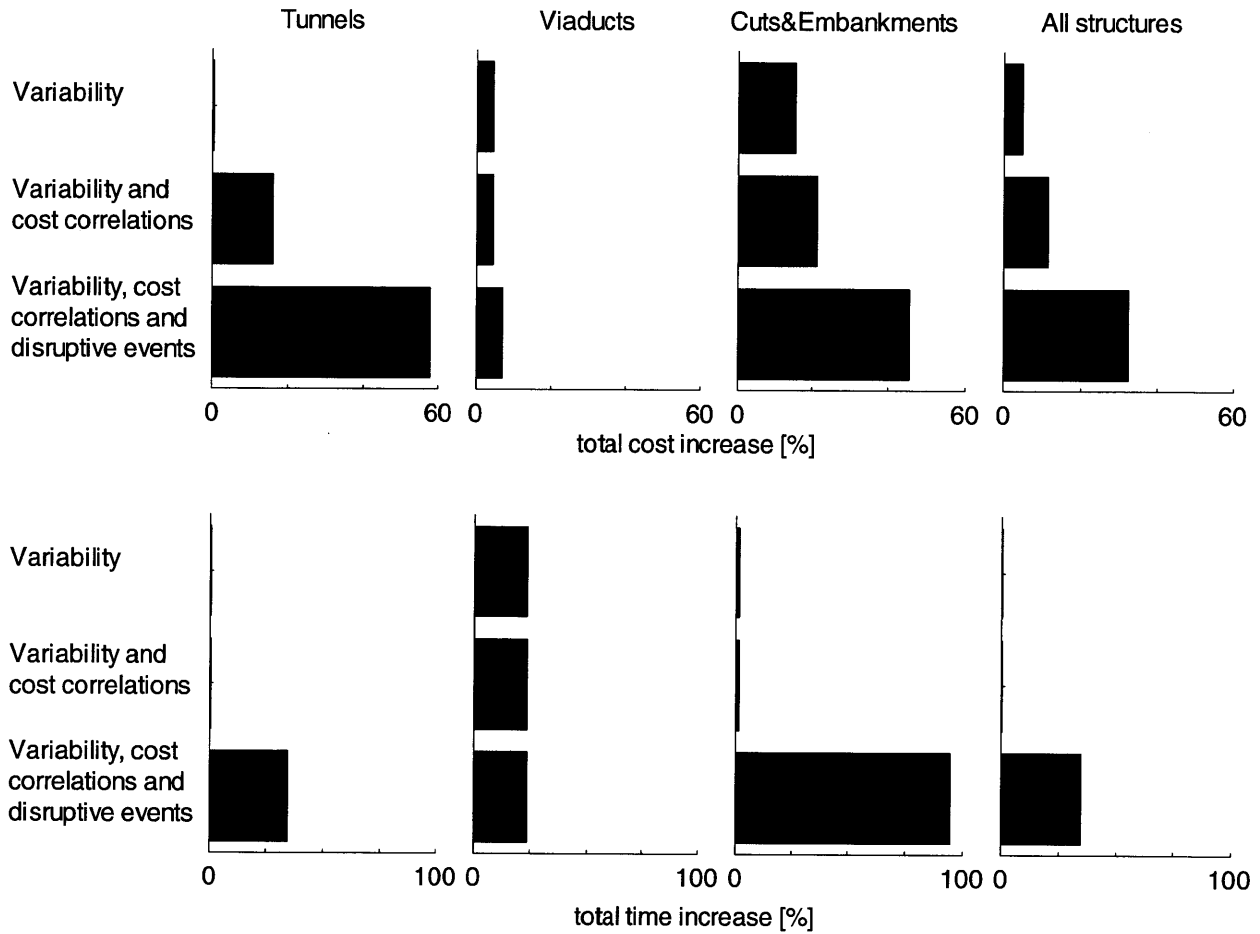


Figure 7-20: Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

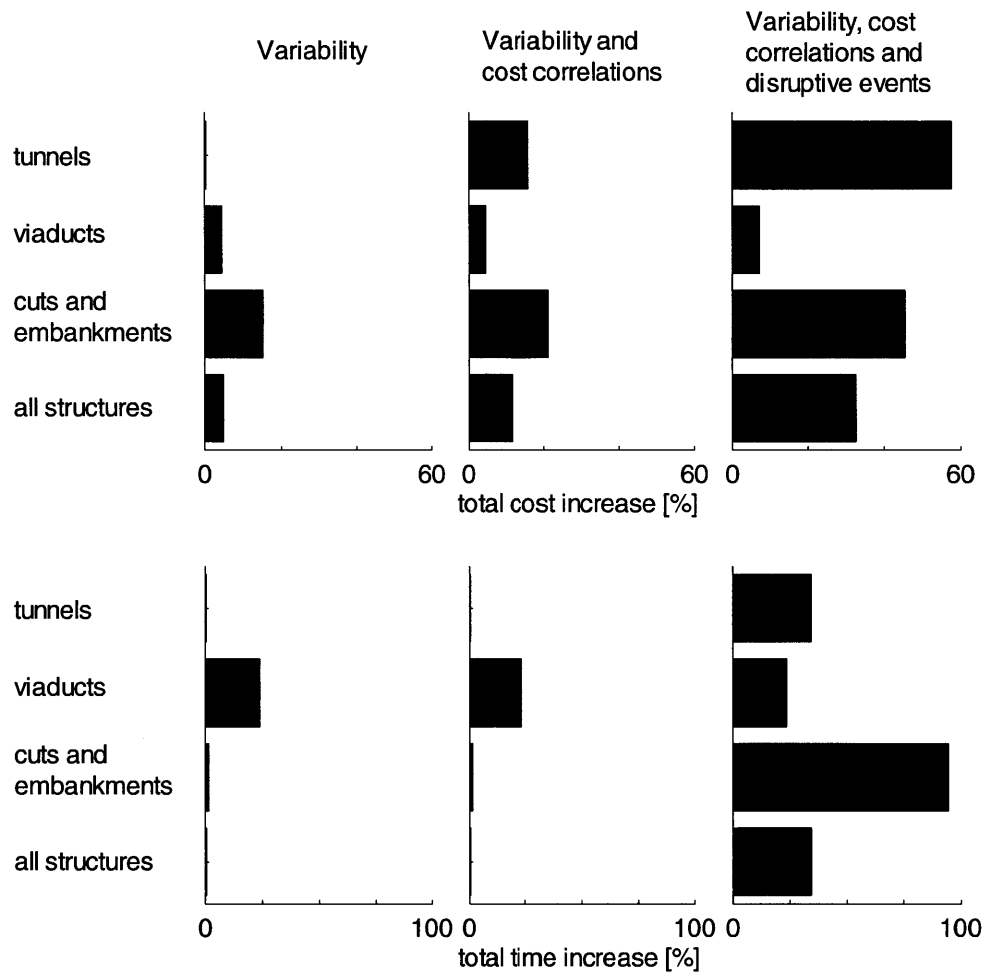


Figure 7-21: Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

21) better shows that, while disruptive events cause large increases in total cost and total time, also variability and cost correlations cause significant increases depending on the structure analyzed. The largest increase due to variability is observed (Figure 7-20 and Table 7.48) for the total cost of cuts and embankments (+15%) and for the total time of viaducts (+23%). The largest increase due to cost correlations is observed for the total cost of tunnels (+16%).

The relevance of a source of uncertainty and its impact depends on the structure:

- For tunnels, variability has an insignificant impact on both total cost and total time, while the total cost increases significantly due to cost correlations (+16%) and disruptive events (+41%) and the total time increases significantly due to disruptive events (+35%).
- For viaducts, variability has the most important impact on total time (+23%) and total cost (+5%), while disruptive events have a limited impact on the total cost (+3%).
- For cuts and embankments, all three sources of uncertainty have some impact on the total cost (variability: +15%; cost correlations: +6%; disruptive events: +24%), while the disruptive events have a truly disruptive impact on the total time (+94%).
- For the entire alignment (all structures), the disruptive events have the largest impact on both total cost (+21%) and total time (+34%).

The different impacts of the sources of uncertainties on the total cost and total time suggests differentiated strategies for the project stakeholders. The tunnel contractor should focus mitigation measures on cost correlations and disruptive events to contain costs and on disruptive events to meet deadlines. The viaduct contractor should focus on reducing the cost and time variability in the construction process to keep both total cost and total time within target. The earthwork (cuts and embankments) contractor must consider all three sources of uncertainty to contain the total cost but can focus on disruptive events to contain the total time. From the perspective of the owner, in order to minimize the total cost of the entire alignment (all structures) the focus is on the disruptive events since they have the largest impact on total cost overall, while in order to minimize total time the attention is on the disruptive events of tunnels since these determine the critical path (section 7.4.3).

The findings and the conclusions are obtained with the analysis of one of the four alignments

(Alignment a-a) of the project. However, similar conclusions are reached in the other three alignments (Appendix D, section D.4).

In conclusion, the impacts of variability, cost correlations, and disruptive events on the total cost and total time distributions vary depending on the structure suggesting the use of differentiated strategies to contain cost overrun and delays for the stakeholders of the project. Most importantly, the cumulative impact of the three sources of uncertainty caused the construction cost and time to increase significantly: the largest increases are observed in the tunnel construction cost (58%) and in the earthwork (cuts and embankments) construction time (94%). The application of the construction and uncertainty models to four alignments of the new Portuguese high speed rail line demonstrates the feasibility of the construction and uncertainty models and their effectiveness in representing the uncertainty at the activity level, in propagating this uncertainty from the individual activity to the cost and time of the whole alignment, and in capturing the cumulative impact of all sources of uncertainty as well as the individual impact of a source of uncertainty on a type of structure (tunnels, viaducts, cuts and embankments) and on the alignment as a whole (all structures).

7.7 Summary and Conclusions

The construction model and the uncertainty model have been applied to the construction of four alignments of the new Portuguese high speed rail line with the simulation tool DAT. The input data was obtained from historical sources and, when these were not available, from expert estimations. With them 1) the deterministic total cost and total time have been calculated, 2) the distributions of the total cost and total time have been simulated for different sources of uncertainty, and 3) the cumulative as well as the individual impact of the sources of uncertainty has been quantified for the different types of structures (tunnels, viaducts, cuts and embankments) and well as for all structures (entire alignment).

Since the total cost and total time of a project are uncertain, they must be represented with probability distributions, whereas single number estimates (deterministic total cost and total time) are inadequate. In the application to the construction of the new Portuguese high speed rail line, the construction model and the uncertainty model are effective in capturing the impact of variability, cost correlations, and two disruptive events. The impact of the three sources of uncertainty

varies from structure to structure: for tunnels the largest cost and time impacts are caused by disruptive events, for viaducts the largest cost and time impacts are caused by variability, for cuts and embankments the largest cost impact is caused by variability and disruptive events and the largest time impact is caused by disruptive events. These findings suggest a structure specific strategy to contain cost and time that focuses on the sources of uncertainty most relevant to the structure. The cumulative impact of the three sources of uncertainty caused significant increases in construction cost and time: the increase in total cost reached 58% in tunnel construction while the increase in total time reached 94% in earthwork (cuts and embankments) construction. Clearly, the analyses presented may deliver different results in other projects, which reinforces the need of uncertainty analysis on a project to project basis.

As the chapter indicates, the application of the construction and uncertainty models is time consuming. It involves four tasks: 1) gathering of data (on the structures of the project and on the uncertainty sources), 2) application of the construction model, 3) application of the uncertainty model and simulation, and 4) analysis of the results. For the application to the new Portuguese high speed rail line, gathering the data was the most time consuming task, while an experienced DAT user can input the structure and uncertainty data of one alignment in two to three weeks. Depending on the size and the complexity of the alignment, simulation and result analysis requires around one week. Albeit time consuming, the application of the construction and uncertainty models and the simulation of the uncertainties in the construction of the rail line provide invaluable insight on the uncertainties that drive project cost and time.

The analyses presented in this chapter were developed through the implementation of the construction- and the uncertainty models in the simulation tool DAT (Decision Aids for Tunneling). In this perspective, these are crucial for the demonstration of the effectiveness of the two models in capturing the impact of the uncertainty sources. Through the implementation of the models and the demonstration of their effectiveness, the DAT are now readily available to be used to model the construction of other rail line projects and similar linear/networked systems.

Chapter 8

Conclusions and Future Research

Cost and time underestimations are widespread in infrastructure construction projects. Among other strategies, improved estimation tools are needed to counteract the frequent underestimations. The author developed a construction model and an uncertainty model, implemented them in the simulation tool DAT (Decision Aids for Tunneling), and applied them to the construction of a new rail line. This chapter summarizes the contributions of the thesis and presents the perspectives for future research.

8.1 Conclusions

This thesis makes four scientific and practical contributions.

First, the research provides an in-depth understanding of the construction process of rail lines. These are modeled as a sequence of four main types of structures: tunnels, viaducts, cuts and embankments. Their construction process is analyzed down to the activity level and the interconnection between activities: single activities are connected into sub-networks modeling the repetitive processes, sub-networks are connected into structure networks modeling the construction of each structure, and structure networks are connected into the construction network modeling the construction of the entire rail line. The proposed construction model considers the relevant aspects determining the construction cost and time of the rail line: the construction costs are equal to the sum of the costs of the individual activities, while the construction time is determined by the order of the performed activities (in sequence/in parallel), by the critical path of the activity network,

and by the material availability. Additionally, the representation of the rail line construction with networks of activities enables one to identify the uncertainties in construction cost and time. This is crucial for the development of the uncertainty model, which analyzes the uncertainty at the activity level.

Second, through the research an in-depth understanding of the uncertainty in the construction cost and time of rail lines is gained. The uncertainty is modeled based on three sources of uncertainty: variability in the construction process, correlations between the costs of repeated activities, and disruptive events. For the first time, these are modeled jointly at the activity level: the cost and the time of an activity are variable; the cost of the activity is correlated with the costs of other activities; and during the activity, one or more disruptive events can occur. The cumulative impact of the three sources of uncertainty is quantitatively captured in the construction cost and -time distributions. Beyond the representation of uncertainty in construction, the uncertainty model has further uses: it is the starting point for mitigation measures and budget allocation, and it can be used throughout a project to update the impact of the uncertainties and to evaluate the effectiveness of countermeasures to mitigate threats.

Third, the construction model and the uncertainty model are implemented in the simulation tool DAT. In order to model the construction of rail lines, these are extended to represent the construction of not only tunnels but also viaducts, cuts and embankments. The variability, the cost correlations and the disruptive events are implemented in the DAT with probability distributions (cost and time variability, and cost and time impacts of disruptive events), the correlation method NORTA and Markov processes (probability of occurrence of the disruptive events). From the perspective of practical applications, the proposed models and their implementation in the DAT represent an innovative tool to model the uncertainties in construction cost and time of rail lines and other linear/networked infrastructure projects.

Fourth, the application of the construction model and the uncertainty model to the construction of four alignments of the new Portuguese high speed rail line clearly shows the effectiveness of the proposed models in capturing the uncertainty in construction cost and time and the feasibility of their application to a rail line project. The construction model was applied to the project by representing the construction of the different alignments with activity networks, simulating the construction process and obtaining the deterministic construction cost and time. As these matched

the cost and time calculated based on the RAVE data (Appendix C), the construction model was validated. The uncertainty model was applied by representing the variability in the activity cost and time, the correlations between the costs of repeated activities and the disruptive events in the construction of the alignments and by simulating their impacts (quantified in terms of the 90th percentiles and deterministic construction cost and time). The individual impact of the three sources of uncertainty varies depending on the structure: for tunnels, the largest cost and time impacts are caused by disruptive events, for viaducts they are caused by variability, while for cuts and embankments the largest cost impact is caused by variability and disruptive events and the largest time impact is caused by disruptive events. Thus, tunnel, viaduct, and earthwork contractors can contain the construction cost and time with differentiated, structure specific strategies. The cumulative impact of the three sources of uncertainty caused the construction cost and time to increase significantly: the largest increases are observed in the tunnel construction cost (58%) and in the earthwork (cuts and embankments) construction time (94%). The construction cost increase is significantly larger than contingency usually applied to transportation construction projects, while the construction time increase almost doubles the construction time. Although the application of the construction and uncertainty models is more time consuming than calculating a single number estimate (deterministic construction cost and time), the insight it provides on the magnitude of the impact and the uncertainty source driving the impact is invaluable. Also the two models can be used to compare different alignments from the perspective of the impact of the sources of uncertainty: due to the route, the structures and the construction conditions (e.g. geological); one alignment might be significantly more impacted by uncertainty than another alignment. Last but not less important, the DAT are a crucial tool in the application of the construction and uncertainty models to the construction of the rail line project. In their current, extended form they can be used to model the uncertainty in the construction of other rail line projects and similar linear/networked systems.

8.2 Current Limitations and Future Perspectives

The development and the implementation of the construction and uncertainty models although satisfactory are subject to some limitations. From these and other considerations, perspectives for future research are presented.

The construction model currently models only one resource: the material excavated in cuts is reused in embankments. A first extension of the resource model could include the reuse of the material excavated in cuts and tunnels for the aggregate of tunnel and viaduct concrete (This is not relevant in the application to the new Portuguese high speed rail line, where the material excavated in tunnels and the excess material from cuts is considered waste to be deposited). A second extension could include a comprehensive model for linear/networked infrastructure projects of the resources, material and workforce, and resource flows with the purpose of representing the idle times and the idle costs if an activity is waiting for resources that are not available at the time.

In its current form, the uncertainty model does not model correlations between the times of activities although these might have a significant impact on the standard deviation of the construction time. Correlations between the times of activities performed in sequence and in parallel are expected. Modeling correlations between the times of parallel activities is complex since it is not known a priori which activities are performed in parallel: this is determined by the times of preceding activities in the network, the resource availability and the occurrence of disruptive events. Modeling correlations between the times of parallel activities requires a construction model that determines for each simulation run which activities are performed in parallel before generating the correlated times.

In the uncertainty model, the uncertainty data not available from historical sources are obtained through expert estimations. In the application of the uncertainty model to the new Portuguese high speed rail line, a tunnel expert, a viaduct expert, and an earthwork (cuts and embankments) expert estimated the needed data. Relying on one expert per type of structure does not consider the uncertainty in the expert estimation: in fact, having different experts may lead to significantly different estimations. Thus, the application of the uncertainty model can be improved by including more than one expert per structure.

The construction model and the uncertainty model have been applied to the construction of the new Portuguese rail line and their effectiveness in capturing the impact of uncertainty on the construction cost and construction time has been shown. However, a proper validation of the proposed models by means of a comparison between the forecast and the observed construction cost and construction time is at the time not possible, since the project is in the pre-construction phase. Also a full-fledged validation would have to include the modeling of the time correlations, a more

comprehensive analysis of disruptive events (i.e. disruptive events with negative as well as positive outcome and not just two disruptive events per type of structure) and the mitigation measures applied by the transportation agency to limit the risks. Clearly, such a validation is affected by the time interval between a project forecast and the actual completion of the infrastructure construction.

Appendix A

Questionnaire for Expert Estimation

The uncertainty model proposed in Chapter 4 and applied to the construction of a new rail line in Chapter 7 is based on historical data and expert opinions. Historical data are available for the cost variability of all structures and for the time variability of tunnels. For the rest of the uncertainties, experts were asked to estimate the needed information (Table A.1). These were obtained in estimation sessions in which experts answered questions about uncertainties based on their experience. Biases, calibration of the experts to avoid biases and questionnaire structure are addressed in section 4.3. For the research work presented in this thesis, a tunnel expert, a viaduct expert and an earthwork (cuts and embankments) expert gave their estimates on time variability, cost correlations and disruptive events in their respective fields of expertise.

In the sequence, the questionnaire of the tunnel expert, the viaduct expert and the earthwork (cuts and embankments) expert are reported here for reference.

Table A.1: Structures (tunnels, viaducts, cuts, and embankments) and uncertainties (variability, cost correlations, and disruptive events) in the construction of rail lines. In this research work, the time variability of viaduct-, cut- and embankment construction as well as the cost correlations and the disruptive events of all structures were not available from historical data and they had therefore to be estimated by experts.

Structures	Uncertainties	variability		cost	disruptive
		cost	time	correlations	events
tunnels		data	data	expert	expert
viaducts		data	expert	expert	expert
earthwork (cuts & embankments)		data	expert	expert	expert

Questionnaire for Tunnel Expert

Cost Correlations

What is the correlation between the cost of constructing a unit length (i) and the cost of constructing the next unit length ($i + 1$)?

$$\rho(i, i + 1) =$$

At which j (j : # of unit length) is the correlation between the cost of constructing a unit length (i) and the cost of constructing a unit length ($i + j$) equal to zero (i.e. the costs are uncorrelated)?

$$\rho(i, i + j) = 0, \text{ for } j =$$

How does the correlation decrease from the ($i + 1$) unit length to the ($i + j$) unit length?

Disruptive Events

Name the two most important disruptive events in tunnel construction.

1)

2)

What is the probability of occurrence of 1)?

What is the probability of occurrence of 2)?

At the minimum what is the cost impact of 1)?

In most cases what is the cost impact of 1)?

At the maximum what is the cost impact of 1)?

At the minimum what is the cost impact of 2)?

In most cases what is the cost impact of 2)?

At the maximum what is the cost impact of 2)?

At the minimum what is the time impact of 1)?

In most cases what is the time impact of 1)?

At the maximum what is the time impact of 1)?

At the minimum what is the time impact of 2)?

In most cases what is the time impact of 2)?

At the maximum what is the time impact of 2)?

Questionnaire for Viaduct Expert

Variability

At the maximum how long does it take to set up the construction site?

At the maximum how long does it take to construct an abutment?

At the maximum how long does it take to construct a technical block?

At the maximum how long does it take to construct a pile set?

At the maximum how long does it take to construct a footing?

At the maximum how long does it take to construct a pier?

At the maximum how long does it take to set up the deck construction?

- ground scaffolding with staging
- movable scaffolding with gantry

At the maximum how long does it take to construct a pier?

At the maximum how long does it take to construct a deck section between two piers (in cast-in-place, reinforced, pre-stressed concrete)?

At the maximum how long does it take to construct the viaduct finishing?

In most cases how long does it take to set up the construction site?

In most cases how long does it take to construct an abutment?

In most cases how long does it take to construct a technical block?

In most cases how long does it take to construct a pile set?

In most cases how long does it take to construct a footing?

In most cases how long does it take to construct a pier?

In most cases how long does it take to set up the deck construction?

- ground scaffolding with staging
- movable scaffolding with gantry

In most cases how long does it take to construct a pier?

In most cases how long does it take to construct a deck section between two piers (in cast-in-place, reinforced, pre-stressed concrete)?

In most cases how long does it take to construct the viaduct finishing?

At the minimum how long does it take to set up the construction site?

At the minimum how long does it take to construct an abutment?

At the minimum how long does it take to construct a technical block?

At the minimum how long does it take to construct a pile set?

At the minimum how long does it take to construct a footing?

At the minimum how long does it take to construct a pier?

At the minimum how long does it take to set up the deck construction?

- ground scaffolding with staging
- movable scaffolding with gantry

At the minimum how long does it take to construct a pier?

At the minimum how long does it take to construct a deck section between two piers (in cast-in-place, reinforced, pre-stressed concrete)?

At the minimum how long does it take to construct the viaduct finishing?

Cost Correlations

What is the correlation between the cost of a pier (i) and the cost of the next pier ($i + 1$)?

$$\rho(i, i + 1) =$$

At which j (j : # of piers) is the correlation between the cost of a pier (i) and the cost of the pier ($i + j$) equal to zero (i.e. the costs are uncorrelated)?

$$\rho(i, i + j) = 0, \text{ for } j =$$

How does the correlation decrease from the ($i + 1$) pier to the ($i + j$) pier?

How does the correlation decrease from the ($i + 1$) deck section to the ($i + j$) deck section?

$$\rho(i, i + 1) =$$

At which j (j : # of deck sections) is the correlation between the cost of a deck section (i) and the cost of the deck section ($i + j$) equal to zero (i.e. the costs are uncorrelated)?

$$\rho(i, i + j) = 0, \text{ for } j =$$

How does the correlation decrease from the ($i + 1$) deck section to the ($i + j$) deck section?

Disruptive Events

Name the two most important disruptive events in viaduct construction.

1)

2)

What is the probability of occurrence of 1)?

What is the probability of occurrence of 2)?

At the minimum what is the cost impact of 1)?

In most cases what is the cost impact of 1)?

At the maximum what is the cost impact of 1)?

At the minimum what is the cost impact of 2)?

In most cases what is the cost impact of 2)?

At the maximum what is the cost impact of 2)?

At the minimum what is the time impact of 1)?

In most cases what is the time impact of 1)?

At the maximum what is the time impact of 1)?

At the minimum what is the time impact of 2)?

In most cases what is the time impact of 2)?

At the maximum what is the time impact of 2)?

Questionnaire for Earthwork (Cuts and Embankments) Expert Variability

At the minimum how many square yards per hour (syd/h) can be cleared?

At the minimum how many cubic yards per hour (cyd/h) can be excavated (mechanical)?

At the minimum how many cubic yards per hour (cyd/h) can be excavated (blasting)?

At the minimum how many cubic yards of capping per hour (cyd/h) can be placed?

At the minimum how many cubic yards of sub-ballast per hour (cyd/h) can be placed?

At the minimum how many cubic yards per hour (cyd/h) can be improved (soil replacement)?

At the minimum how many cubic yards per hour (cyd/h) can be filled?

In most cases how many square yards per hour (syd/h) can be cleared?

In most cases how many cubic yards per hour (cyd/h) can be excavated (mechanical)?

In most cases how many cubic yards per hour (cyd/h) can be excavated (blasting)?

In most cases how many cubic yards of capping per hour (cyd/h) can be placed?

In most cases how many cubic yards of sub-ballast per hour (cyd/h) can be placed?

In most cases how many cubic yards per hour (cyd/h) can be improved (soil replacement)?

In most cases how many cubic yards per hour (cyd/h) can be filled?

At the maximum how many square yards per hour (syd/h) can be cleared?

At the maximum how many cubic yards per hour (cyd/h) can be excavated (mechanical)?

At the maximum how many cubic yards per hour (cyd/h) can be excavated (blasting)?

At the maximum how many cubic yards of capping per hour (cyd/h) can be placed?

At the maximum how many cubic yards of sub-ballast per hour (cyd/h) can be placed?

At the maximum how many cubic yards per hour (cyd/h) can be improved (soil replacement)?

At the maximum how many cubic yards per hour (cyd/h) can be filled?

Cost Correlations

What is the correlation between the cost of excavating the (i) cyd and the cost of excavating the next cyd ($i + 1$)?

$$\rho(i, i + 1) =$$

At which j (j : # of cyd excavated) is the correlation between the cost of excavating the (i) cyd and the cost of excavating the ($i + j$) cyd equal to zero (i.e. the costs are uncorrelated)?

$$\rho(i, i + j) = 0, \text{ for } j =$$

How does the correlation decrease from the ($i + 1$) cyd to the ($i + j$) cyd?

What is the correlation between the cost of filling the (i) cyd and the cost of filling the next cyd ($i + 1$)?

$$\rho(i, i + 1) =$$

At which j (j : # of cyd filled) is the correlation between the cost of filling the (i) cyd and the cost of filling the ($i + j$) cyd equal to zero (i.e. the costs are uncorrelated)?

$$\rho(i, i + j) = 0, \text{ for } j =$$

How does the correlation decrease from the ($i + 1$) cyd to the ($i + j$) cyd?

Disruptive Events

Name the two most important disruptive events in cut/embankment construction.

1)

2)

What is the probability of occurrence of 1)?

What is the probability of occurrence of 2)?

At the minimum what is the cost impact of 1)?

In most cases what is the cost impact of 1)?

At the maximum what is the cost impact of 1)?

At the minimum what is the cost impact of 2)?

In most cases what is the cost impact of 2)?

At the maximum what is the cost impact of 2)?

At the minimum what is the time impact of 1)?

In most cases what is the time impact of 1)?

At the maximum what is the time impact of 1)?

At the minimum what is the time impact of 2)?

In most cases what is the time impact of 2)?

At the maximum what is the time impact of 2)?

Appendix B

Probability Concepts

Chapter 5 Correlations investigated correlations in the construction of rail lines. While reviewing correlation measures and models to simulation such correlations, the following probability concepts need to be applied:

1. Variance of the sum of cost variables,
2. Independent versus uncorrelated variables,
3. Correlation methods generating samples with desired Pearson correlation, and
4. Positive definiteness of a correlation matrix.

These concepts are explained in the following.

B.1 Variance of the Sum of Cost Variables

The standard deviation of the sum of cost variables is underestimated if the costs are assumed independent (i.e. uncorrelated) when they are in fact positively correlated. The impact of positive correlation on the variance (and the standard deviation $\sigma = \sqrt{Var}$) of the sum of costs can be explained with probability theory. The variance of the sum of costs X and Y is equal to (see e.g. Bertsekas & Tsitsiklis 2002):

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y). \quad (B.1)$$

By substituting the covariance with

$$\text{Cov}(X, Y) = \rho(X, Y) \sqrt{\text{Var}(X) \text{Var}(Y)}, \quad (\text{B.2})$$

we obtain

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\rho(X, Y) \sqrt{\text{Var}(X) \text{Var}(Y)}, \quad (\text{B.3})$$

where $\rho(X, Y)$ is the correlation between costs X and Y . Equation B.3 clearly shows that a positive correlation between costs X and Y increases the variance of the sum of the costs X and Y .

This probability concept is important when calculating the total cost of rail line construction. When cost elements of the rail line construction are positively correlated, one can expect the variance of the total cost to be larger than the sum of cost elements variances. If one disregards the correlation between cost elements, the variance (and the standard deviation) of the total cost is underestimated.

B.2 Independent Versus Uncorrelated Variables

Independent random variables are not equivalent to uncorrelated random variables. Independence implies absence of correlation, but absence of correlation does not imply independence. Thus, two independent random variables are uncorrelated. However, two uncorrelated random variables are not necessarily independent.

Definition B.1 *Two events A and B are said to be independent if*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \mathbf{P}(B). \quad (\text{B.4})$$

If in addition, $\mathbf{P}(B) > 0$, independence is equivalent to the condition

$$\mathbf{P}(A|B) = \mathbf{P}(A) \quad (\text{B.5})$$

(Bertsekas & Tsitsiklis 2002).

The case of two uncorrelated random variables, which are not independent, is illustrated in the following example (Bertsekas & Tsitsiklis 2002).

Example B.1 The random variables (X, Y) take the values $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$, each with probability $1/4$ (see Figure B-1). The correlation between the two random variables is 0. Thus, the two random variables are uncorrelated. However, the two random variables are not independent. E.g. the event A occurs when $X = 1$, the event B occurs when $Y = 0$. Then,

$$\mathbf{P}(X = 1) = \mathbf{P}(A) = 1/4$$

$$\mathbf{P}(Y = 0) = \mathbf{P}(B) = 1/2$$

$$\mathbf{P}(A \cap B) = \mathbf{P}(X = 1, Y = 0) = 1/4$$

so that

$$\mathbf{P}(A)\mathbf{P}(B) \neq \mathbf{P}(A \cap B).$$

Thus, the two random variables are not independent.

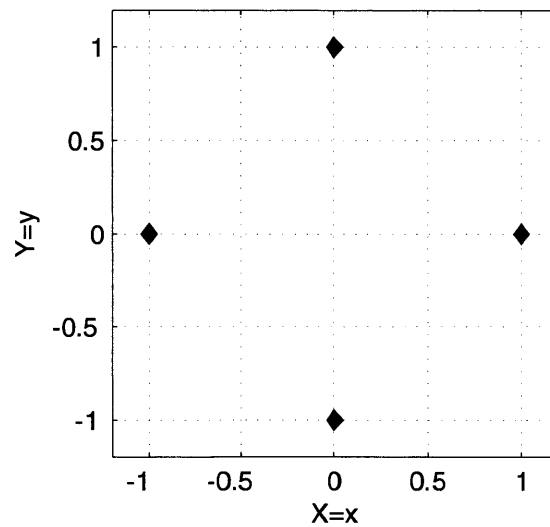


Figure B-1: Uncorrelated but not independent variables X and Y . Independent variables are uncorrelated, whereas uncorrelated variables are not necessarily independent.

B.3 Correlation Methods generating Samples with Desired Pearson Correlation

In chapter 5, section 5.2, the best suited correlation measure to represent correlations in the construction model was sought. The Pearson correlation and two other correlation measures were presented. One pitfall of the Pearson correlation was that it is not invariant under non-linear strictly increasing transformations, such as the cumulative distribution function. As a consequence, another correlation measure was selected to represent correlations in the construction model.

Despite this pitfall, methods have been developed to generate samples with the desired Pearson correlations: the method by Cario & Nelson (1997), NORTA, was explained in section 5.3, the methods by Li & Hammond (1975) and by Lurie & Goldberg (1998) are explained below. The goal of the two methods is generating random samples with desired Pearson correlations.

The method by Li & Hammond (1975) is based on the following steps:

1. Generate samples from independent Gaussian random variables, (Y_1, \dots, Y_n) .
2. Apply the linear transformation matrix G to the independent Gaussian random variables (Y_1, \dots, Y_n) in order to obtain Gaussian random variables, (V_1, \dots, V_n) , with Pearson correlation coefficients $\rho_{V_i V_j}$.
3. Apply the non-linear transformations h_i to the correlated Gaussian random variables (V_1, \dots, V_n) in order to obtain random variables, (Z_1, \dots, Z_n) with desired probability distributions and desired correlation coefficients $\rho_{Z_i Z_j}$.

Step 1 allows one to generate samples that do not have the desired correlation matrix or the desired probability distributions. Step 2 allows one to generate samples with desired correlation matrix but with probability distributions different from the desired ones. Step 3 allows one to generate a sample with desired correlation matrix and desired probability distributions. The linear transformation matrix G and the non-linear transformations h_i are determined depending on the desired correlation coefficients, $\rho_{Z_i Z_j}$, and the desired probability distributions for random variables, (Z_1, \dots, Z_n) .

Li & Hammond (1975) observe that every valid Pearson correlation coefficient matrix for (Z_1, \dots, Z_n) does not necessarily correspond to a valid Pearson correlation coefficient matrix for

(V_1, \dots, V_n) . For a Pearson correlation coefficient matrix to be valid in the method by Li & Hammond (1975), the matrix must be positive definite. Li & Hammond (1975) constructed an example where a valid Pearson correlation matrix R for (Z_1, \dots, Z_n) does not correspond to a valid Pearson correlation matrix for (V_1, \dots, V_n) .

Example B.2 *Let (Z_1, Z_2, Z_3) be uniform random variables with Pearson correlation coefficients $\rho_{Z_1Z_2}, \rho_{Z_2Z_3}, \rho_{Z_1Z_3}$ equal to $-0.4, 0.2$ and 0.8 , respectively. These values yield a positive definite correlation matrix. An analytical expression that relates the Pearson correlation coefficients $\rho_{Z_1Z_2}, \rho_{Z_2Z_3}, \rho_{Z_1Z_3}$ with the Pearson correlation coefficients $\rho_{V_1V_2}, \rho_{V_2V_3}, \rho_{V_1V_3}$ of two uniform distributed random variables is*

$$\rho_{V_iV_j} = 2 \sin\left(\frac{\pi}{6}\rho_{Z_iZ_j}\right), \quad i \neq j. \quad (\text{B.6})$$

From expression B.6, one obtains $\rho_{V_1V_2}, \rho_{V_2V_3}, \rho_{V_1V_3}$ equal to $-0.4158, 0.2091$, and 0.8135 , respectively. These values do not yield a positive definite correlation matrix.

The method by Lurie & Goldberg (1998) is a further development of the method by Li & Hammond (1975). In the instances where a valid Pearson correlation coefficient matrix for (Z_1, \dots, Z_n) does not correspond to a valid Pearson correlation coefficient matrix for (V_1, \dots, V_n) , Lurie & Goldberg (1998) propose a method where a positive definite input correlation matrix for (V_1, \dots, V_n) can be chosen in such a way to obtain a positive definite correlation matrix "close" to the desired correlation matrix for (Z_1, \dots, Z_n) .

B.4 Positive Definiteness

NORTA is a correlation model that can generate samples with desired correlation matrix and desired marginal distributions. To perform the first step of the NORTA model, the Pearson correlation matrix, R , must be positive definite. $x^T Ax > 0$ for all nonzero real vectors x is a sufficient condition for the real symmetric matrix A to be positive definite (see e.g. Strang 2006). R must be positive definite because the algorithm to generate correlated samples includes the calculation of the lower triangular matrix P such that $R = PP^T$. Only if R is positive definite P consists of real numbers (versus imaginary numbers).

A 2×2 correlation matrix is always positive definite. A 3×3 or higher dimensionality correlation matrix might or might not be positive definite. It has been observed that the probability of

obtaining a positive definite correlation matrix when randomly generating the correlation matrix' entries decreases with increasing dimensionality (Ghosh & Henderson 2003, Touran 1993). If the Spearman correlation has been estimated by an expert rather than calculated from a good-sized data set, it is possible that the expert estimated an inconsistent correlation matrix so that the calculated Pearson correlation matrix is not positive definite. The question what to do if the calculated Pearson correlation matrix is not positive definite poses itself.

A not positive definite correlation matrix means that its entries do not yield a meaningful correlation matrix. Example B.3 is a constructed example to clarify this point.

Example B.3 *Random variables X_1, X_2, X_3 are correlated. An expert estimates the Spearman correlation matrix. The Pearson correlation matrix calculated from the estimated Spearman correlation matrix is*

$$R = \begin{bmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0.9 \\ 0 & 0.9 & 1 \end{bmatrix}.$$

R is a correlation matrix (ones on the diagonal, symmetric and entries within $[-1, 1]$). R is not a positive definite matrix. One can argue that the matrix entries do not yield a meaningful correlation matrix. In fact, if X_1 and X_2 are highly correlated ($\rho_{12} = 0.9$) and X_2 and X_3 are also highly correlated ($\rho_{23} = 0.9$), it is not reasonable that X_1 and X_3 are uncorrelated ($\rho_{13} = 0$).

Thus, a non positive definite Pearson correlation matrix should be understood as an indication for lack of meaningfulness of the estimated Spearman correlation matrix. If the calculated Pearson correlation matrix is not positive definite, some possible options are:

- Reassess the estimated Spearman correlation matrix with the expert.
- Reduce gradually the matrix entries until the correlation matrix is positive definite (Touran 1993). This option arises from the observation that matrices with large correlation coefficients tend to be not positive definite. The disadvantage of this approach is that the generated random variables will not have the desired correlation matrix.
- Eliminate some matrix entries aiming at achieving a positive definite correlation matrix (Touran & Wiser 1992). This option arises from the observation that the probability of obtaining a positive definite correlation matrix, when randomly generating the correlation matrix'

entries, decreases with increasing dimensionality (Ghosh & Henderson 2003, Touran 1993). Again the disadvantage is that the generated random variables will not have the desired correlation matrix.

- Find a positive definite correlation matrix close to the desired correlation matrix with the correction method by Ghosh & Henderson (2002).

Appendix C

Structures of the New Portuguese High Speed Rail Line

Chapter 7 presented the application of the construction model and the uncertainty model to the construction of four alignments of the new Portuguese high speed rail line. The construction model and the uncertainty model require a large amount of input data and of calculations/simulations. These are presented in this appendix.

C.1 List of structures

The four alignments of the new Portuguese high speed rail line are modeled with a sequence of four types of structures: tunnels, viaducts, cuts and embankments. The data on the structures include start position, end position and length of all the structures in alignments A-a and A-c (RAVE 2006d, RAVE 2006c, RAVE 2006b) and in alignments B-a and B-b (RAVE 2008d, RAVE 2008c, RAVE 2008b). However, after combining the data, gaps and overlaps between structures emerge. The overlaps are in the order of few dozens meters with an extreme case of 950m, whilst the gaps are in the order of few hundred meters with an extreme case of 1.8km. These discrepancies are overcome using the following heuristics:

1. Gap between a tunnel and a cut: the cut adapts to the tunnel and is lengthened.
2. Gap between a viaduct and an embankment: the embankment adapts to the viaduct and is lengthened.

3. Gap between a cut and an embankment: the length is evenly distributed between the cut and the embankment.
4. Gap between two cuts: since RAVE (2006b) omits embankments with height less than 4m (alignments A-a and A-c) and RAVE (2008b) omits embankments with height less than 10m (alignments B-a and B-b), the gap between two cuts is replaced with an embankment of 4m (alignments A-a and A-c) or 10m height (alignments B-a and B-b). It is assumed that the construction of the gap embankment does not require soil improvement.
5. Gap between two embankments: since RAVE (2006b) omits cuts with depth less than 4m (alignments A-a and A-c) and RAVE (2008b) omits cuts with depth less than 10m (alignments B-a and B-b), the gap between two embankments is replaced with a cut of 4m (alignments A-a and A-c) or 10m depth (alignments B-a and B-b).
6. Gap between a cut and a viaduct: the cut is not lengthened to adapt to the viaduct since usually viaducts are preceded and followed by embankments. The gap is replaced with an embankment of 4m (alignments A-a and A-c) or 10m (alignments B-a and B-b) height.
7. Gap between an embankment and a tunnel: the embankment is not lengthened to adapt to the tunnel since usually tunnels are preceded and followed by cuts. The gap is replaced with a cut of 4m (alignments A-a and A-c) or 10m (alignments B-a and B-b) depth.
8. Overlap between a cut and an embankment: the length is evenly distributed between the two structures.

The obtained start positions, end positions and lengths of the structures in the four alignments are listed in Tables C.1 to C.4.

C.2 Tunnels

The construction of the tunnels in the four alignments of the new Portuguese high speed rail line was presented in chapter 7, section 7.3. In the following, the ground parameter states, the ground classes and the construction methods to model the construction of tunnels are discussed in more detail.

Table C.1: List of structures in alignment A-a. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure at the beginning of the alignment.

Structure	start [m]	end [m]	length [m]	structure	start [m]	end [m]	length [m]
D1	0	754	754	C11	19,730	21,220	1,490
T1	754	1,750	996	E12	21,220	21,620	400
C1	1,750	2,720	970	C12	21,620	22,530	910
E1	2,720	2,900	180	E13	22,530	22,700	170
V1	2,900	3,665	765	V4	22,700	25,400	2,700
E2	3,665	4,000	335	E14	25,400	25,460	60
C2	4,000	4,700	700	C13	25,460	26,050	590
E3	4,700	5,250	550	E15	26,050	26,900	850
C3	5,250	8,515	3,265	C14	26,900	27,850	950
T2	8,515	9,970	1,455	E16	27,850	28,700	850
C4	9,970	10,215	245	C15	28,700	29,610	910
E4	10,215	10,350	135	E17	29,610	31,850	2,240
V2	10,350	11,050	700	C16	31,850	32,450	600
E5	11,050	11,110	60	T4	32,450	32,950	500
C5	11,110	11,230	120	C17	32,950	33,230	280
E6	11,230	11,720	490	E18	33,230	33,274	44
C6	11,720	12,000	280	V5	33,274	33,669	395
E7	12,000	12,210	210	E19	33,669	33,740	71
C7	12,210	12,818	608	C18	33,740	34,100	360
T3	12,818	13,500	682	E20	34,100	35,560	1,460
C8	13,500	13,610	110	C19	35,560	36,200	640
E8	13,610	13,683	73	T5	36,200	38,850	2,650
V3	13,683	18,535	4,852	C20	38,850	38,980	130
E9	18,535	18,590	55	E21	38,980	39,650	670
C9	18,590	18,750	160	C21	39,650	40,050	400
E10	18,750	18,950	200	T6	40,050	43,707	3,657
C10	18,950	19,420	470	C22	43,707	44,500	793
E11	19,420	19,730	310	E22	44,500	46,500	1,100

Table C.2: List of structures in alignment A-c. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure at the beginning of the alignments.

Structure	start [m]	end [m]	length [m]	structure	start [m]	end [m]	length [m]
D1	0	760	760	E11	21,896	22,646	750
T1	760	1,750	990	C10	22,646	23,166	520
C1	1,750	3,090	1,340	E12	23,166	25,486	2,320
E1	3,090	3,300	210	C11	25,486	26,891	1,405
C2	3,300	4,670	1,370	E13	26,891	27,541	650
E2	4,670	5,180	510	C12	27,541	28,066	525
C3	5,180	5,620	440	E20	28,066	28,084	18
E3	5,620	5,980	360	C13	28,084	28,659	575
C4	5,980	6,500	520	E14	28,659	30,899	2,240
V1	6,500	6,600	100	C14	30,899	31,499	600
E4	6,600	7,450	850	T3	31,499	31,999	500
C21	7,450	8,170	720	C15	31,999	32,279	280
E5	8,170	8,300	130	E15	32,279	32,323	44
C5	8,300	10,050	1,750	V4	32,323	32,718	395
E6	10,050	11,225	1,175	E16	32,718	32,789	71
C6	11,225	12,050	825	C16	32,789	33,149	360
E7	12,050	12,100	50	E17	33,149	34,609	1,460
V2	12,100	13,020	920	C17	34,609	35,249	640
E8	13,020	15,300	2,280	T4	35,249	37,899	2,650
C7	15,300	15,964	664	C18	37,899	38,029	130
T2	15,964	17,675	1,711	E18	38,029	38,699	670
C8	17,675	18,066	391	C19	38,699	39,099	400
E9	18,066	19,416	1,350	T5	39,099	42,756	3,657
V3	19,416	19,666	250	C20	42,756	43,549	793
E10	19,666	21,016	1,350	E19	43,549	44,649	1,100
C9	21,016	21,896	880				

Table C.3: List of structures in alignment B-a. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure at the beginning of the alignment.

Structure	start [m]	end [m]	length [m]	structure	start [m]	end [m]	length [m]
D1	0	1,000	1,000	E11	36,250	36,500	250
E1	1,000	2,025	1,025	C26	36,500	36,700	200
C1	2,025	3,200	1,175	E12	36,700	37,100	400
T1	3,200	3,350	150	C27	37,100	37,150	50
C2	3,350	3,598	248	T8	37,150	37,350	200
T2	3,598	5,841	2,243	C28	37,350	37,400	50
C3	5,841	7,200	1,359	E13	37,400	37,700	300
E2	7,200	9,640	2,440	C14	37,700	38,000	300
V1	9,640	9,970	330	E24	38,000	38,100	100
E3	9,970	11,430	1,460	C15	38,100	39,100	1,000
V2	11,430	13,950	2,520	C16	39,100	39,550	450
E4	13,950	15,875	1,925	E25	39,550	39,648	98
C4	15,875	16,300	425	V6	39,648	40,662	1,014
E5	16,300	17,050	750	E26	40,662	40,700	38
C24	17,050	17,900	850	C17	40,700	41,103	403
E6	17,900	19,125	1,225	T9	41,103	45,658	4,555
C5	19,125	20,737	1,612	C18	45,658	46,950	1,292
T3	20,737	22,240	1503	E14	46,950	48,500	1,550
T4	22,240	22,500	260	C29	48,500	48,650	150
C25	22,500	23,000	500	T10	48,650	48,850	200
E7	23,000	23,500	500	C19	48,850	49,100	250
V3	23,500	23,620	120	E15	49,100	49,999	899
E21	23,620	23,650	30	V7	49,999	50,278	279
C6	23,650	23,950	300	E27	50,278	50,300	22
E22	23,950	24,300	350	C20	50,300	52,717	2,417
C7	24,300	24,579	279	T11	52,717	54,966	2,249
T5	24,579	27,600	3,021	C21	54,966	55,110	144
C8	27,600	27,850	250	E16	55,110	55,950	840
E23	27,850	28,100	250	V8	55,950	56,250	300
C9	28,100	28,540	440	E17	56,250	57,254	1,004
V4	28,540	28,745	205	V9	57,254	58,274	1,020
E8	28,745	29,650	905	E18	58,274	59,400	1,126
C10	29,650	29,800	150	C30	59,400	60,100	700
T6	29,800	30000	200	E19	60,100	61,200	1,100
C11	30,000	31,334	1,334	C31	61,200	61,550	350
T7	31,334	32,645	1,311	T12	61,550	61,700	150
C12	32,645	33,200	555	C32	61,700	61,900	200
E9	33,200	33,773	573	E20	61,900	62,500	600
V5	33,773	35,393	1,620	C22	62,500	63,750	1,250
E10	35,393	35,650	257	E28	63,750	63,900	150
C13	35,650	36,250	600	C23	63,900	66,500	2,600

Table C.4: List of structures in alignment B-b. "T" stands for tunnel, "V" for viaduct, "C" for cut, and "E" for embankment. "D1" is a dummy structure.

Structure	start [m]	end [m]	length [m]	structure	start [m]	end [m]	length [m]
D1	0	1,000	1,000	V6	33,940	36,788	2848
E1	1,000	2,025	1,025	E12	36,788	37,000	212
C1	2,025	3,200	1,175	C13	37,000	37,400	400
T1	3,200	3,350	150	E13	37,400	39,745	2,075
C2	3,350	3,598	248	C14	39,745	41,350	1,875
T2	3,598	5,841	2,243	E27	41,350	41,370	20
C3	5,841	7,200	1,359	V7	41,370	41,481	111
E2	7,200	9,640	2,440	E28	41,481	41,500	19
V1	9,640	9,970	330	C15	41,500	43,150	1,650
E3	9,970	11,430	1,460	E29	43,150	43,212	62
V2	11,430	13,650	2,220	V8	43,212	44,352	1,140
E22	13,650	14,250	600	E14	44,352	45,200	848
C4	14,250	15,675	1,425	V9	45,200	45,500	300
E4	15,675	15,700	25	E30	45,500	45,550	50
C25	15,700	16,100	400	C16	45,550	46,070	520
E5	16,100	17,435	1,335	E31	46,070	46,077	7
V3	17,435	19,373	1,938	V10	46,077	46,149	72
E6	19,373	19,600	227	E32	46,149	46,150	1
C5	19,600	20,010	410	C17	46,150	46,387	237
T3	20,010	20,441	431	T7	46,387	50,600	4,213
C6	20,441	20,580	139	C18	50,600	51,300	700
E23	20,580	20,611	31	T8	51,300	51,500	200
V4	20,611	21,451	840	C19	51,500	51,750	250
E7	21,451	22,050	599	E15	51,750	52,653	903
C26	22,050	22,360	310	V11	52,653	52,932	279
E8	22,360	22,750	390	C20	52,932	54,654	1,722
C27	22,750	23,550	800	T9	54,654	57,620	2,966
E9	23,550	24,000	450	C21	57,620	57,800	180
C7	24,000	24,450	450	E16	57,800	58,100	300
T4	24,450	24,737	287	V12	58,100	59,084	984
C8	24,737	25,250	513	E17	59,084	59,540	456
E24	25,250	25,297	47	V13	59,540	60,940	1,400
V5	25,297	25,427	130	E18	60,940	61,525	585
E25	25,427	25,500	73	C22	61,525	63,050	1,525
C9	25,500	25,754	254	E19	63,050	64,300	1,250
T5	25,754	26,239	485	C29	64,300	64,550	250
C10	26,239	27,300	1,061	T10	64,550	64,700	150
E26	27,300	28,050	750	C30	64,700	64,900	200
C11	28,050	28,637	587	E20	64,900	65,500	600
T6	28,637	30,400	1,763	C23	65,500	66,800	1,300
C12	30,400	30,900	500	E33	66,800	66,900	100
E10	30,900	33,500	2,600	C24	66,900	69,525	2,625
C28	33,500	33,650	150	E21	69,525	69,790	265
E11	33,650	33,940	290				

C.2.1 Ground Parameter States and Ground Classes

Ground parameter states and ground classes of all tunnels in the four alignments are summarized in Table C.5. Note that the geologist assigned ground class medium to tunnels T6 in alignment A-a (RAVE 2006d), which is also tunnel T5 in alignment A-c, although according to its ground parameters states (Table C.5) and the ground class definition (Table 7.29), they should be assigned ground class poor. Here, the ground class poor is assigned to tunnel T6 in alignment A-a and tunnel T5 in alignment A-c.

C.2.2 Construction Methods

The construction method, the tunnel length, the construction cost and time of all tunnels in the four alignments are summarized in Table C.6.

Since alignments A-a and A-c coincide in the last part of the Porto-Braga axis, tunnels T4, T5 and T6 in alignment A-a are the same as tunnels T3, T4 and T5 in alignment A-c. Similarly, tunnels T1, T2, T10 and T12 in alignment B-a are the same as tunnels T1, T2, T8 and T10 in alignment B-b.

C.3 Viaducts

The construction of the viaducts in the four alignments of the new Portuguese high speed rail line was presented in chapter 7, section 7.3. In the following, the activity networks and the calculation of the construction cost and time are discussed in more detail.

C.3.1 Activity Networks

The activity network of a viaduct depicts the sequence of activities during the construction of the viaduct. The activity networks of the 31 viaducts in the four alignments are presented in the following (Figures C-1 to C-31). The notation differs slightly from the notation in chapter 7 (Table C.7).

Table C.5: List of tunnels in the four alignments with the ground parameters states and the ground classes. GP1 is ground parameter overburden, GP2 granit, GP3 meta-sediment, GP4 perpendicular faults, GP5 oblique faults, GP6 acute-angled faults, GP7 parallel faults, and GP8 transversal faults.

Align.	tunnel	ground parameters								ground class
		GP1	GP2	GP3	GP4	GP5	GP6	GP7	GP8	
A-a	T1	$< 2D$	yes	no	no	no	no	no	no	reduced overburden
	T2	$\geq 2D$	yes	no	yes	yes	no	yes	no	poor
	T3	$\geq 2D$	no	yes	no	yes	no	no	no	medium
	T4	$\geq 2D$	yes	no	no	no	no	no	no	good
	T5	$\geq 2D$	yes	no	yes	yes	no	yes	no	poor
	T6	$\geq 2D$	yes	yes	no	yes	no	yes	no	poor
A-c	T1	$< 2D$	yes	no	no	no	no	no	no	reduced overburden
	T2	$\geq 2D$	yes	yes	no	yes	no	no	no	medium
	T3	$\geq 2D$	yes	no	no	no	no	no	no	good
	T4	$\geq 2D$	yes	no	yes	yes	no	yes	no	poor
	T5	$\geq 2D$	yes	yes	no	yes	no	yes	no	poor
B-a	T1	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T2	$\geq 2D$	yes	yes	yes	no	no	no	no	medium
	T3	$\geq 2D$	no	yes	yes	yes	no	no	no	poor
	T4	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T5	$\geq 2D$	no	yes	no	yes	yes	no	no	poor
	T6	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T7	$\geq 2D$	no	yes	no	no	no	no	no	medium
	T8	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T9	$\geq 2D$	yes	yes	no	yes	yes	yes	no	poor
	T10	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T11	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T12	$\geq 2D$	yes	yes	no	no	no	yes	no	poor
B-b	T1	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T2	$\geq 2D$	yes	yes	yes	no	no	no	no	medium
	T3	$\geq 2D$	yes	yes	yes	yes	no	no	no	medium
	T4	$\geq 2D$	no	yes	nk	nk	nk	nk	nk	poor
	T5	$\geq 2D$	yes	no	no	yes	yes	yes	no	poor
	T6	$\geq 2D$	yes	yes	no	yes	yes	no	yes	medium
	T7	$\geq 2D$	no	yes	no	yes	yes	no	no	poor
	T8	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden
	T9	$\geq 2D$	yes	yes	no	no	no	yes	no	poor
	T10	$< 2D$	nk	nk	nk	nk	nk	nk	nk	reduced overburden

Table C.6: Construction method, tunnel length, construction cost and time of all tunnels in the four alignments.

Alignment	tunnel	construction method	length [m]	cost [10 ⁶ euro]	time [days]
A-a	T1	cut & cover	996	10.6	374
	T2	NATM slow	1,455	21.6	1,200
	T3	NATM medium	682	8.2	409
	T4	NATM fast	500	5.1	225
	T5	NATM slow	2,650	39.3	2,186
	T6	NATM medium	3,657	44.1	2,194
A-c	T1	cut & cover	990	10.5	371
	T2	NATM medium	1,711	20.6	1,027
	T3	NATM fast	500	5.1	225
	T4	NATM slow	2,650	39.3	2,186
	T5	NATM medium	3,657	44.1	2,194
B-a	T1	cut & cover	150	1.6	56
	T2	NATM medium	2,243	27.0	673
	T3	NATM slow	1,503	22.3	620
	T4	cut & cover	260	2.8	97
	T5	NATM slow	3,021	44.8	1,247
	T6	cut & cover	200	2.1	75
	T7	NATM medium	1,311	15.8	394
	T8	cut & cover	200	2.1	75
	T9	NATM slow	4,555	67.5	1,879
	T10	cut & cover	200	2.1	75
	T11	NATM slow	2,249	33.3	928
	T12	cut & cover	150	1.6	56
B-b	T1	cut & cover	150	1.6	56
	T2	NATM medium	2,243	27.0	673
	T3	NATM medium	431	5.2	259
	T4	NATM slow	287	4.3	237
	T5	NATM slow	485	7.2	400
	T6	NATM medium	1,763	21.2	529
	T7	NATM slow	4,213	62.5	1,738
	T8	cut & cover	200	2.1	75
	T9	NATM slow	2,966	44.0	1,223
	T10	cut & cover	150	1.6	56

Table C.7: Notation of the activities in the viaduct activity networks.

Activity	notation here	notation in chapter 7
abutment	V_A	a
deck section	V_d or V_D	ds
pier	V_p or V_P or V_pP	p
foundation footing	V_f	ff
foundation pile set		fp
sloped pier	V_sp or V_spSP	sp
technical block	V_b	tb
finishing	V_fi	fi
pre-work	V_pw	pw
pre-deck	V_pd	pd
construction method	V_cp or V_CP	cm
dummy	d	d

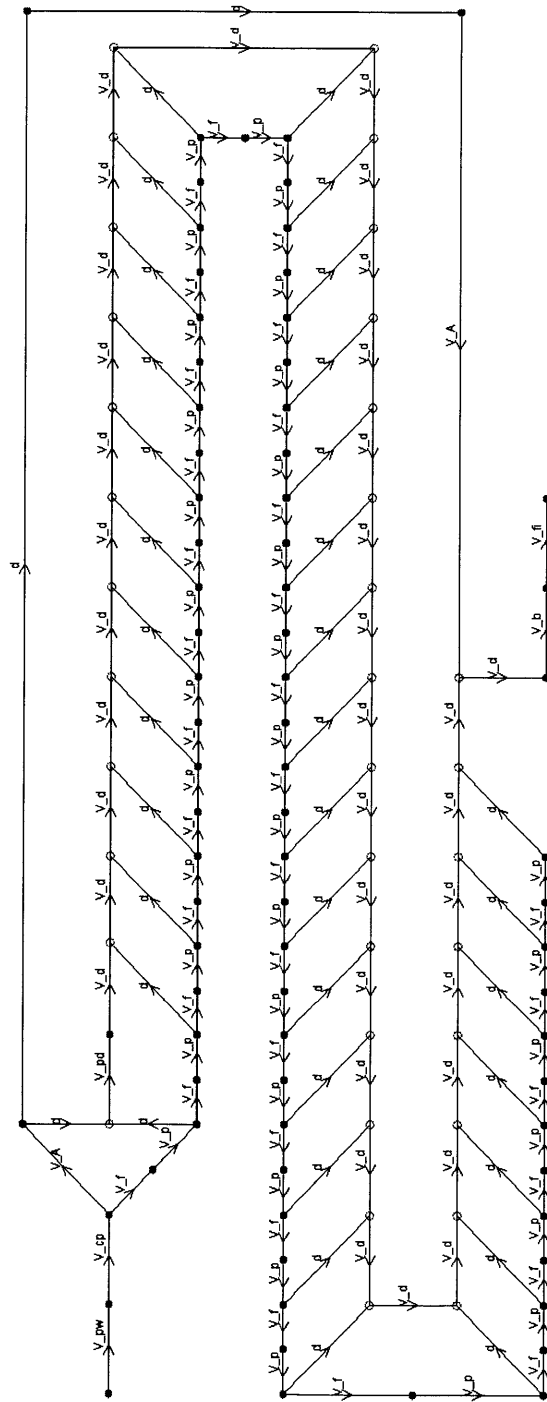


Figure C-1: Activity network of viaduct Moinhos in alignment A-a.

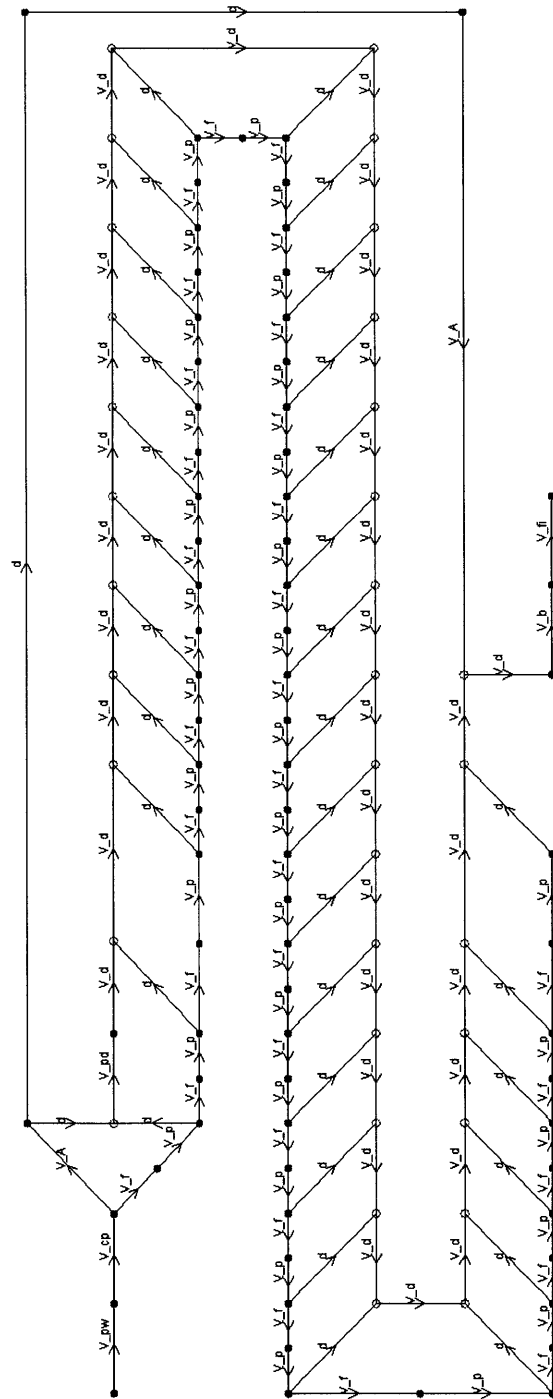


Figure C-2: Activity network of viaduct Aldeia in alignment A-a.

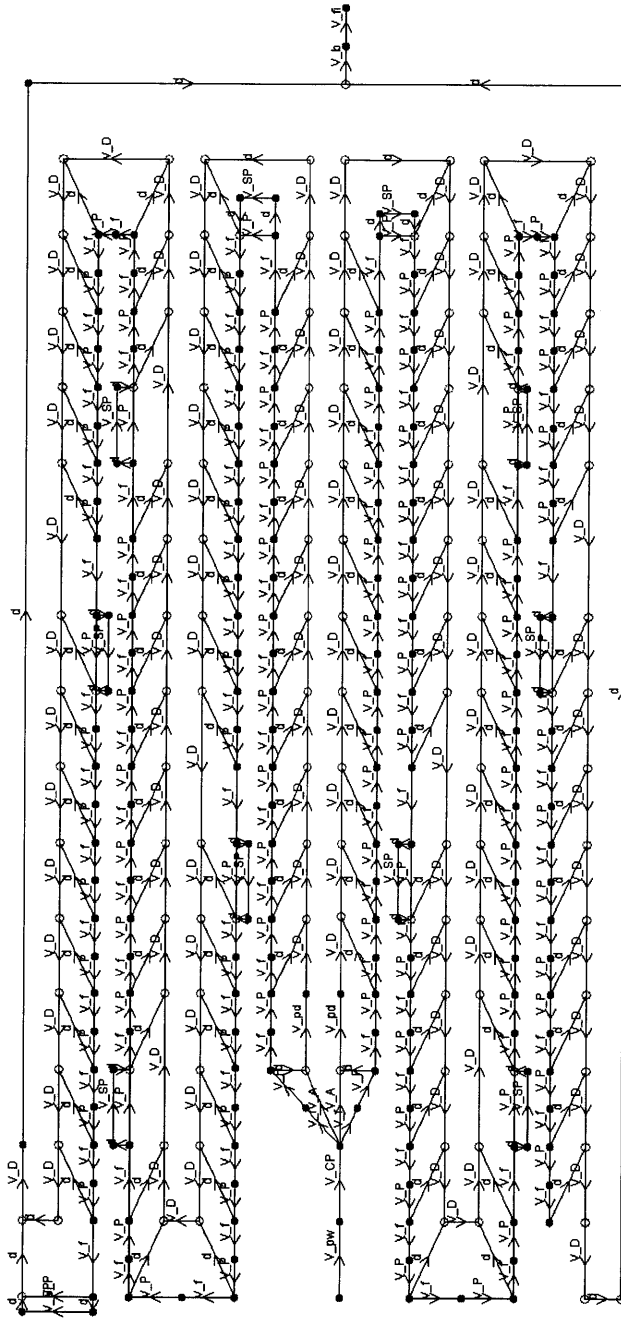


Figure C-3: Activity network of viaduct Ave in alignment A-a. It is the only viaduct constructed from both ends due to its extended length.

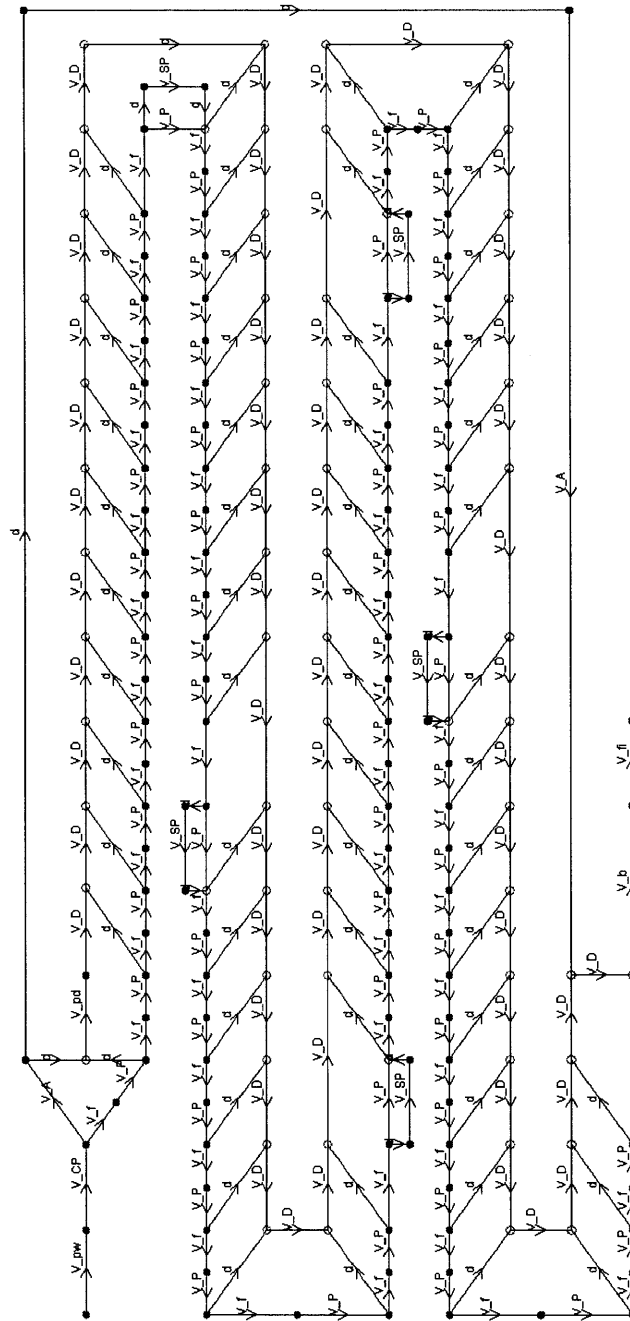


Figure C-4: Activity network of viaduct Este in alignment A-a.

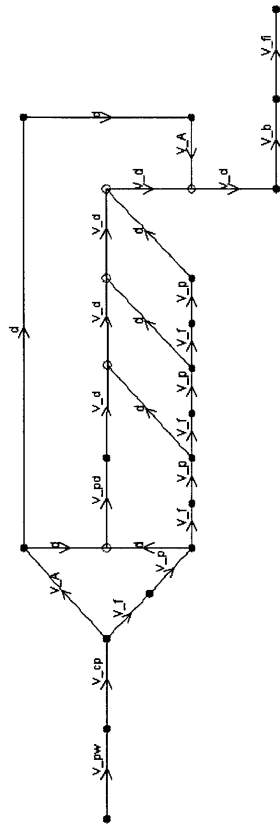


Figure C-6: Activity network of viaduct Laje in alignment A-c.

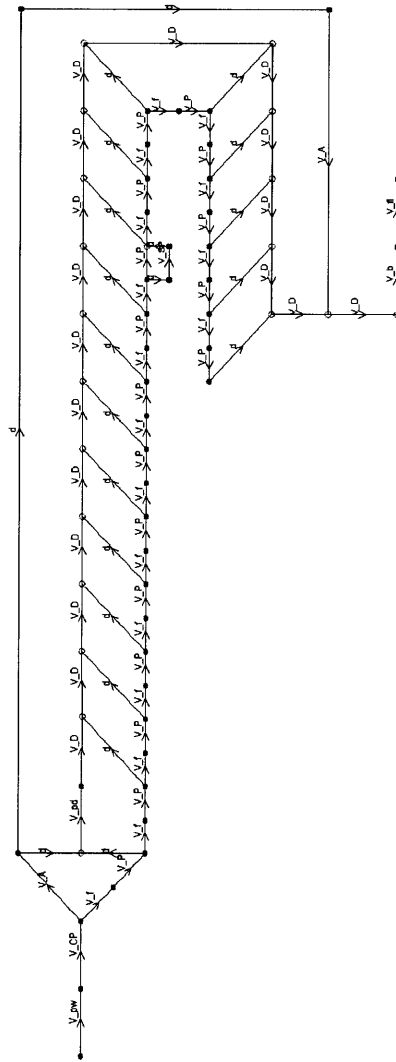


Figure C-7: Activity network of viaduct Ave in alignment A-c.

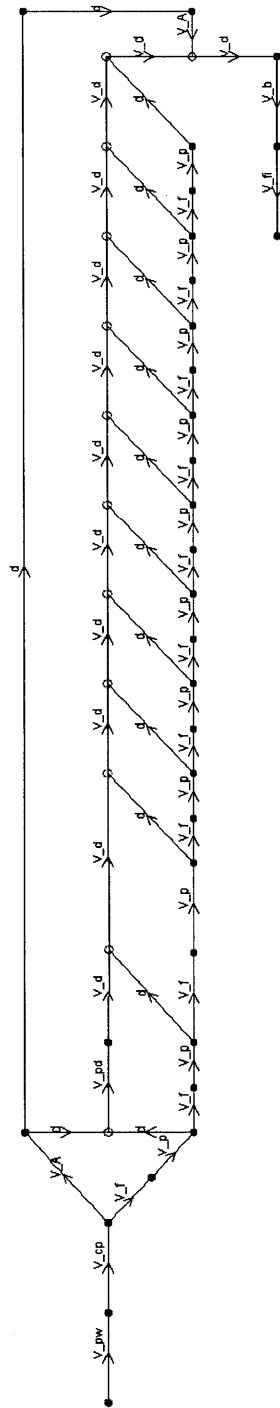


Figure C-8: Activity network of viaduct Este in alignment A-c.

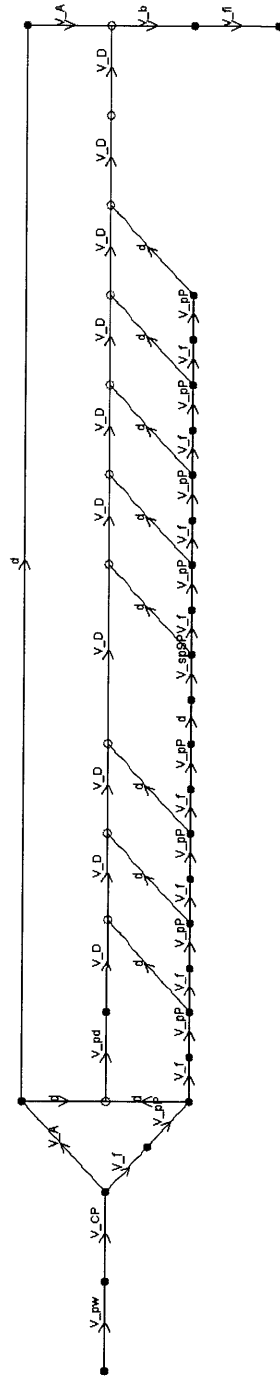


Figure C-9: Activity network of viaduct Cambeses in alignment A-c.

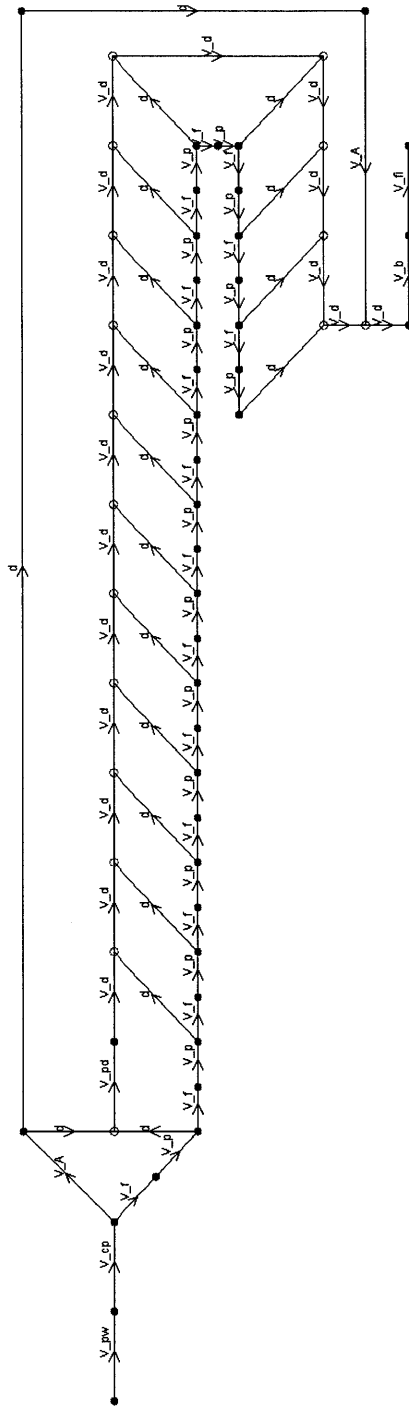


Figure C-10: Activity network of viaduct Mau in alignment B-a.

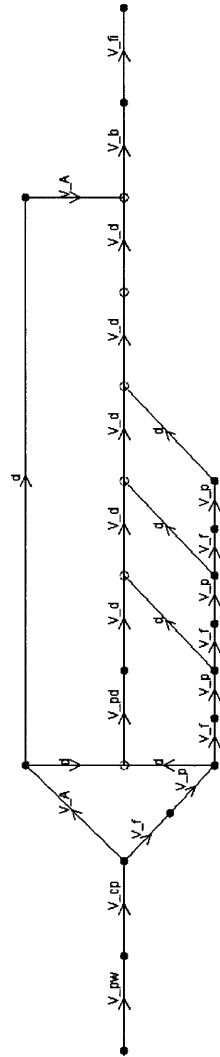


Figure C-12: Activity network of viaduct Neiva in alignment B-a.

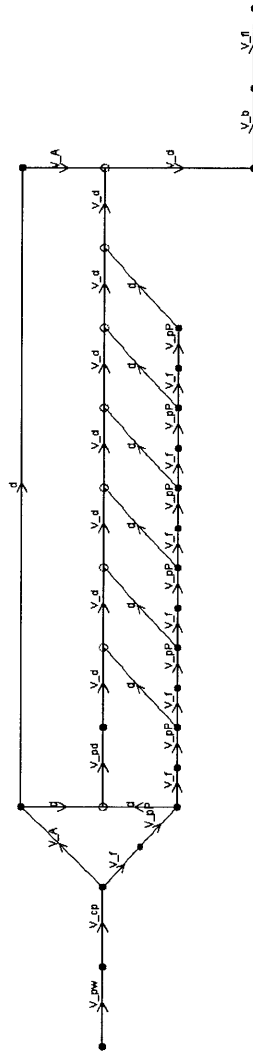


Figure C-13: Activity network of viaduct Outeiro in alignment B-a.

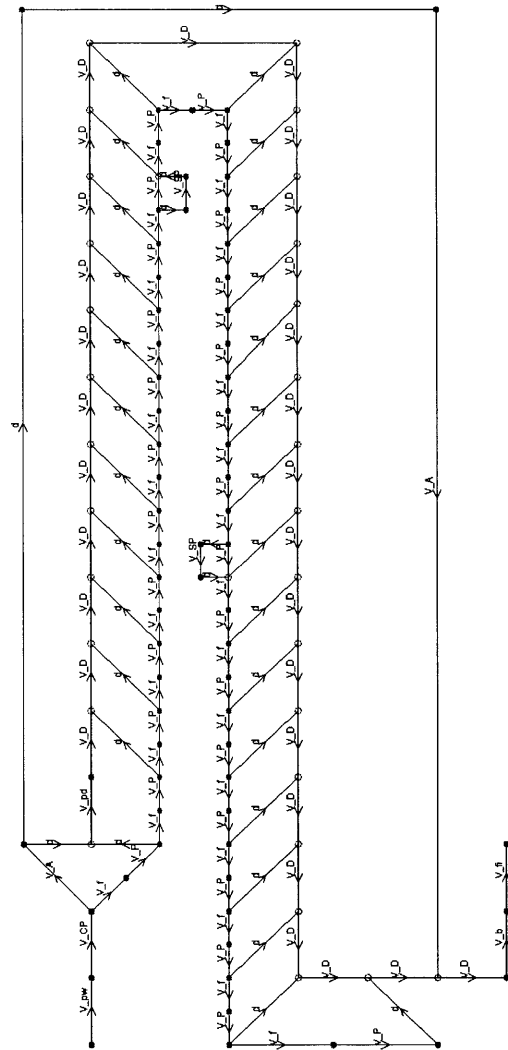


Figure C-14: Activity network of viaduct Lima in alignment B-a.

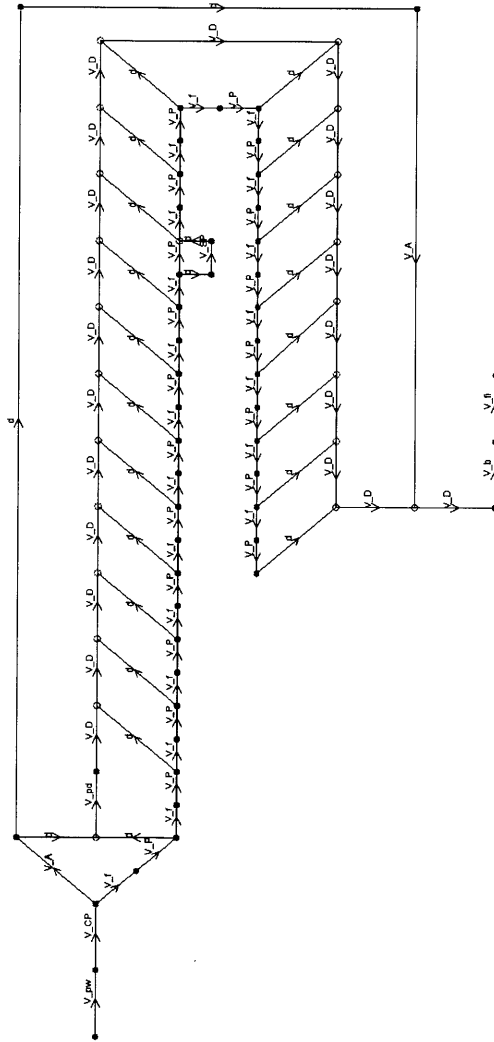


Figure C-15: Activity network of viaduct Labruja in alignment B-a.

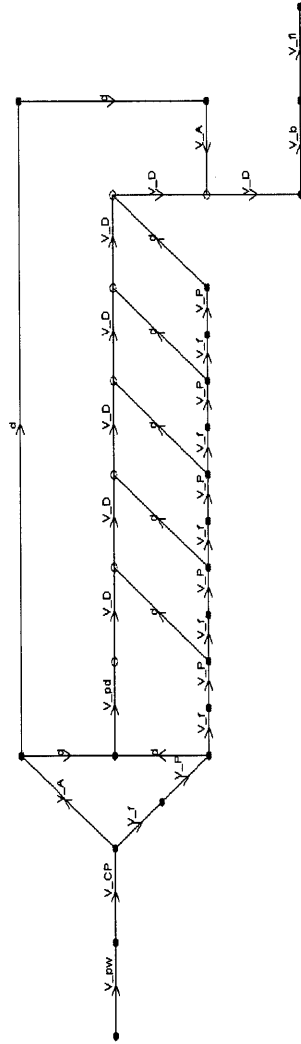


Figure C-16: Activity network of viaduct Coura in alignment B-a.

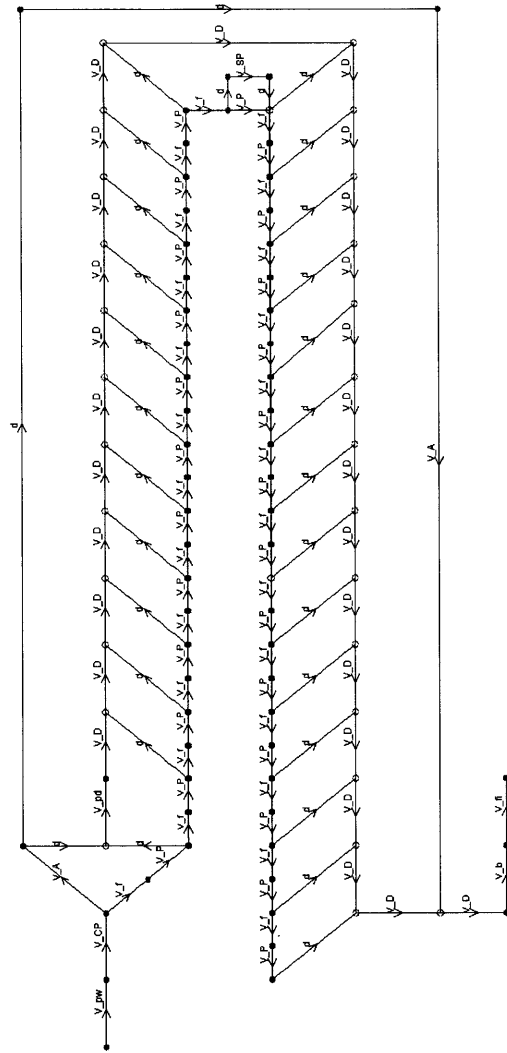


Figure C-18: Activity network of viaduct Boriz in alignment B-a.

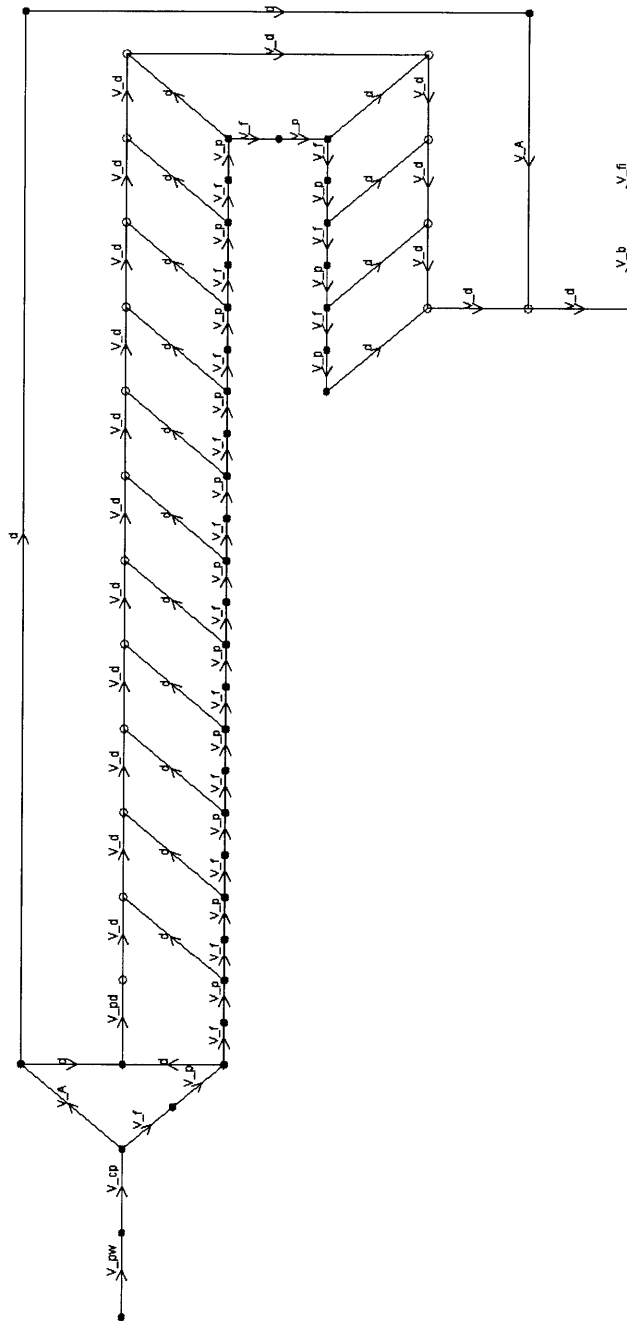


Figure C-19: Activity network of viaduct Mau in alignment B-b.

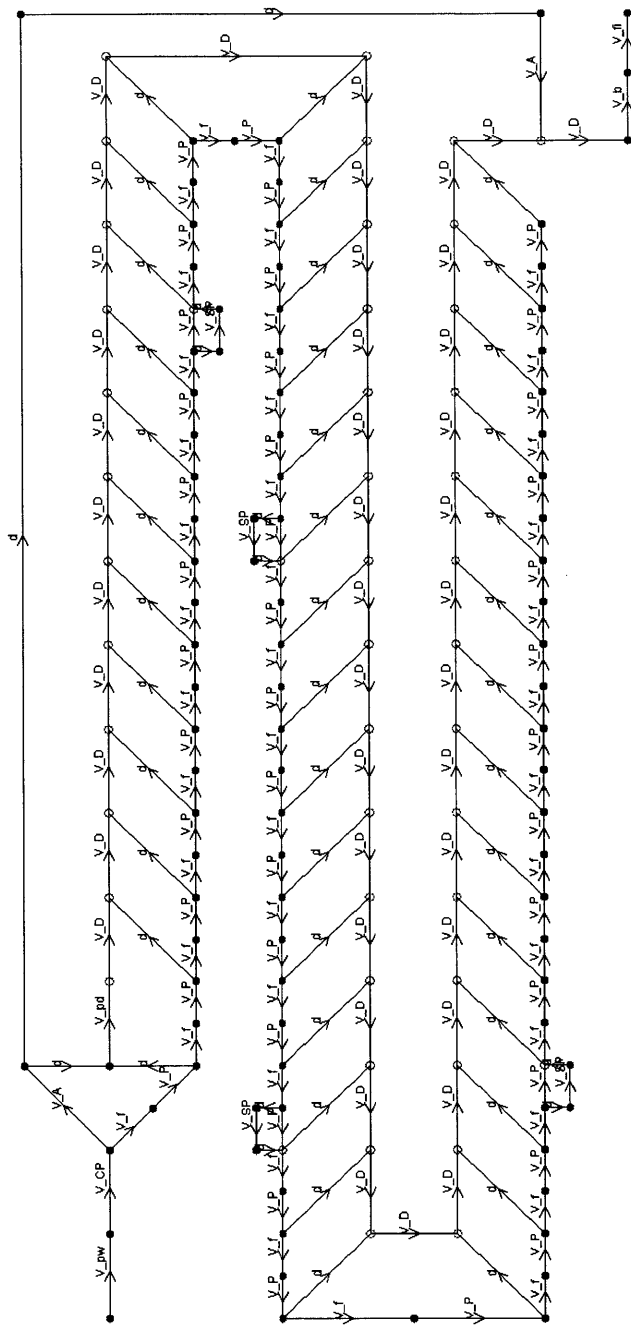


Figure C-20: Activity network of viaduct Cavado in alignment B-b.

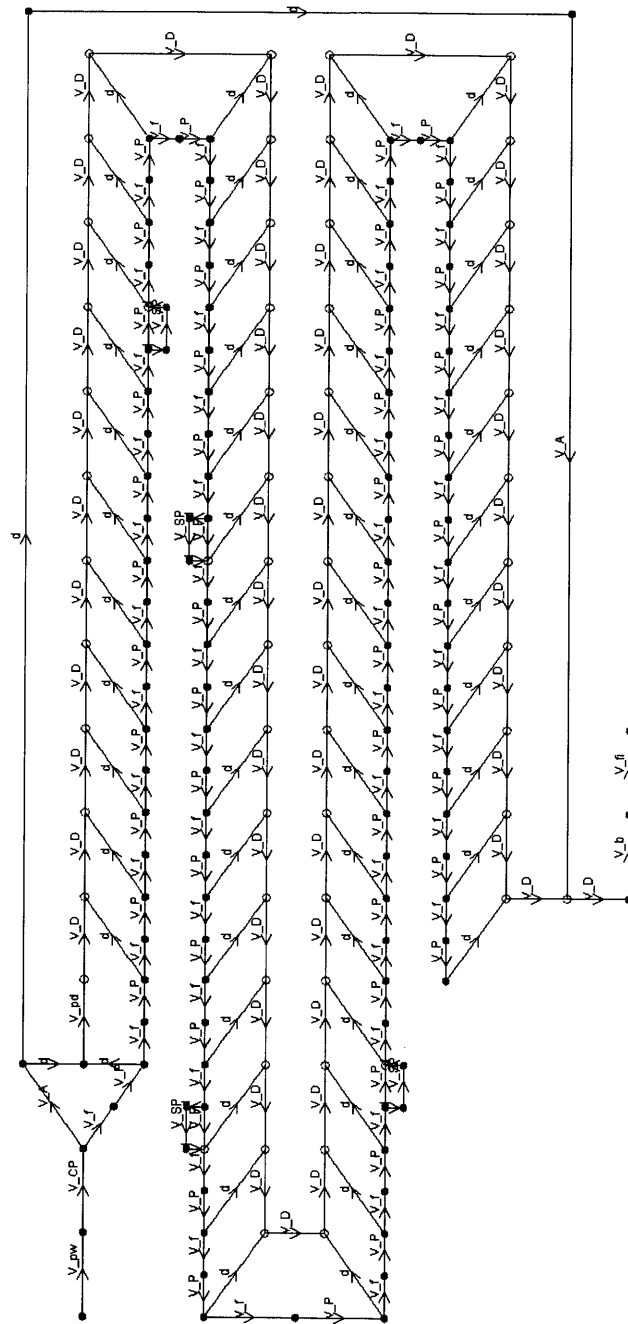


Figure C-21: Activity network of viaduct Leiras in alignment B-b.

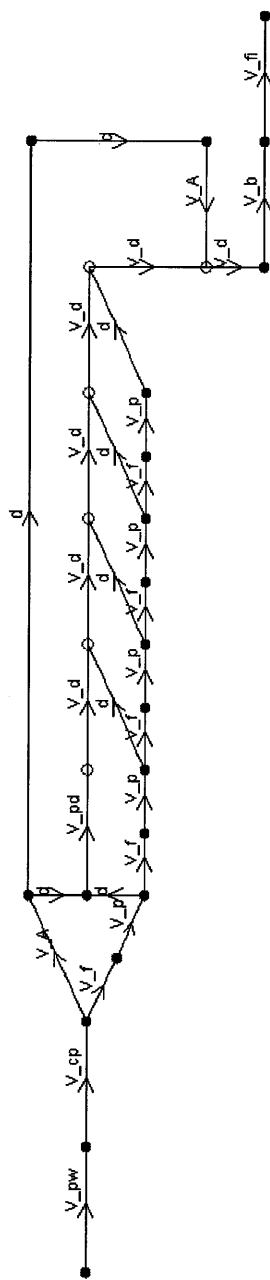


Figure C-23: Activity network of viaduct Pombarinhos in alignment B-b.

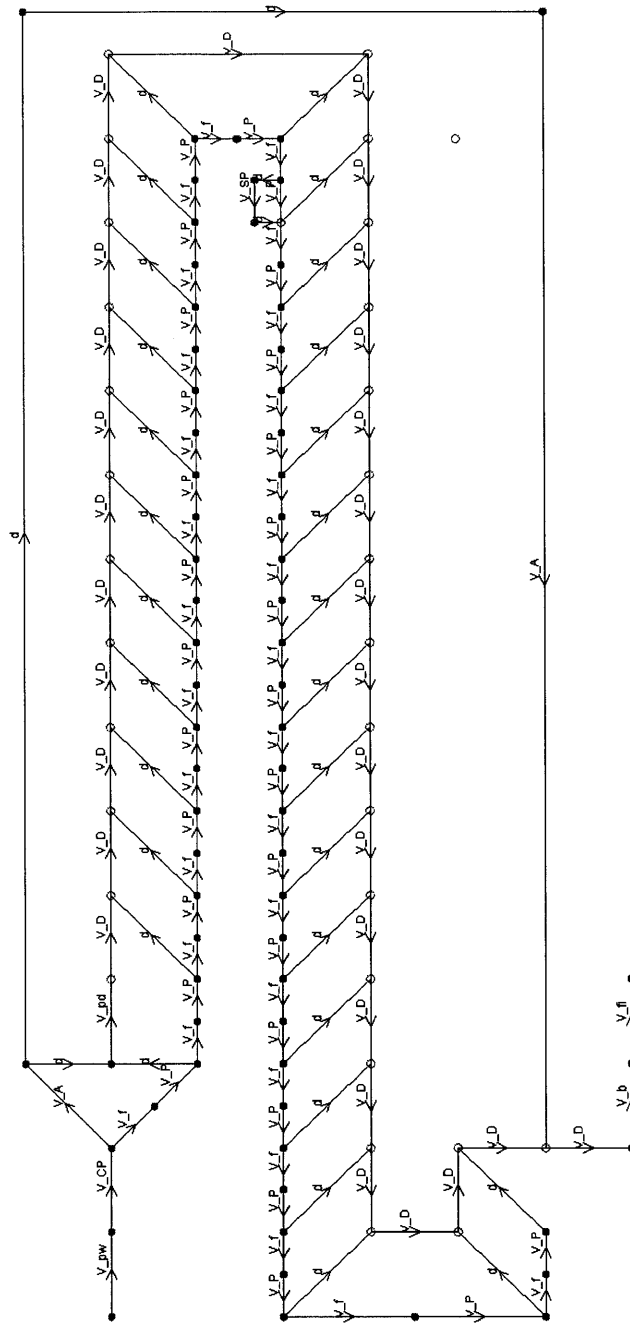


Figure C-26: Activity network of viaduct Cabracao in alignment B-b.

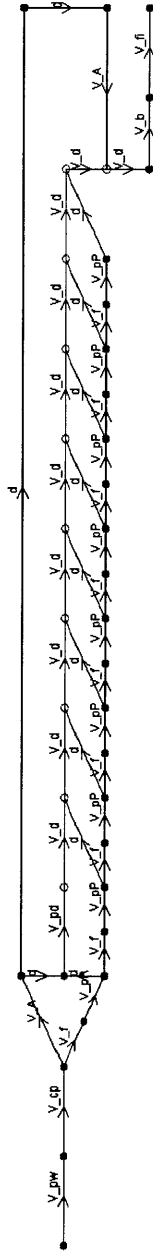


Figure C-27: Activity network of viaduct Bouca in alignment B-b.

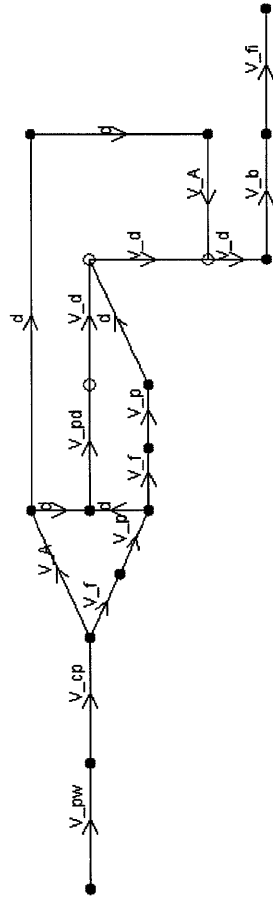


Figure C-28: Activity network of viaduct Formigoso in alignment B-b.

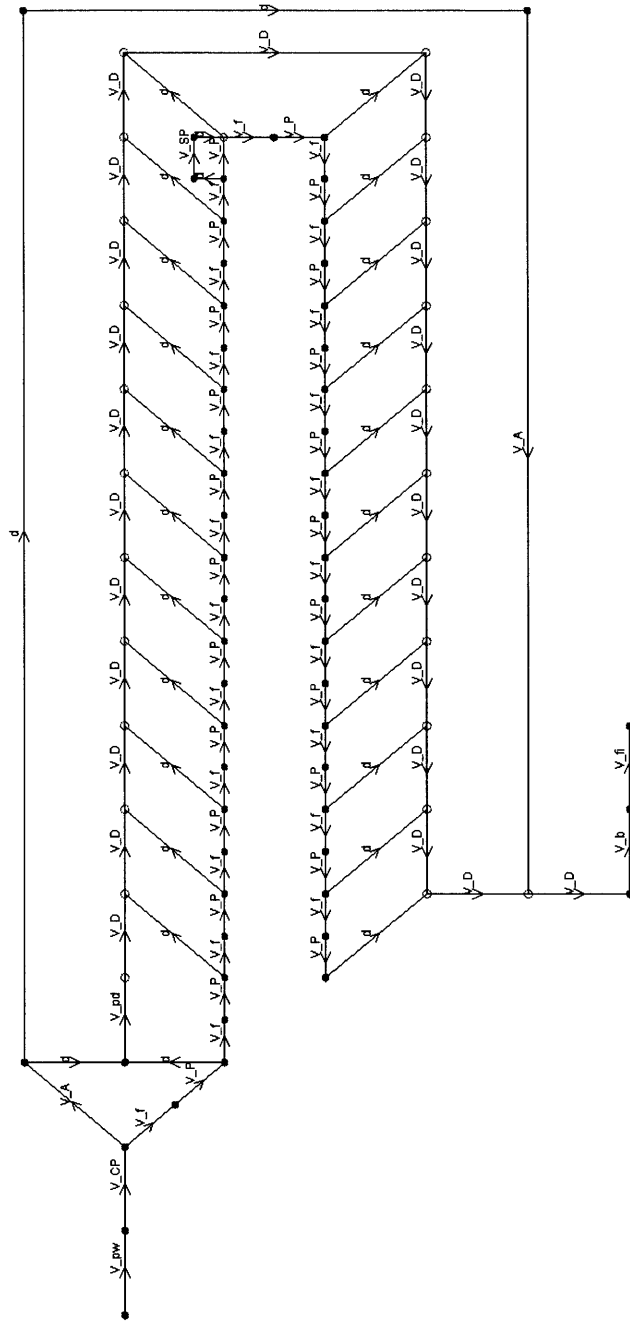


Figure C-30: Activity network of viaduct Bogalheiro in alignment B-b.

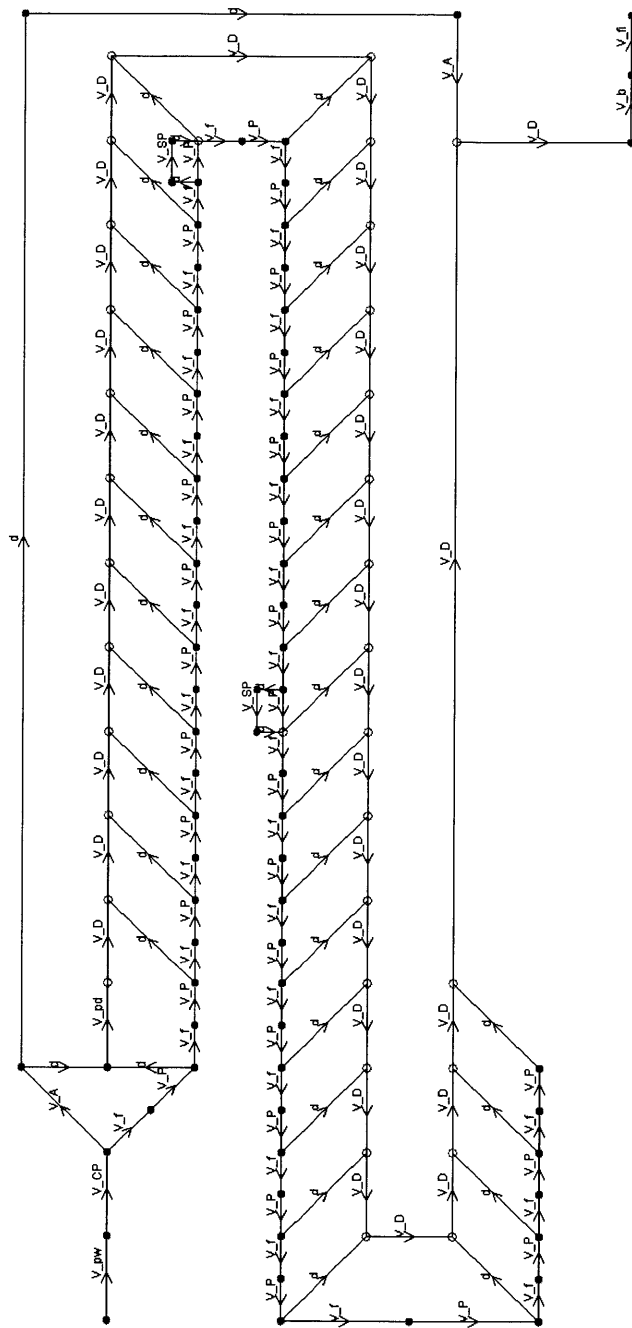


Figure C-31: Activity network of viaduct Boriz in alignment B-b.

C.3.2 Construction Cost and Time

The geometric parameters, the cost and the time of all viaducts in the four alignments are summarized in Table C.8. For the construction cost and time of the following viaducts one has to note:

- viaduct Ave (alignment A-a) is almost twice as long as the second longest viaduct of the four alignments. To reduce the construction time, the viaduct is constructed starting from both ends, allowing to halve the deck construction time. The construction of the viaduct starting from both ends is reflected in the activity network (Figure C-3), as well as in the additional cost for two gantries (double cost for activity construction method).
- viaduct Cambeses (alignment A-a and A-c). Although the span length is $l = 42$ m and the pier height is $h = 20$ m, the pier cost expression for ($l \leq 35$ m, $h > 15$ m) is used (Leite 2009).

The calculated costs differ from the construction costs in the source (RAVE 2006c, RAVE 2008c) since they are estimated with cost expressions provided by the design engineer. The discrepancy is quantified with the ratio between the calculated cost and the cost in the sources (Table C.9). Most of the ratios are approximately 100%, indicating a good agreement. There is a preponderance of underestimations ($< 100\%$) over overestimations ($\geq 100\%$). The extremes are viaduct Neiva (V3) in alignment B-a with a ratio of 66% and viaduct Formigoso (V10) in alignment B-b with a ratio of 143%. The ratios significantly deviating from 100% were discussed with the design engineer. Through a detailed analysis of the viaduct drawings (RAVE 2006c, RAVE 2008c), the differences could be explained with special structural elements, which the cost expressions do not take into account (Leite 2009):

- Viaduct Ave (alignment A-c) includes a special structure consisting of arched piers for crossing the river Ave.
- Viaduct Lima (alignments B-a) includes one special 60m long span and a special structure consisting of arched piers of considerable height for crossing the river Lima.
- Viaduct Neiva (alignment B-b) includes a special structure consisting of a longer and deeper span to cross a highway, which requires a pushing apparatus and an auxiliary metallic structure.

- Viaduct Formigoso (alignment B-b) is the shortest viaduct (72m) in the four alignments, for which a solution with cheaper abutments, no bearings or expansion joints was preferred.

For the four viaducts with special structures, cost adjustments are introduced to better match the calculated costs with the costs in the sources. The magnitude of the cost adjustment is equal to the difference between the calculated cost and the cost in the sources (Table C.10).

C.4 Cuts and Embankments

The construction of the cuts and embankments in the four alignments of the new Portuguese high speed rail line was presented in chapter 7, section 7.3. In the following, the volumes and the construction cost and time of the cuts and embankments are discussed in more detail.

C.4.1 Volumes of the Cuts and Embankments

The earthwork (cuts and embankments) construction cost and time are functions of the cost per unit volume, the production rates and the volumes of the cuts and embankments. These are determined by the difference between the natural landscape and the new rail line profile and obtained by means of a DTM system (Menezes 2010). Since they are not available from the sources (RAVE 2006b, RAVE 2008b), they are estimated based on the geometric parameters of the cuts and embankments (Tables C.11, C.12 and C.13).

In the following, the calculations of the cut volumes and the embankment volumes are explained. The cut volume includes the clearing and the excavating volumes, which are removed, and the capping and the sub-ballast volumes, which are added (Figure C-32a).

- The clearing volume is assumed to be a prism (Table C.12a) with length L , thickness t_{clear} , and width varying between the cut width at the bottom, w_c , and the maximal width, w_{max} .
- The excavating volume, which is removed after clearing, is equal by approximation to the cut volume minus the clearing volume. The cut volume consists of two central wedges and four lateral pyramids (Table C.12b). The wedge base width is equal to the cut width at the bottom, w_c , the wedge base height is equal to the cut depth, D , the wedge height is equal

Table C.8: Construction cost and time of the viaducts in the four alignments. They are calculated with cost expressions (Table 7.13) and time expressions (Table 7.16) based on geometric parameters. *Delta* is the cost adjustment (Table C.10). Cambeses is the same viaduct in alignments A-a and A-c, similarly to Mau and Coura in alignments B-a and B-b. On the other hand, Ave and Este in alignments A-a and A-c, and Cavado, Neiva, Lima and Boriz in alignments B-a and B-b are different viaducts.

Al.	viaduct	L [m]	l [m]	h [m]	d [m]	α [%]	N –	S –	Δ [10 ⁶ euro]	cost	time [days]
A-a	V1 Moinhos	765	25	15	11	70	35	0		6.12	655
	V2 Aldeia	700	24	14	8	70	33	0		5.38	626
	V3 Ave	4,852	51	43	10	40	104	11		66.39	978
	V4 Este	2,700	52	36	11	40	56	5		33.53	1,004
	V5 Cambeses	395	42	20	16.5	100	10	1		4.65	328
A-c	V1 Laje	100	23	6	0	0	5	0		1.25	225
	V2 Ave	920	53	35	10	80	18	1	+2.90	14.30	447
	V3 Este	250	25	11	4	10	12	0		2.22	326
	V4 Cambeses	395	42	20	16.5	100	10	1		4.65	328
B-a	V1 Mau	330	20	10	12	100	17	0		2.77	397
	V2 Cavado	2,520	55.4	20	15	100	49	5		28.7	903
	V3 Neiva	120	26	10	10	100	5	0		1.49	226
	V4 Outeiro	205	27.5	20	0	0	8	0		2.22	269
	V5 Lima	1,620	60	40	7	50	29	2	+5.99	27.31	611
	V6 Labruja	1,014	52	38	0	0	21	1		12.40	490
	V7 Coura	279	43	16	5	30	7	0		3.06	284
	V8 Reguengo	300	26	20	4	30	12	0		2.90	326
	V9 Boriz	1,020	40	28	18	90	27	1		12.19	574
B-b	V1 Mau	330	20	10	12	100	17	0		2.77	397
	V2 Cavado	2,220	56.9	20	15	100	42	4		25.49	801
	V3 Leiras	1,938	38	25	5	100	54	4		19.78	965
	V4 Neiva	840	30	13	12	100	28	0	+3.27	9.73	558
	V5 Pombarinhos	130	23.5	13	3	20	6	0		1.52	240
	V6 Lima	2,848	47.5	30	14	80	64	6		34.19	1,118
	V7 Moinho Velho	111	25	14	0	0	5	0		1.43	226
	V8 Cabracao	1,140	40	27	4	20	30	1		11.58	618
	V9 Bouca	300	31.5	20	0	0	10	0		2.91	298
	V10 Formigoso	72	30	8	0	0	3	0	–0.35	0.81	197
	V11 Coura	279	43	16	5	30	7	0		3.06	284
	V12 Bogalheiro	984	44	20	4	30	24	1		9.39	532
	V13 Boriz	1,400	46.5	28	18	90	32	2		17.07	650

Table C.9: The calculated construction costs of the viaducts and the construction costs in the sources are compared.

Alignment	viaduct	calculated cost [10 ⁶ euro]	(RAVE 2006c) [10 ⁶ euro]	ratio [%]
A-a	V1 Moinhos	6.12	6.30	97
	V2 Aldeia	5.38	5.70	94
	V3 Ave	66.39	71.20	93
	V4 Este	33.53	37.80	89
	V5 Cambeses	4.65	4.20	111
A-c	V1 Laje	1.25	1.10	114
	V2 Ave	11.40	14.30	80
	V3 Este	2.22	2.10	106
	V4 Cambeses	4.65	4.20	111
B-a	V1 Mau	2.77	3.45	80
	V2 Cavado	28.7	30.38	95
	V3 Neiva	1.49	1.41	106
	V4 Outeiro	2.22	2.27	97
	V5 Lima	21.32	27.31	78
	V6 Labruja	12.40	15.56	80
	V7 Coura	3.06	3.27	94
	V8 Reguengo	2.90	3.33	87
	V9 Boriz	12.19	13.30	92
B-b	V1 Mau	2.77	3.45	80
	V2 Cavado	25.49	26.89	95
	V3 Leiras	19.78	23.63	84
	V4 Neiva	6.46	9.73	66
	V5 Pombarinhos	1.52	1.31	116
	V6 Lima	34.19	41.15	83
	V7 Moinho Velho	1.43	1.30	110
	V8 Cabracao	11.58	16.06	72
	V9 Bouca	2.91	3.43	85
	V10 Formigoso	1.16	0.81	143
	V11 Coura	3.06	3.27	94
	V12 Bogalheiro	9.39	11.38	82
	V13 Boriz	17.07	18.31	93

Table C.10: Viaduct cost adjustments. Due to special structures in the viaducts Ave, Lima, Neiva, and Formigoso, cost adjustments are introduced to better match the calculated costs to the costs in the sources.

Viaduct	alignment	calculated cost [10 ⁶ euro]	cost adjustment, Δ [10 ⁶ euro]	cost in the sources [10 ⁶ euro]	
V3	Ave	A-c	11.40	+2.90	14.30
V5	Lima	B-a	21.32	+5.99	27.31
V4	Neiva	B-b	6.46	+3.27	9.73
V10	Formigoso	B-b	1.16	-0.35	0.81

Table C.11: The geometric parameters of cuts and embankments determine the volumes of these.

Structure	parameter	parameter description
cut (Table C.12)	D	cut depth
	$w_c = 20$ m	cut width at bottom
	e	percentage of cut volume excavated with blasting
embankment (Table C.13)	H	embankment height
	$w_e = 14$ m	embankment width at top
	t	thickness of embankment improving layer
	S	length of embankment improving layer
cut and embankment (Tables C.12 and C.13)	L	length
	w_{max}	maximal width
	α	slope steepness
	t_{clear}	thickness of clearing layer
	$t_{capping} = 0.6$ m	thickness of capping layer
	$t_{sub-ballast} = 0.3$ m	thickness of sub-ballast layer
	$w_{capping} = 14$ m	width of capping layer
$w_{sub-ballast} = 14$ m	width of sub-ballast layer	

Table C.12: Expressions to calculate the cut volume based on geometric parameters (Table C.11).
a) clearing volume, b) excavating volume, c) capping volume, and d) sub-ballast volume.

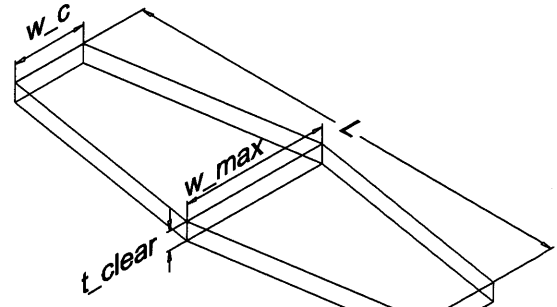
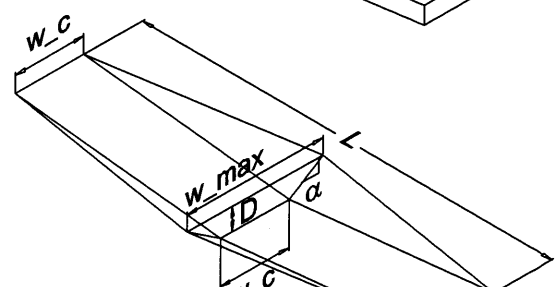
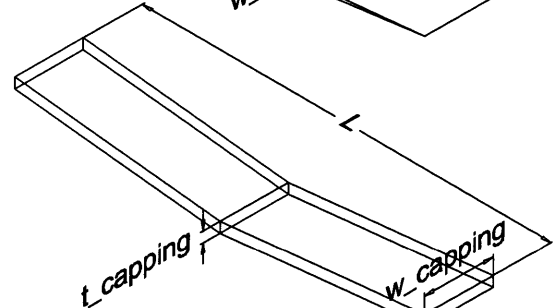
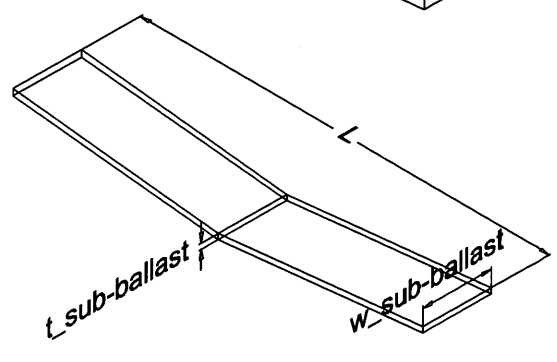
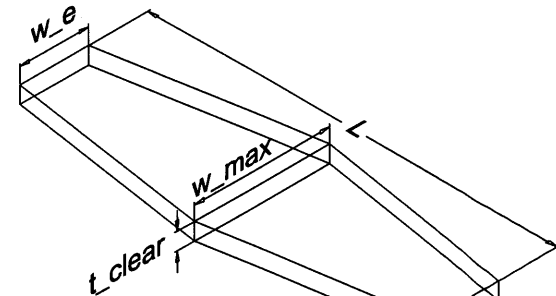
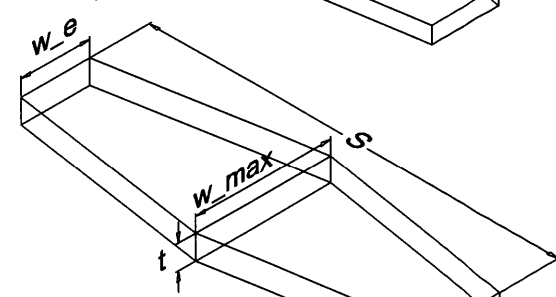
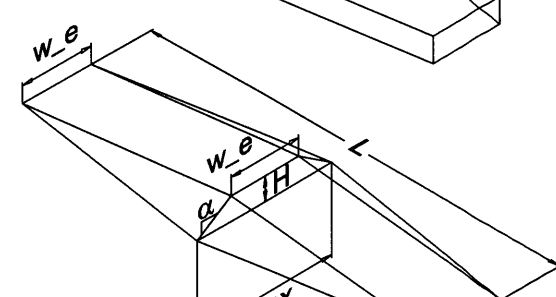
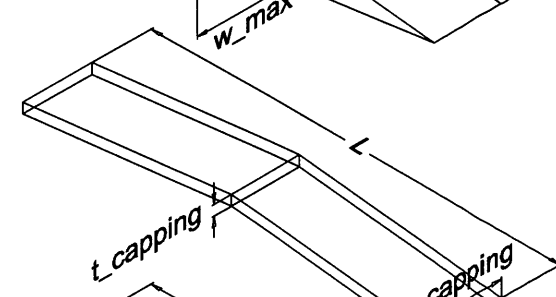
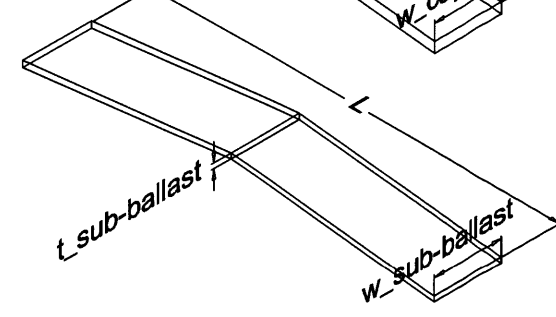
<p>a)</p> 	<p>Clearing volume: $L \times t_{clear} \times \left(\frac{w_c + w_{max}}{2} \right)$</p>
<p>b)</p> 	<p>Cut volume (2 wedges + 4 pyramids): $2 \left(w_c \times D \times \frac{L}{2} \times \frac{1}{2} \right) +$ $4 \left(D \times \alpha D \times \frac{1}{2} \times \frac{L}{2} \times \frac{1}{3} \right)$ <p>simplified: $\frac{1}{6} D \times L (3 \times w_c + 2\alpha D)$</p> </p>
<p>c)</p> 	<p>Excavating volume: $\frac{1}{6} D \times L (3 \times w_c + 2\alpha D) -$ $L \times t_{clear} \times \left(\frac{w_c + w_{max}}{2} \right)$</p> <p>Capping volume: $w_{capping} \times t_{capping} \times L$</p>
<p>d)</p> 	<p>Sub-ballast volume: $w_{sub-ballast} \times t_{sub-ballast} \times L$</p>

Table C.13: Expressions to calculate the embankment volume based on geometric parameters (Table C.11). a) clearing volume, b) improving volume, c) filling volume, d) capping volume, and e) sub-ballast volume.

<p>a)</p> 	<p>Clearing volume:</p> $L \times t_{clear} \times \left(\frac{w_e + w_{max}}{2} \right)$
<p>b)</p> 	<p>Improving volume:</p> $S \times t \times \left(\frac{w_e + w_{max}}{2} \right)$
<p>c)</p> 	<p>Embankment volume (2 wedges + 4 pyramids):</p> $2 \left(w_e \times H \times \frac{L}{2} \times \frac{1}{2} \right) +$ $4 \left(H \times \alpha H \times \frac{1}{2} \times \frac{L}{2} \times \frac{1}{3} \right)$ <p>simplified:</p> $\frac{1}{6} H \times L (3 \times w_e + 2\alpha H)$
<p>d)</p> 	<p>Filling volume:</p> $\frac{1}{6} H \times L (3 \times w_e + 2\alpha H) +$ $L \times t_{clear} \times \left(\frac{w_e + w_{max}}{2} \right)$ <p>Capping volume:</p> $w_{capping} \times t_{capping} \times L$
<p>e)</p> 	<p>Sub-ballast volume:</p> $w_{sub-ballast} \times t_{sub-ballast} \times L$

to half of the cut length, $L/2$. The pyramid base width is equal to the cut depth, D , the pyramid base height is equal to αD , the pyramid height is equal to half of the cut length, $L/2$. If e % of the excavating volume is removed with blasting, $(1 - e)$ % is removed with mechanical means.

- The capping volume consists of two prisms (Table C.12c), whose width and thickness are $w_{capping}$ and $t_{capping}$, respectively. The prism length is approximated to half of the cut length, $L/2$.
- The sub-ballast volume also consists of two prisms (Table C.12d), whose width and thickness are $w_{sub-ballast}$ and $t_{sub-ballast}$, respectively. The prism length is also approximated to half of the cut length, $L/2$.

The embankment volume includes the clearing volume, which is removed, the improving volumes, which is replaced, and the filling, the capping and the sub-ballast volumes, which are added (Figure C-32b).

- The clearing volume is assumed to be a prism (Table C.13a) with length L , thickness t_{clear} , and width varying between the embankment width at the top, w_e , and the maximal width, w_{max} .
- The improving volume is assumed to be a prism (Table C.13b) with length S , thickness t , and width varying between the embankment width at top, w_e , and the maximal width, w_{max} .
- The filling volume, which is added after clearing and improving, is equal by approximation to the embankment volume plus the clearing volume. The embankment volume consists of two central wedges and four lateral pyramids (Table C.13c). The wedge base width is equal to the embankment width at the top, w_e , the wedge base height is equal to the embankment height, H , the wedge height is equal to half of the embankment length, $L/2$. The pyramid base width is equal to the embankment height, H , the pyramid base height is equal to αH , the pyramid height is equal to half of the embankment length, $L/2$.
- Capping and sub-ballast volumes are calculated the same way as for the cuts (compare Tables C.12c and C.13d, and C.12d and C.13e).

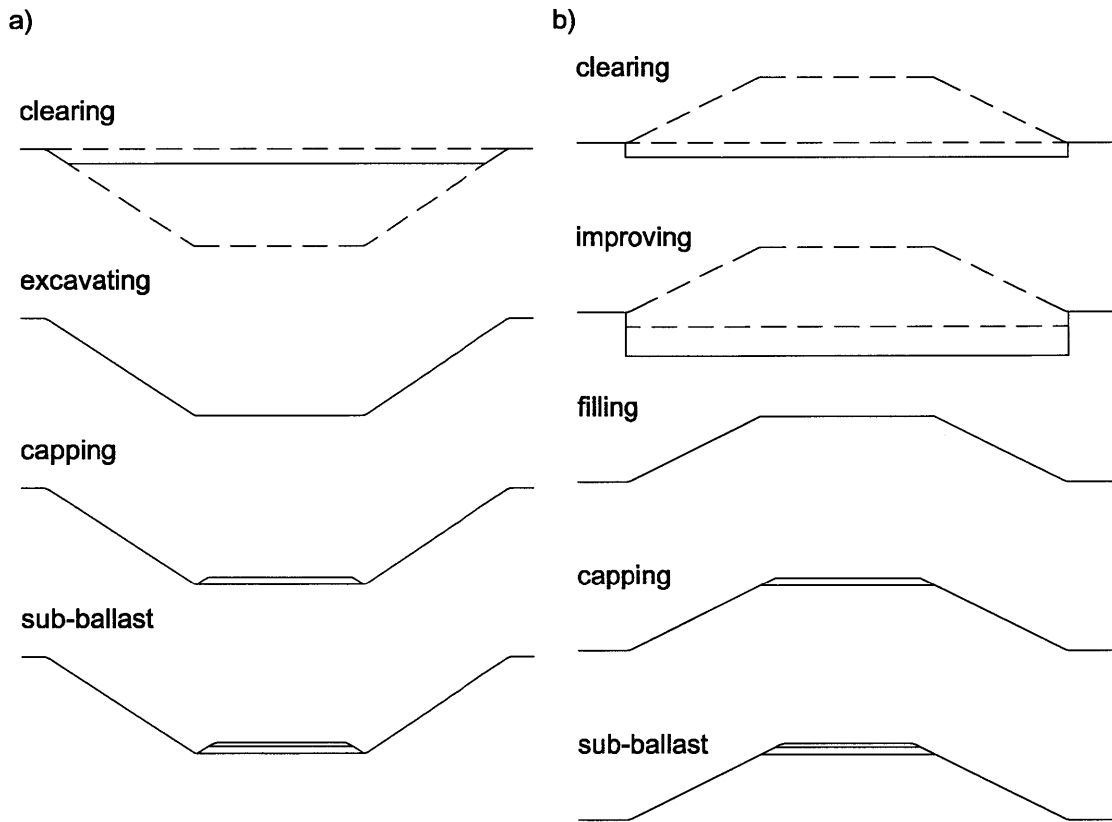


Figure C-32: Construction of cuts and embankments. a) cuts: first the top layer is cleared, then the cut is excavated, finally the capping is placed followed by the sub-ballast. b) embankments: the top layer is cleared, if needed the ground under the embankment is improved, then the embankment is filled, finally the capping is placed followed by the sub-ballast. The cost and time of these processes depend on the volumes to be removed (clearing and excavating), to be replaced (improving), and to be added (filling, capping, and sub-ballast).

The cut depth, D , the embankment height, H , the slope steepness, α , and the percentage of cut volume excavated with blasting, e , are averages or approximations of the varying cut depth, embankment height, slope steepness and percentage of cut volume excavated with blasting. They are discussed in the following.

The cut depth, D , is calculated with a weighted average

$$D = \frac{1}{4} (D_l + 2D_c + D_r), \quad (\text{C.1})$$

where D_l , D_r , and D_c are the cut depth on the left, on the right, and in the center, respectively

(Figure C-33a). Similarly, the embankment height, H , is calculated with a weighted average

$$H = \frac{1}{4} (H_l + 2H_c + H_r), \quad (\text{C.2})$$

where H_l , H_r , and H_c are the embankment height on the left, on the right, and in the center, respectively (Figure C-33b).

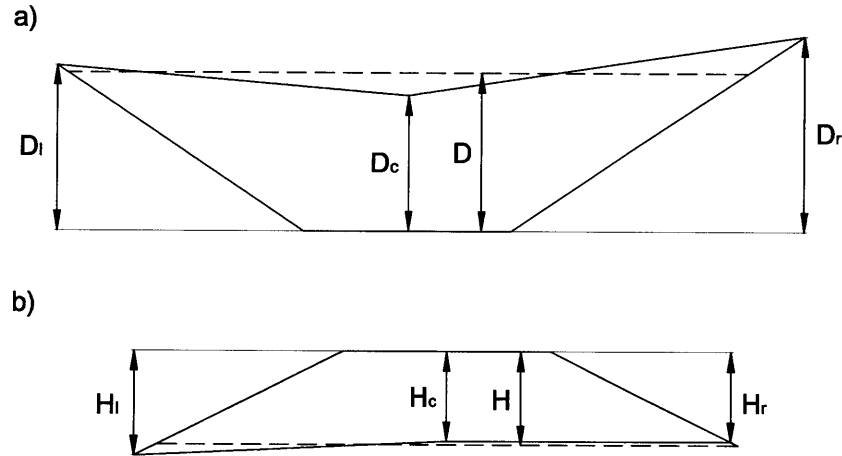


Figure C-33: a) the cut depth, D , and b) the embankment height, H , are weighted averages of the depths, respectively the heights, at the sides and in the center (equations C.1 and C.2).

The three cut depths (D_l, D_c, D_r) and the three embankment heights (H_l, H_c, H_r) of the cuts and embankments in alignments A-a and A-c are available from the source (RAVE 2006b). However, only the central depth, D_c , and the central height, H_c , of the cuts and embankments in alignments B-a and B-b are available from the source (RAVE 2008b). In order to estimate the cut depths, D , and the embankment heights, H , of the cuts and embankments in alignments B-a and B-b, the ratios $\frac{D}{D_c}$ and $\frac{H}{H_c}$ of the cuts and embankments in alignments A-a and A-c are calculated: $\frac{D}{D_c} = 1.12$ and $\frac{H}{H_c} = 1.01$. The depths of the cuts in alignments B-a and B-b are estimated with $D = 1.12 \times D_c$, while the heights of the embankments are estimated with $H = H_c$. The depths and the heights of the cuts and embankments in the four alignments are listed in Tables C.17 to C.24.

The slope steepness, α , is an approximation of the cut and embankment slopes, which in most cases include benches that increase the stability of the structure. The cut slopes have six geometries

(Figure C-34):

- CS1 has a slope of 1 : 1.5, corresponding to $\alpha = 1.5$;
- CS2 includes 3m wide benches every 8m of cut depth and a slope of 1 : 1.5 between benches. The cut slope including the benches is 8 : 15, corresponding to 1.875, which is rounded to $\alpha = 1.9$;
- CS3 includes four 3m wide benches every 8m of cut depth and a slope of 1 : 1.5 between benches, with the exception of the slope between the first and the second bench, which is 1 : 1. The cut slope including the benches is 4 : 7, corresponding to 1.75, which is rounded to $\alpha = 1.8$;
- CS4 includes 3m wide benches every 8m of cut depth and a slope of 1 : 1 between benches. The cut slope including the benches is 8 : 11, corresponding to 1.375, which is rounded to $\alpha = 1.4$;
- CS5 has a slope of 1.5 : 1, corresponding to $\alpha = 0.667$, which is rounded to $\alpha = 0.7$;
- CS6 corresponds to a vertical wall so that $\alpha = 0$.

The embankment slopes have two geometries (Figure C-35):

- ES1 has a slope of 1 : 2, corresponding to $\alpha = 2.0$;
- ES2 includes 2m wide benches every 10m of embankment height and a slope of 1 : 2 between benches. The embankment slope including the benches is 5 : 11, corresponding to $\alpha = 2.2$.

The slope steepness of the cuts and embankments in the four alignments are listed in Tables C.17 to C.24

The percentage of cut volume excavated with blasting, e , in alignments A-a and A-c is expressed with a range (RAVE 2006b), while in alignments B-a and B-b is equal to a single value (RAVE 2008b). For consistency, here the percentages of cut volume excavated with blasting, e , are set equal to the values in RAVE (2008b) for all alignments (Table C.14).

The cut and embankment volumes are calculated based on their geometric parameters and the volume expressions in Tables C.12 and C.13. The geometric parameters and the calculated volumes of all cuts and embankments in the four alignments are listed in Tables C.17 to C.24.

Table C.14: For consistency, the percentage of cut volume excavated with blasting, e [%] from RAVE (2008b) are used for all cuts and embankments.

Percentage of cut volume excavated with blasting, e [%]		
RAVE (2006b)	RAVE (2008b)	used here
0 – 5	5	5
5 – 25	25	25
25 – 50	50	50
50 – 75	75	75

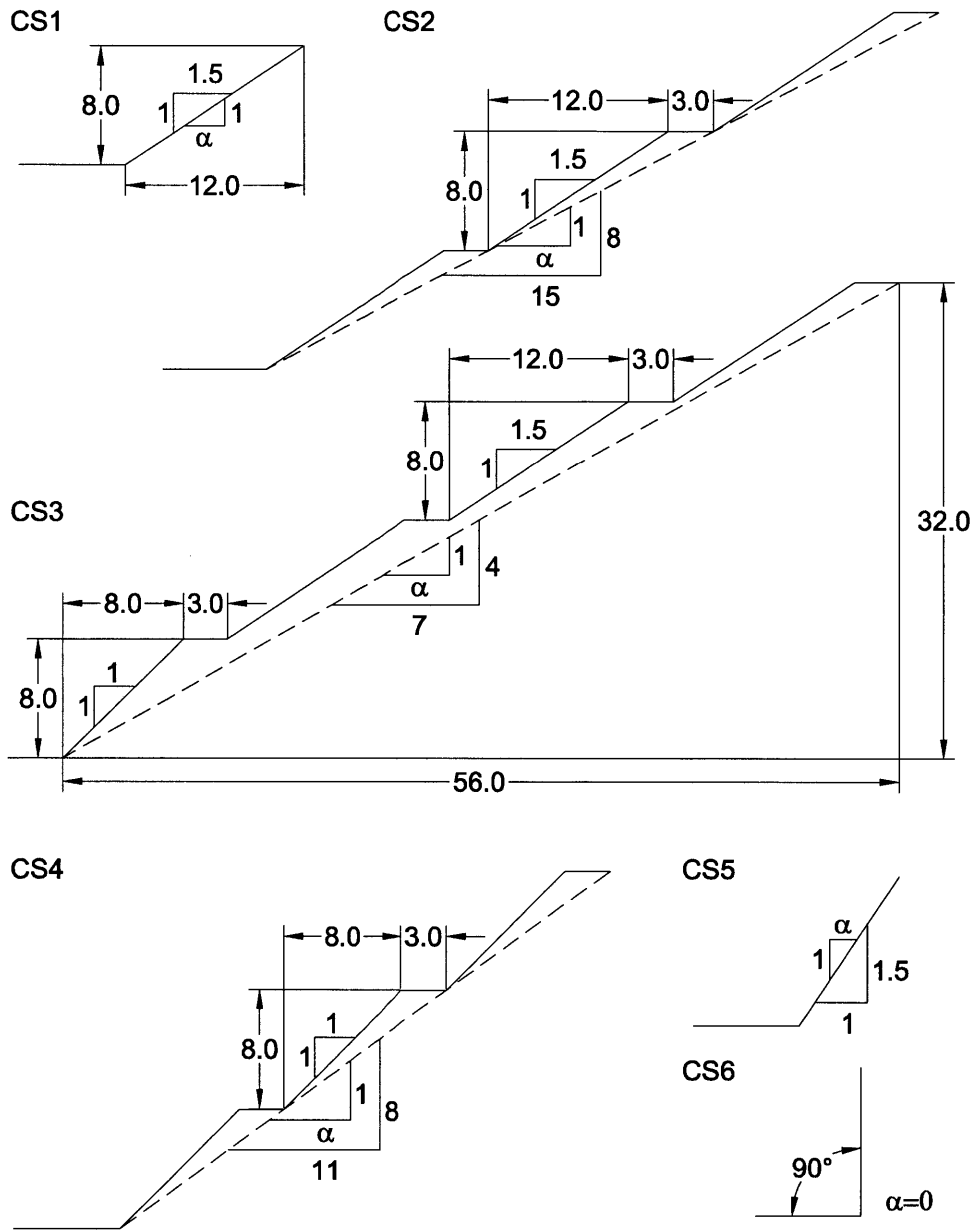


Figure C-34: Six geometries of cuts. They correspond to a slope steepness α varying between 0 for the vertical wall (CS6) and $\frac{15}{8} \approx 1.9$ (CS2).

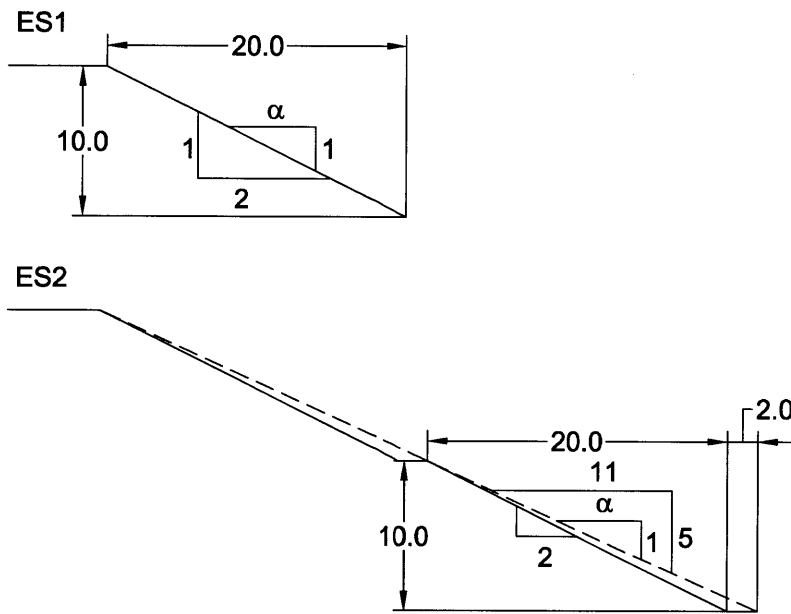


Figure C-35: Two geometries of embankments. They correspond to a slope steepness α varying between 2.0 (ES1) and $\frac{11}{5} = 2.2$ (ES2).

The sum of the clearing, excavating, capping, sub-ballast, improving, and filling volumes in the four alignments are compared with the volumes in the sources (RAVE 2006b, RAVE 2008b), showing that some good agreement is accompanied by some differences (Table C.15). These discrepancies are quantified with the ratio between the calculated volumes and the volumes in the sources.

Table C.15: The volumes calculated with the expressions in Tables C.12 and C.13 are compared with the volumes from the sources: for alignments A-a and A-c (RAVE 2006b) and for alignments B-a and B-b (RAVE 2008b). The ratios between calculated volumes and volumes from the sources show a good agreement in the case of clearing, excavating, and filling. Conversely, improving, capping and sub-ballast volumes require more scrutiny.

Volume	align.	calculated [m ³]	(RAVE 2006b) [m ³]	(RAVE 2008b)	ratio [%]
clearing	A-a	872,144	754,430		116
	A-c	1,009,580	1,278,034		79
	B-a	929,388		992,865	94
	B-b	950,790		1,041,737	91
excavating	A-a	9,763,744	8,561,477		114
	A-c	8,720,079	8,652,454		101
	B-a	9,682,903		11,068,344	87
	B-b	11,651,619		11,989,744	97
capping	A-a	214,150	279,020		77
	A-c	274,818	483,594		57
	B-a	353,220		735,075	48
	B-b	363,804		766,289	47
sub-ballast	A-a	107,075	124,575		86
	A-c	137,409	222,601		62
	B-a	176,610		326,822	54
	B-b	181,902		340,755	53
improving	A-a	363,951	110,648		329
	A-c	776,009	211,519		367
	B-a	293,114		557,086	53
	B-b	309,130		589,895	52
filling	A-a	2,753,406	2,779,957		99
	A-c	4,370,579	4,815,066		93
	B-a	5,899,139		5,983,131	99
	B-b	5,578,349		6,934,676	80

Although the ratios of the clearing, the excavating, and the filling volumes are balanced on both sides of the ideal ratio of 100%, which can be considered to be satisfactory, the same does not apply to the improving, the capping and the sub-ballast volumes. The improving volumes are characterized by highly varying ratios between minima around 50% in alignments B-a and B-b and

maxima of over 300% in alignments A-a and A-c. The calculated improving volumes are obtained with the expression in Table C.13b, recalled here

$$S \times \left(\frac{w_e + w_max}{2} \right) \times t. \quad (C.3)$$

In this expression, the geometric parameter S (length of improving layer) and t (thickness of improving layer) are available from the sources (RAVE 2006b, RAVE 2008b) for all embankments in the four alignments (Tables C.21 to C.24). Conversely, the width of the improving layers are not available from the sources. In expression C.3, the width is assumed to be equal to the average of the embankment width at the top, w_e , and the embankment maximum width, w_max . RAVE (2006b) might assume that the width of the improving layer is equal to the embankment width at top, w_e ; in fact, improving volumes calculated with this width yield a ratio of 98% in alignment A-a and of 117% in alignment A-c. Differently, RAVE (2008b) might assume that the width of the improving layer is equal to the embankment maximal width, w_max ; in fact, improving volumes calculated with this width yield a ratio of 90% in alignment B-a and of 88% in alignment B-b. Following these analyses, an average width between the embankment width at top, w_e , and the embankment maximal width, w_max , is an appropriate choice for the estimation of the width of the improving layers. The improving volumes of the embankments in the four alignments are listed in Tables C.21 to C.24.

The differences between calculated volumes of capping and sub-ballast and the volumes from the sources (RAVE 2006b, RAVE 2008b) are more difficult to explain. The geometric parameters and the volumes in the sources are contradictory. The capping volume expression (Tables C.12c and C.13d), recalled here

$$L \times w_capping \times t_capping, \quad (C.4)$$

is a function of the cut or embankment length, L , the width and the thickness of the capping layer, $w_capping$ and $t_capping$. These are equal to $w_capping = 14\text{m}$ and $t_capping = 0.6\text{m}$ (Table C.11). In order to check the capping volumes in the sources, the total length of cuts and embankments is backcalculated:

$$L_{back} = \frac{V_capping}{w_capping \times t_capping} = \frac{V_capping}{14\text{m} \times 0.6\text{m}}, \quad (C.5)$$

where $V_capping$ is the total capping volume in the alignment. The backcalculated length is

greater than the alignment length in three of the four alignments (Table C.16).

The sub-ballast volume expression (Tables C.12d and C.13e), recalled here

$$L \times w_{sub-ballast} \times t_{sub-ballast}, \quad (C.6)$$

is a function of the cut or embankment length, L , the width and the thickness of the sub-ballast layer, $w_{sub-ballast}$ and $t_{sub-ballast}$. These are equal to $w_{sub-ballast} = 14\text{m}$ and $t_{sub-ballast} = 0.3\text{m}$ (Table C.11). In order to check the sub-ballast volumes in the sources, the total length of cuts and embankments is backcalculated

$$L_{back} = \frac{V_{sub-ballast}}{w_{sub-ballast} \times t_{sub-ballast}} = \frac{V_{sub-ballast}}{14\text{m} \times 0.3\text{m}}, \quad (C.7)$$

where $V_{sub-ballast}$ is the total sub-ballast volume in the alignment. Again the backcalculated length is greater than the alignment length in three of the four alignments (Table C.16). Following these analyses, clearly the geometrical parameters and the capping/sub-ballast volumes in the sources are in contradiction. Due to this contradiction, the calculated volumes of capping and sub-ballast in Table C.15 are preferred.

The calculated volumes of capping and the sub-ballast of the cuts and embankments in the four alignments, are listed in Tables C.17 to C.24.

Table C.16: Given, the volumes, the widths and the thicknesses of the capping and sub-ballast layers (RAVE 2006b, RAVE 2008b), the total length of cuts and embankments is back calculated. In three alignments, it is larger than the alignment length. Clearly, the data in the sources are contradictory.

	Alignment	volume [m^3]	backcalculated length [m]		alignment length [m]
capping	A-a	279,020	33,217	<	45,600
	A-c	483,594	57,571	>	44,649
	B-a	735,075	87,509	>	66,500
	B-b	766,289	91,225	>	69,790
sub-ballast	A-a	124,575	29,661	<	45,600
	A-c	222,601	53,000	>	44,649
	B-a	326,822	77,815	>	66,500
	B-b	340,755	81,132	>	69,790

The geometric parameters, the clearing volume, the excavating volume, the capping volume,

and the sub-ballast volume of the cuts in the four alignments are listed in Tables C.17 to C.20. The geometric parameters, the clearing volume, the improving volume, the filling volume, the capping volume, and the sub-ballast volume of the embankments in the four alignments are listed in Tables C.21 to C.24.

Table C.17: Alignment A-a: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.

Cut	L	D	α	w_{max}	e	t_{clear}	volume [$10^3 m^3$]			
	[m]	[m]	[-]	[m]	[%]	[m]	V1	V2	V3	V4
C1	970	20.5	1.9	97.9	50	0.6	34.3	422.7	8.1	4.0
C2	700	8.0	1.5	44	5	0.6	13.4	65.0	5.9	2.9
C3	3,265	25.3	1.9	116.1	25	0.6	133.4	2,016.2	27.4	13.7
C4	245	21.3	1.9	100.9	50	0.6	8.9	113.7	2.1	1.0
C5	120	6.5	1.5	39.5	5	0.6	2.1	8.2	1.0	0.5
C6	280	4.5	1.5	33.5	5	0.6	4.5	10.9	2.4	1.2
C7	608	32.0	1.8	135.2	75	0.6	28.3	539.8	5.1	2.6
C8	110	30.0	1.9	134.0	50	0.6	5.1	90.6	.0	.5
C9	160	6.0	1.5	38.0	5	0.6	2.8	9.7	1.3	0.6
C10	470	21.5	1.9	101.7	50	0.6	17.2	221.5	3.9	2.0
C11	1,490	31.0	1.9	137.8	50	0.6	70.5	1,298.2	12.5	6.3
C12	910	28.3	1.9	127.5	25	0.6	40.3	678.8	7.6	3.8
C13	590	33.3	1.8	139.9	50	0.6	28.3	560.7	5.0	2.5
C14	950	28.5	1.8	122.6	75	0.6	40.6	693.1	8.0	4.0
C15	910	14.0	1.9	73.2	25	0.6	25.4	214.9	7.6	3.8
C16	600	33.8	1.9	148.4	50	0.6	30.3	606.6	5.0	2.5
C17	280	32.5	1.9	143.5	25	0.6	13.7	264.6	2.4	1.2
C18	360	10.8	1.5	52.4	5	0.6	7.8	52.1	3.0	1.5
C19	640	39.8	1.8	163.3	50	0.6	35.2	827.8	5.4	2.7
C20	130	26.8	1.9	121.8	25	0.6	5.5	88.5	1.1	0.5
C21	400	14.0	1.9	73.2	25	0.6	11.2	94.5	3.4	1.7
C22	793	35.8	1.9	156	50	0.6	41.9	885.7	6.7	3.3

Table C.18: Alignment A-c: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.

Cut	L	D	α	w_{max}	e	t_{clear}	volume [$10^3 m^3$]			
	[m]	[m]	[$-$]	[m]	[$\%$]	[m]	V1	V2	V3	V4
C1	1,340	16.0	1.9	80.8	25	0.6	40.5	391.1	11.3	5.6
C2	1,370	17.8	1.9	87.6	25	0.6	44.2	474.6	11.5	5.8
C3	440	10.0	1.5	50.0	5	0.6	9.2	56.8	3.7	1.8
C4	520	15.3	1.9	78.1	25	0.6	15.3	141.3	4.4	2.2
C5	1,750	19.0	1.0	92.2	25	0.6	58.9	673.7	14.7	7.4
C6	825	30.5	1.9	135.9	75	0.6	38.6	699.1	6.9	3.5
C7	664	34.0	1.9	149.2	50	0.6	33.7	678.4	5.6	2.8
C8	391	35.0	1.8	146.0	50	0.6	19.5	404.8	3.3	1.6
C9	880	22.5	1.9	105.5	5	0.6	33.2	447.0	7.4	3.7
C10	520	16.0	1.9	80.8	5	0.6	15.7	151.8	4.4	2.2
C11	1,405	35.8	1.8	148.9	75	0.6	71.1	1,512.3	11.8	5.9
C12	525	13.0	1.9	69.4	5	0.6	14.1	110.4	4.4	2.2
C13	575	14.0	1.9	73.2	25	0.6	16.1	135.9	4.8	2.4
C14	600	33.8	1.9	148.4	50	0.6	30.3	606.6	5.0	2.5
C15	280	32.5	1.9	143.5	25	0.6	13.7	264.6	2.4	1.2
C16	360	10.8	1.5	52.4	5	0.6	7.8	52.1	3.0	1.5
C17	640	39.8	1.8	163.3	50	0.6	35.2	827.8	5.4	2.7
C18	130	26.8	1.9	121.8	25	0.6	5.5	88.5	1.1	0.5
C19	400	14.0	1.9	73.2	25	0.6	11.2	94.5	3.4	1.7
C20	793	35.8	1.9	156	50	0.6	41.9	885.7	6.7	3.3
C21	720	4.0	1.5	32.0	5	0.6	11.2	23.3	6.0	3.0

Table C.19: Alignment B-a: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.

Cut	L	D	α	w_{max}	e	t_{clear}	volume [$10^3 m^3$]			
	[m]	[m]	[-]	[m]	[%]	[m]	V1	V2	V3	V4
C1	1,175	26.9	0.7	57.7	25	0.5	22.8	491.7	9.9	4.9
C2	248	28	0.7	59.2	25	0.5	4.9	109.9	2.1	1.0
C3	1,359	22.4	1.9	105.1	0	0.5	42.5	693.7	11.4	5.7
C4	425	11.2	1.9	62.6	0	0.5	8.8	72.6	3.6	1.8
C5	1,612	22.4	0.7	51.4	25	0.5	28.8	521.0	13.5	6.8
C6	300	11.2	1.9	62.6	0	0.5	6.2	51.2	2.5	1.3
C7	279	16.8	1.4	67.0	50	0.5	6.1	77.6	2.3	1.2
C8	250	16.8	1.9	83.8	25	0.5	6.5	80.2	2.1	1.0
C9	440	22.4	1.9	105.1	0	0.5	13.8	224.6	3.7	1.9
C10	150	13.4	1.9	70.9	0	0.4	2.7	34.4	1.3	0.6
C11	1,334	32.5	1.9	143.5	0	0.4	43.6	1,282.3	11.2	5.6
C12	555	20.2	1.9	96.8	0	0.4	13.0	242.6	4.7	2.3
C13	600	29.1	1.4	101.5	25	0.4	14.6	397.1	5.0	2.5
C14	300	13.4	1.9	70.9	25	0.4	5.5	68.9	2.5	1.3
C15	1,000	33.6	0.7	67	50	0.4	17.4	582.0	8.4	4.2
C16	450	22.4	1.9	105.1	25	0.4	11.3	232.5	3.8	1.9
C17	403	20.2	1.9	96.8	25	0.4	9.4	176.1	3.4	1.7
C18	1,292	25.8	1.9	118.0	25	0.4	35.7	842.3	10.9	5.4
C19	250	16.8	1.9	83.8	25	0.4	5.2	81.5	2.1	1.1
C20	2,417	24.6	1.9	113.5	25	0.4	64.5	1,456.5	20.3	10.2
C21	144	15.7	1.9	79.7	25	0.4	2.9	42.2	1.2	0.6
C22	1,250	20.2	1.9	96.8	25	0.4	29.2	546.3	10.5	5.3
C23	2,600	16.8	1.9	83.8	25	0.4	54.0	847.6	21.8	10.9
C24	850	11.2	1.9	62.6	0	0.5	17.6	145.2	7.1	3.6
C25	500	11.2	1.9	62.6	0	0.5	10.3	85.4	4.2	2.1
C26	200	11.2	1.9	62.6	0	0.4	3.3	35.0	1.7	0.8
C27	50	11.2	1.9	62.6	0	0.4	0.8	8.8	0.4	0.2
C28	50	11.2	1.9	62.6	0	0.4	0.8	8.8	0.4	0.2
C29	150	11.2	1.9	62.6	0	0.4	2.5	26.2	1.3	0.6
C30	700	11.2	1.9	62.6	0	0.4	11.6	122.4	5.9	2.9
C31	350	11.2	1.9	62.6	0	0.4	5.8	61.2	2.9	1.5
C32	200	11.2	1.9	62.6	0	0.4	3.3	35.0	1.7	0.8

Table C.20: Alignment B-b: cut volumes. The clearing volume (V1), the excavating volume (V2), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.12) and the geometric parameters of the cut. With the costs per unit volume and the production rates, they determine the construction cost and time of the cuts in the alignment.

Cut	L	D	α	w_{max}	e	t_{clear}	volume [$10^3 m^3$]			
	[m]	[m]	[-]	[m]	[%]	[m]	V1	V2	V3	V4
C1	1,175	26.9	0.7	57.7	25	0.5	22.8	491.7	9.9	4.9
C2	248	28	0.7	59.2	25	0.5	4.9	109.9	2.1	1.0
C3	1,359	22.4	1.9	105.1	0	0.5	42.5	693.7	11.4	5.7
C4	1,425	24.6	1.9	113.5	0	0.5	47.5	849.2	12.0	6.0
C5	410	22.4	0.7	51.4	0	0.5	7.3	132.5	3.4	1.7
C6	139	22.4	0.7	51.4	25	0.5	2.5	44.9	1.2	0.6
C7	450	22.4	0	20.0	0	0.5	4.5	96.3	3.8	1.9
C8	513	24.6	0	20	0	0.5	5.1	121.1	4.3	2.2
C9	254	22.4	0	20	25	0.5	2.5	54.4	2.1	1.1
C10	1,061	22.4	1.9	105.1	25	0.5	33.2	541.6	8.9	4.5
C11	587	26.9	1.9	122.2	0	0.5	20.9	406.0	4.9	2.5
C12	500	26.9	1.9	122.2	0	0.5	17.8	345.9	4.2	2.1
C13	400	20.2	1.9	96.8	25	0.5	11.7	172.5	3.4	1.7
C14	1,875	53.8	0.7	95.3	50	0.3	32.5	2,242.6	15.8	7.9
C15	1,650	22.4	1.4	82.7	25	0.3	25.4	730.5	13.9	6.9
C16	520	38.1	0.7	73.3	75	0.3	7.3	367.0	4.4	2.2
C17	237	26.9	1.9	122.2	75	0.3	5.1	167.3	2.0	1.0
C18	700	26.9	0.7	57.7	25	0.3	8.1	298.3	5.9	2.9
C19	250	17.9	1.9	88.0	25	0.4	5.4	90.1	2.1	1.1
C20	1,722	20.2	1.9	96.8	25	0.4	40.2	752.7	14.5	7.2
C21	180	39.2	1.9	169.0	25	0.4	6.8	238.9	1.5	0.8
C22	1,525	24.6	1.9	113.5	25	0.4	40.7	918.9	12.8	6.4
C23	1,300	20.2	1.9	96.8	25	0.4	30.4	568.2	10.9	5.4
C24	2,625	16.8	1.9	83.8	0	0.4	54.5	855.7	22.1	11.0
C25	400	11.2	1.9	62.6	0	0.5	8.3	68.3	3.4	1.7
C26	310	11.2	1.9	62.6	0	0.5	6.4	52.9	2.6	1.3
C27	800	11.2	1.9	62.6	0	0.5	16.5	136.6	6.7	3.4
C28	150	11.2	1.9	62.6	0	0.6	3.7	25.0	1.3	0.6
C29	150	11.2	1.9	62.6	0	0.4	4.1	43.7	2.1	1.1
C30	200	11.2	1.9	62.6	0	0.4	3.3	35.0	1.7	0.8

Table C.21: Alignment A-a: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment.

Emb.	L	H	α	w_{max}	S	t	t_{clear}	volume [$10^3 m^3$]				
	[m]	[m]	[-]	[m]	[m]	[m]	[m]	V1	V5	V6	V3	V4
E1	180	6.0	2.0	38.0	0	0	0.6	2.8	0	14.7	1.5	0.8
E2	335	12.0	2.2	66.8	0	0	0.6	8.1	0	71.6	2.8	1.4
E3	550	19.0	2.2	97.6	130	1.2	0.6	18.4	8.7	237.2	4.6	2.3
E4	135	6.3	2.0	39.2	0	0	0.6	2.2	0	11.7	1.1	0.6
E5	60	7.3	2.0	43.2	0	0	0.6	1.0	0	6.2	0.5	0.3
E6	490	12.0	2.2	66.8	190	2.0	0.6	11.9	15.4	104.8	4.1	2.1
E7	210	10.3	2.0	55.2	200	1.9	0.6	4.4	13.1	34.4	1.8	0.9
E8	73	9.8	2.0	53.2	0	0	0.6	4.4	13.1	34.4	1.8	0.9
E9	55	10.0	2.0	54.0	0	0	0.6	1.1	0	8.6	0.5	0.2
E10	200	13.8	2.2	74.7	60	0.8	0.6	1.1	0	8.6	0.5	0.2
E11	310	5.0	2.0	34.0	280	1.8	0.6	4.5	12.1	20.4	2.6	1.3
E12	400	15.5	2.2	82.2	90	1.6	0.6	11.5	6.9	125.4	3.4	1.7
E13	170	9.8	2.0	53.2	0	0	0.6	3.4	0	26.0	1.4	0.7
E14	60	9.0	2.0	50.0	0	0	0.6	1.2	0	8.2	0.5	0.3
E15	850	13.3	2.2	72.5	100	1.2	0.6	22.1	5.2	211.5	7.1	3.6
E16	850	22.5	2.2	113.0	420	3.5	0.6	32.4	93.3	481.8	7.1	3.6
E17	2,240	14.0	2.2	75.6	921	2.0	0.6	60.2	82.5	601.7	18.8	9.4
E18	44	9.5	2.0	52.0	0	0	0.6	0.9	0	6.4	0.4	0.2
E19	71	11.0	2.0	58.0	0	0	0.6	1.5	0	12.8	0.6	0.3
E20	1,460	15.0	2.2	80.0	1,230	2.0	0.6	41.2	115.6	435.4	12.3	6.1
E21	670	9.5	2.0	52.0	180	1.5	0.6	13.3	8.9	98.1	5.6	2.8
E22	1,100	10.0	2.0	54.0	0	0	0.6	22.4	0	172.8	9.2	4.6

Table C.22: Alignment A-c: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment.

Emb.	L	H	α	w_{max}	S	t	t_{clear}	volume [$10^3 m^3$]				
	[m]	[m]	[-]	[m]	[m]	[m]	[m]	V1	V5	V6	V3	V4
E1	210	5.0	2.0	34.0	170	3.7	0.6	3.0	15.1	13.9	1.8	0.9
E2	510	13.8	2.2	74.7	160	1.4	0.6	13.6	9.4	134.1	4.3	2.1
E3	360	7.0	2.0	42.0	0	0	0.6	6.0	0	35.4	3.0	1.5
E4	850	8.0	2.0	46.0	200	3.3	0.6	15.3	19.8	99.2	7.1	3.6
E5	130	4.8	2.0	33.2	105	2.5	0.6	1.8	6.2	8.2	1.1	0.5
E6	1,175	6.0	2.0	38.0	260	3.7	0.6	18.3	25.0	95.9	9.9	4.9
E7	50	9.5	2.0	52.0	0	0	0.6	1.0	0	7.3	0.4	0.2
E8	2,280	16.0	2.2	84.4	1,950	2.0	0.6	67.3	191.9	750.7	19.2	9.6
E9	1,350	12.0	2.0	62.0	300	2.0	0.6	30.8	22.8	273.8	11.3	5.7
E10	1,350	12.0	2.0	62.0	300	2.0	0.6	30.8	22.8	273.8	11.3	5.7
E11	750	10.3	2.0	55.2	0	0	0.6	15.6	0	122.7	6.3	3.2
E12	2,320	16.0	2.2	84.4	1,124	3.0	0.6	68.5	165.9	763.9	19.5	9.7
E13	650	22.0	2.2	110.8	0	0	0.6	24.3	0	355.1	5.4	2.7
E14	2,240	14.0	2.2	75.6	921	2.0	0.6	60.2	82.5	601.7	18.8	9.4
E15	44	9.5	2.0	52.0	0	0	0.6	0.9	0	6.4	0.4	0.2
E16	71	11.0	2.0	58.0	0	0	0.6	1.5	0	12.8	0.6	0.3
E17	1,460	15.0	2.2	80.0	1,230	2.0	0.6	41.2	115.6	435.4	12.3	6.1
E18	670	9.5	2.0	52.0	180	1.5	0.6	13.3	8.9	98.1	5.6	2.8
E19	1,100	10.0	2.0	54.0	0	0	0.6	22.4	0	172.8	9.2	4.6
E20	18	4.0	2.0	30.0	0	0	0.6	0.2	0	0.9	0.1	0.1

Table C.23: Alignment B-a: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment.

Emb.	L	H	α	w_{max}	S	t	t_{clear}	volume [$10^3 m^3$]				
	[m]	[m]	[-]	[m]	[m]	[m]	[m]	V1	V5	V6	V3	V4
E1	1,025	13.0	2.2	71.2	0.5	250	1.0	21.8	10.7	242.1	8.6	4.3
E2	2,440	16.0	2.2	84.4	420	3.0	0.5	60.2	62.0	791.4	20.5	10.2
E3	1,460	10.0	2.0	54.0	120	2.0	0.5	24.8	8.2	224.4	12.3	6.1
E4	1,925	16.0	2.2	84.4	0	0	0.5	47.4	0	624.3	16.2	8.1
E5	750	19.0	2.2	97.6	300	1.5	0.5	20.9	25.1	319.2	6.3	3.2
E6	1,225	20.0	2.2	102.0	250	2.0	0.5	35.5	29.0	566.4	10.3	5.1
E7	500	11.0	2.0	58.0	250	3.0	0.5	9.0	27.0	87.8	4.2	2.1
E8	905	11.0	2.2	102.0	100	1.0	0.5	26.2	5.8	418.4	7.6	3.8
E9	573.5	13.0	2.2	71.2	0	0	0.4	9.8	0	133.0	4.8	2.4
E10	256.5	10.0	2.0	54.0	0	0	0.4	3.5	0	38.5	2.2	1.1
E11	250	12.0	2.0	62.0	0	0	0.4	3.8	0	48.8	2.1	1.1
E12	400	10.0	2.0	54.0	0	0	0.4	5.4	0	60.1	3.4	1.7
E13	300	14.0	2.2	75.6	150	0.6	0.4	5.4	4.0	77.9	2.5	1.3
E14	1,550	15.0	2.2	80.0	600	0.6	0.4	29.1	16.9	447.6	13.0	6.5
E15	899	17.0	2.2	88.8	700	1.5	0.4	18.5	54.0	316.0	7.6	3.8
E16	840	15.0	2.2	80.0	0	0	0.4	15.8	0	242.6	7.1	3.5
E17	1,004	18.0	2.2	93.2	350	2.0	0.4	21.5	37.5	386.6	8.4	4.2
E18	1,126	13.0	2.2	71.2	0	0	0.4	19.2	0	261.2	9.5	4.7
E19	1,100	16.0	2.2	84.4	50	1.0	0.4	21.4	2.4	351.4	9.2	4.6
E20	600	11.0	2.0	58.0	150	1.0	0.4	8.6	5.4	103.2	5.0	2.5
E21	30	10.0	2.0	54.0	0	0	0.5	0.5	0	4.6	0.3	0.1
E22	350	10.0	2.0	54.0	0	0	0.5	6.0	0	53.8	2.9	1.5
E23	250	10.0	2.0	54.0	100	1.5	0.5	4.3	5.1	38.4	2.1	1.1
E24	100	10.0	2.0	54.0	0	0	0.4	1.4	0	15.0	0.8	0.4
E25	98	10.0	2.0	54.0	0	0	0.4	1.3	0	14.7	0.8	0.4
E26	38	10.0	2.0	54.0	0	0	0.4	0.5	0	5.7	0.3	0.2
E27	22	10.0	2.0	54.0	0	0	0.4	0.3	0	3.3	0.2	0.1
E28	150	10.0	2.0	54.0	0	0	0.4	2.0	0	22.5	1.3	0.6

Table C.24: Alignment B-b: embankment volumes. The clearing volume (V1), the improving volume (V5), the filling volume (V6), the capping volume (V3), and the sub-ballast volume (V4) are calculated with volume expressions (Table C.13) and the geometric parameters of the embankment. With the costs per unit volume and the production rates, they determine the construction cost and time of the embankments in the alignment.

Emb.	L	H	α	w_{max}	S	t	t_{clear}	volume [$10^3 m^3$]				
	[m]	[m]	[-]	[m]	[m]	[m]	[m]	V1	V5	V6	V3	V4
E1	1,025	13.0	2.2	71.2	0.5	250	1.0	21.8	10.7	242.1	8.6	4.3
E2	2,440	16.0	2.2	84.4	420	3.0	0.5	60.2	62.0	791.4	20.5	10.2
E3	1,460	10.0	2.0	54.0	120	2.0	0.5	24.8	8.2	224.4	12.3	6.1
E4	25	16.0	2.2	84.4	0	0	0.5	0.60	0	8.1	0.2	0.1
E5	1,335	18.0	2.2	93.2	150	1.2	0.5	35.8	9.6	521.2	11.2	5.6
E6	227	10.0	2.0	54.0	0	0	0.5	3.9	9	34.9	1.9	1.0
E7	598.5	10.0	2.0	54.0	0	0	0.5	10.2	0	92.0	5.0	2.5
E8	390	10.0	2.0	54.0	200	2.5	0.5	6.6	17.0	59.9	3.3	1.6
E9	450	10.0	2.0	54.0	200	1.0	0.5	7.6	6.8	69.1	3.8	1.9
E10	2,600	15.0	2.2	80.0	1,350	2.0	0.6	73.3	126.9	775.3	21.8	10.9
E11	290	12.0	2.0	62.0	140	4.0	0.6	6.6	21.3	58.8	2.4	1.2
E12	212	11.0	2.0	58.0	0.5	0	0.5	3.8	0	37.2	1.8	0.9
E13	2,075	18.0	2.2	93.2	0	0	0.5	55.6	0	810.1	17.4	8.7
E14	848	25.0	2.2	124.0	0	0	0.3	17.6	0	554.6	7.1	3.6
E15	902.9	11.0	2.0	58.0	700	1.5	0.3	13.0	37.8	155.4	7.6	3.8
E16	300	12.0	2.0	62.0	0	0	0.4	4.6	0	58.6	2.5	.13
E17	456	18.0	2.2	93.2	0	0	0.4	9.8	0	175.6	3.8	1.9
E18	585	19.0	2.2	97.6	0	0	0.4	13.1	0	245.7	4.9	2.5
E19	1,250	11.0	2.0	58.0	50	1.0	0.4	18.0	1.8	215.1	10.5	5.2
E20	600	11.0	2.0	58.0	150	1.0	0.4	8.6	5.4	103.2	5.0	2.5
E21	265	15.0	2.2	80.0	0	0	0.4	5.0	0	76.5	2.2	1.1
E22	600	10.0	2.0	54.0	0	0	0.5	10.2	0	92.2	5.0	2.5
E23	32	10.0	2.0	54.0	0	0	0.5	0.5	0	4.8	0.3	0.1
E24	47	10.0	2.0	54.0	0	0	0.5	0.8	0	7.2	0.4	0.2
E25	73	10.0	2.0	54.0	0	0	0.5	1.2	0	11.2	0.6	0.3
E26	750	10.0	2.0	54.0	0	0	0.5	12.7	1.7	115.2	6.3	3.1
E27	20	10.0	2.0	54.0	0	0	0.3	0.2	0	2.9	0.2	0.1
E28	19	10.0	2.0	54.0	0	0	0.3	0.2	0	2.8	0.2	0.1
E29	62	10.0	2.0	54.0	0	0	0.3	0.6	0	9.1	0.5	0.3
E30	50	10.0	2.0	54.0	0	0	0.3	0.5	0	7.3	0.4	0.2
E31	7	10.0	2.0	54.0	0	0	0.3	0.1	0	1.0	0.1	0.0
E32	1	10.0	2.0	54.0	0	0	0.3	0.0	0	0.1	0.0	0.0
E33	100	10.0	2.0	54.0	0	0	0.4	1.4	0	15.0	0.8	0.4

C.4.2 Construction Cost and Time

The total cost of the cuts and the embankments in an alignment is equal to the sum of the clearing, excavating, capping, sub-ballast, improving and filling costs of all cuts and embankments (Table 7.40). The total time is equal to the sum of clearing, excavating, capping, sub-ballast, improving and filling times on the critical path of the construction network (Figure C-36). The construction networks of the four alignments are shown in Figures C-37 to C-40.

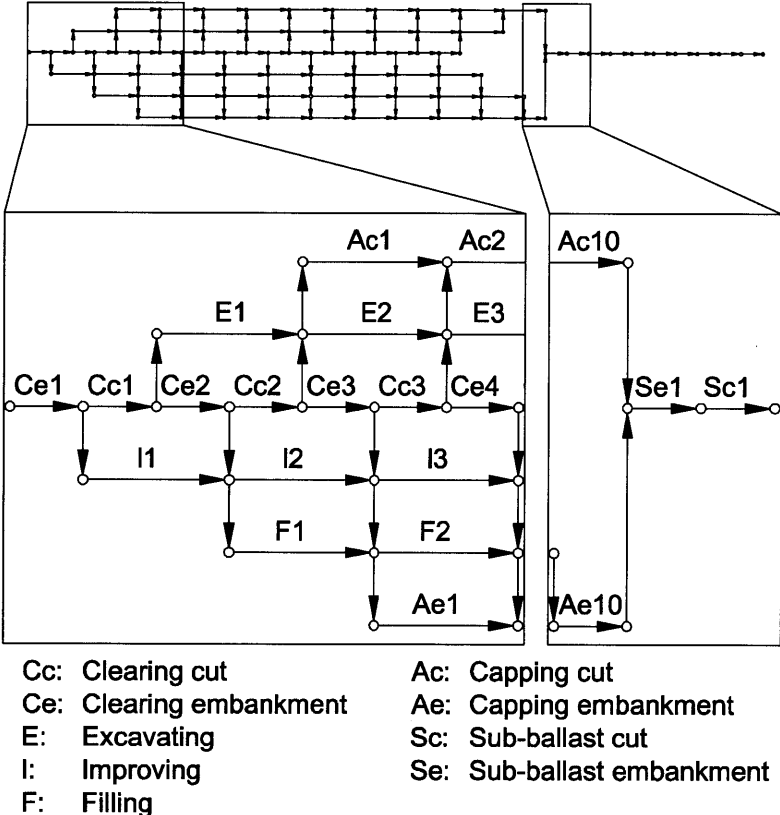


Figure C-36: Construction network of a sequence of cuts and embankments. The total construction time is equal to the sum of clearing, excavating, improving, filling, capping, and sub-ballast times on the critical path of the construction network.

The critical path of the construction network is determined by the sequence of the activities and their times and by the material availability. In fact, if the material used by the activity is not available, the activity must wait until it becomes available. The material handling in the cuts and embankments construction was presented in chapter 3. It is applied to the four alignments: the real

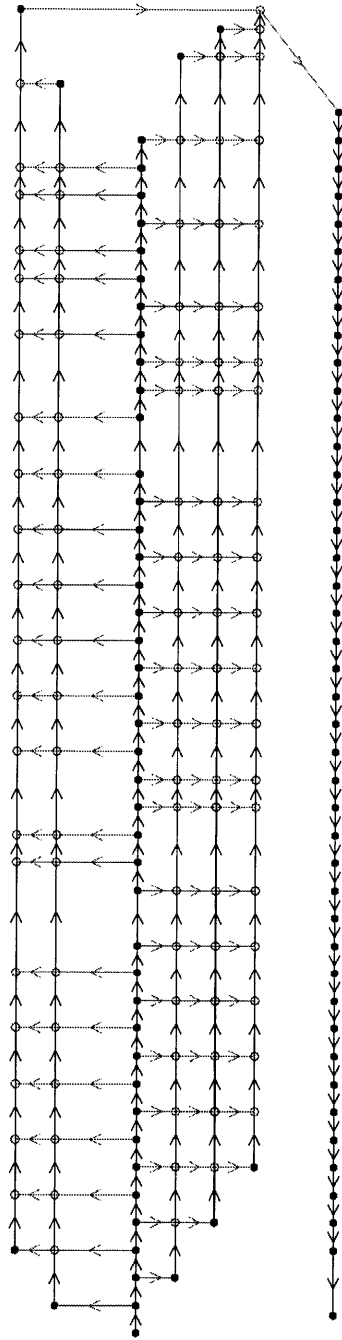


Figure C-38: Construction network of cuts and embankments in alignment A-c.

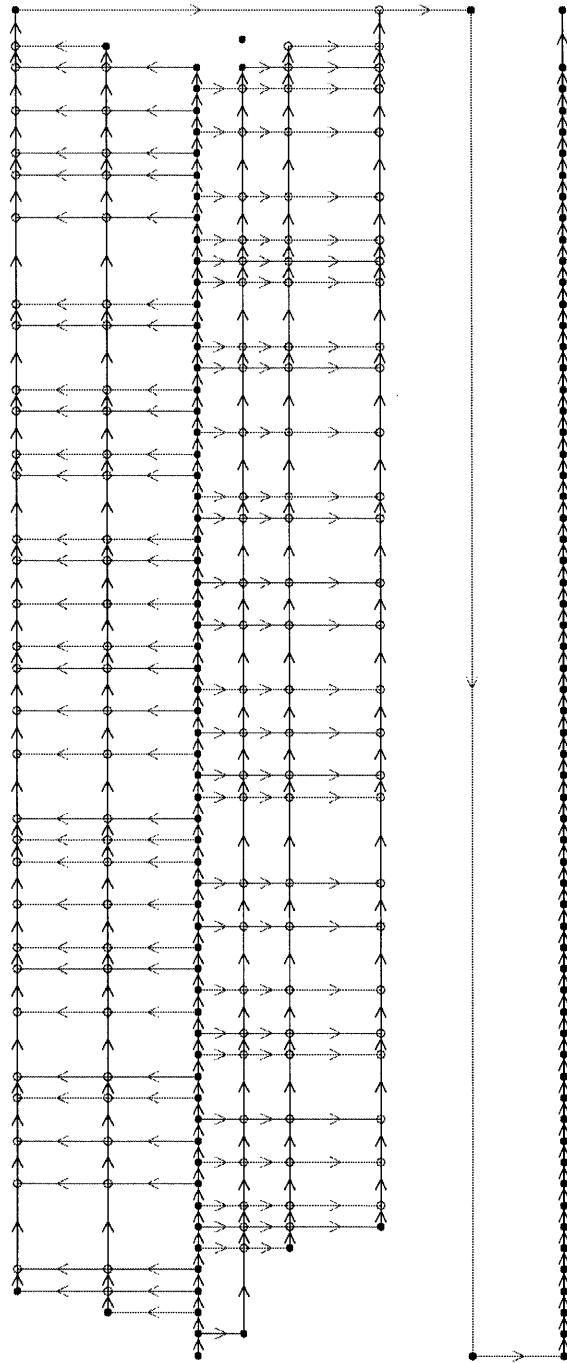


Figure C-39: Construction network of cuts and embankments in alignment B-a.

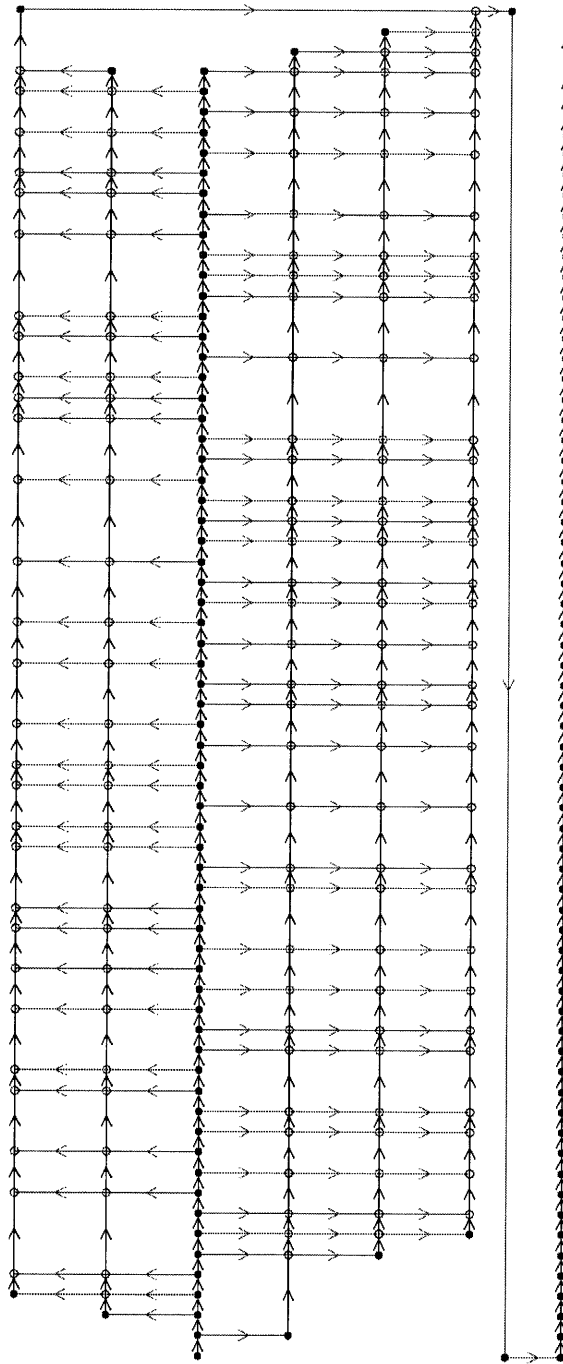


Figure C-40: Construction network of cuts and embankments in alignment B-b.

resource is called "soilrock" while the virtual resource is called "startc". Additionally, the material handling model is expanded to include the reuse of the excavated material in the improving activity. The details of the additional resource are:

- Activity: improving.

Included processes: transporting the excavated material (soilrock) from the virtual repository to the embankment, filling the embankment, transporting the virtual material (startc) to a virtual repository.

Resource: soilrock (used)

Resource: startc (produced).

The material handling including the activity improving is explained in detail in the following. The virtual resource (Figure C-41), startc, is produced by the activity clearing, only in embankments (clearing of cuts does not produce virtual material), in the quantity equivalent to the improving volume (V_{im}), and by the activity improving in the quantity equivalent to the filling volume (V_{fi}). It is used by the activity excavating in the quantity equivalent to the excavation volume (V_{ex}). The produced material is delivered to the repository in steps at every unit length of embankment, totalling the volume ($V_{im} + V_{fi}$) at the end of clearing and improving. The used material is retrieved from the repository also in steps at every unit length of cut, totalling the volume V_{ex} at the end of excavating. By delivering the virtual resource startc, the activities clearing and improving trigger the activity excavating. This uses the virtual resource startc in the quantity equivalent to the excavating volume (V_{ex}) and produces the resource soilrock in the quantity also equivalent to the excavating volume (V_{ex}). This is used by the activities improving and filling in the quantity equivalent to the improving volume (V_{im}) and filling volume (V_{fi}), respectively. Similarly to the virtual resource startc, the produced material is delivered to the repository in steps at every unit length of cut, totalling the volume V_{ex} at the end of excavating. The used material is retrieved from the repository also in steps at every unit length of embankment, totalling the volume ($V_{im} + V_{fi}$) at the end of filling.

The material handling model controls the mass balance between the cuts and embankments in the four alignments. The mass balance compares materials in the loose volume. The loose volume of in situ and compacted materials can be calculated with loosening, ρ_l , and compacting, ρ_c (formulas

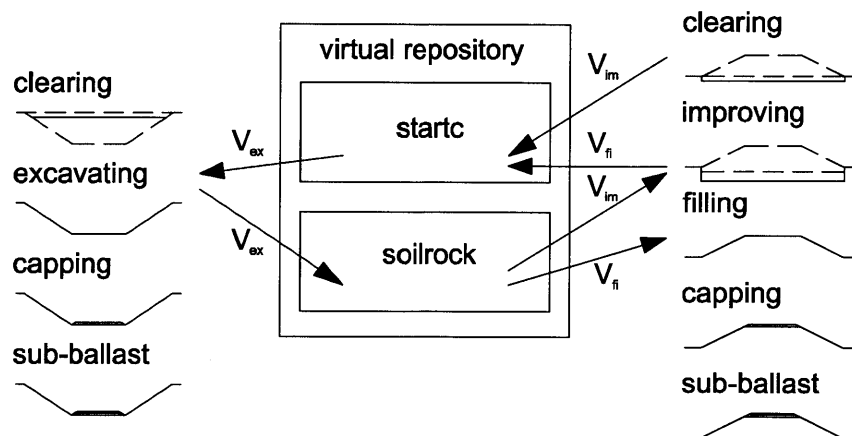


Figure C-41: Material handling. The virtual resource (Figure C-41), *startc*, is produced by the activity *clearing*, only in embankments (*clearing* of cuts does not produce virtual material), in the quantity equivalent to the *improving* volume (V_{im}), and by the activity *improving* in the quantity equivalent to the *filling* volume (V_{fi}). It is used by the activity *excavating* in the quantity equivalent to the *excavation* volume (V_{ex}). By delivering the virtual resource *startc*, the activities *clearing* and *improving* trigger the activity *excavating*. This uses the virtual resource *startc* in the quantity equivalent to the *excavating* volume (V_{ex}) and produces the resource *soilrock* in the quantity also equivalent to the *excavating* volume (V_{ex}). This is used by the activities *improving* and *filling* in the quantity equivalent to the *improving* volume (V_{im}) and *filling* volume (V_{fi}), respectively.

3.10 and 3.11). These are set based on soil and rock descriptions (RAVE 2006a, RAVE 2008a), according to the range suggested in Girmscheid (2005): the loosening factors of rock and soil are set equal to 0.6 and 0.8, respectively, while the compacting factor is set equal to 0.8 for both rock and soil since the filled material consists of a compound of rock and soil. Since the rock and the soil are mainly excavated with blasting and mechanical means, respectively, the blasting and the mechanical excavations produce *soilrock* loose volumes computed with $\rho_l = 0.6$ and $\rho_l = 0.8$, respectively.

The mass balance in the four alignments are in excess of material, meaning that the material remaining in the repository is deposited (section 3.3). Another implication is that the excavation of the cuts may not wait for the embankments to be ready to receive the material, since at the end of the project a substantial amount of excavated material will not be utilized, and will be deposited. To model this degree of freedom, the initial level of the virtual resource *startc* in the repository is set equal to the amount of loose volume that will be excavated but not reused. At the end of

construction, the virtual resource startc is completely consumed by the excavation of all the cuts, while the amount of the resource soilrock in the repository is equal to the volume of the excavated material that is not reused. The amount of the resource soilrock in the repository at the end of construction is then deposited (Table C.25).

Table C.25: Material excavated in cuts that is not reused in the construction of embankments for the four alignments. The material is transported from the virtual repository to the final destination.

Alignment	amount deposited [$10^6 m^3$]
A-a	10.1
A-c	6.1
B-a	5.1
B-b	8.4

Appendix D

Uncertainty in the Construction of the New Portuguese High Speed Rail Line

Chapter 7 presented the results of the application of the uncertainty model to the construction of an alignment (Alignment A-a) of the new Portuguese high speed rail line. The results for the other alignments (A-c, B-a, and B-b) are summarized here.

D.1 Variability

In this section, the results comparing 1) the deterministic total cost and total time and 2) the simulation modeling the variability in the construction process are presented for alignments A-c, B-a, and B-b. As for alignment A-a, while the deterministic total cost and total time are represented by a point, the results from the simulation of the variability are a scattergram of points, one for each simulation. The scattergrams are the clear manifestation that the uncertainty in total cost and total time is caused by variability. Thus, the uncertainty cannot be expressed with a point (deterministic total cost and total time) but rather with a range of possible outcomes. More detailed findings and conclusions are discussed in section 7.4.1.

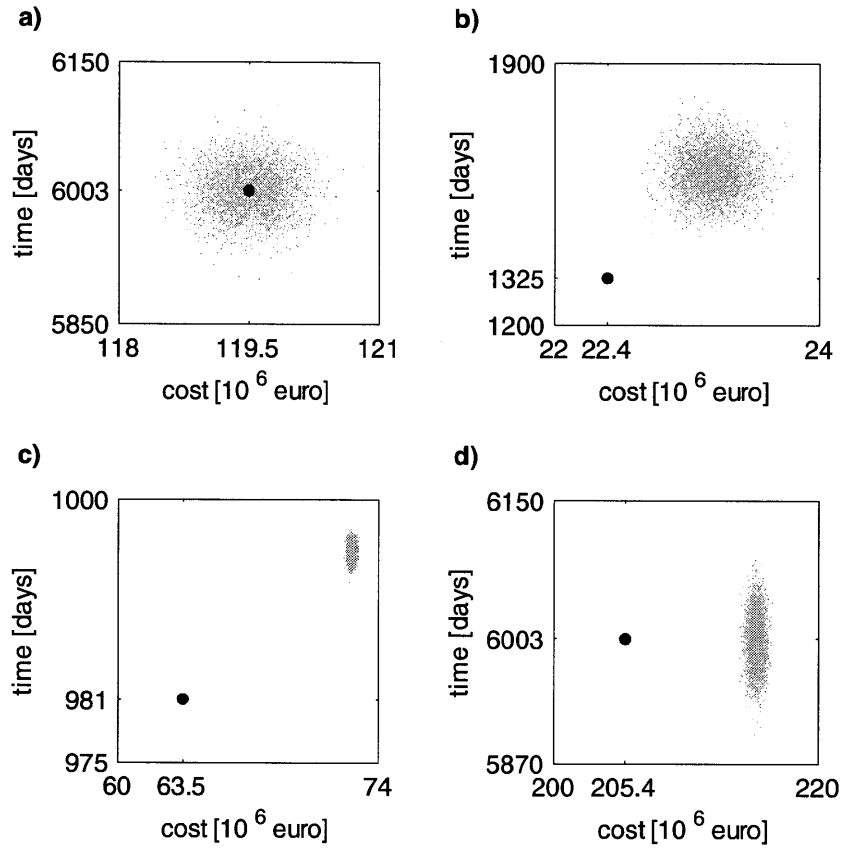


Figure D-1: Alignment A-c, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

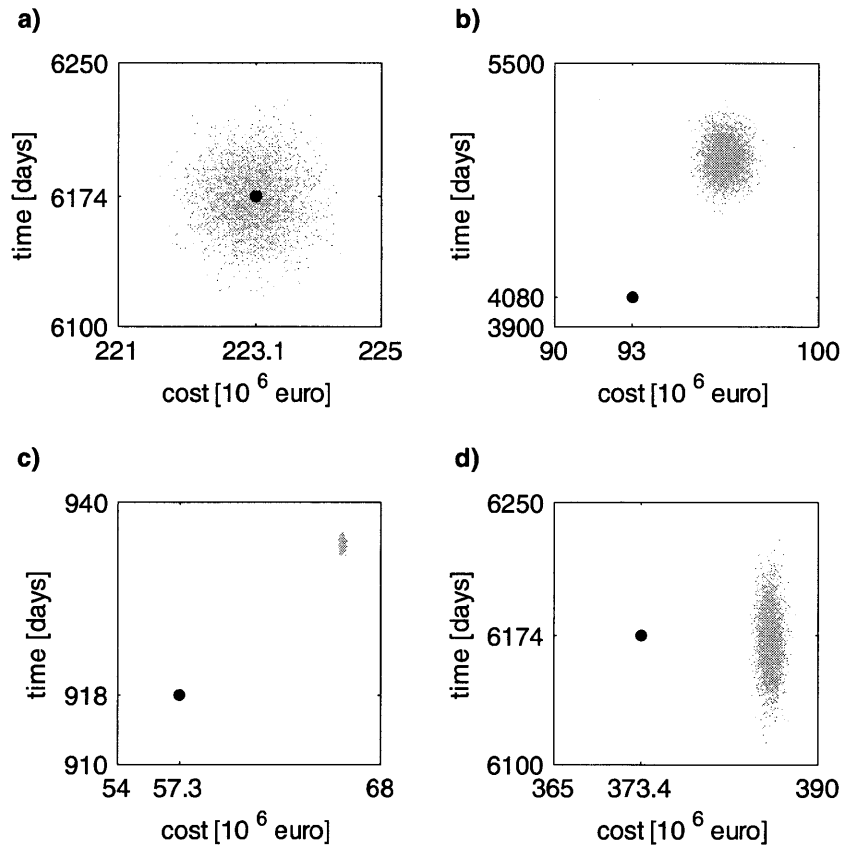


Figure D-2: Alignment B-a, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

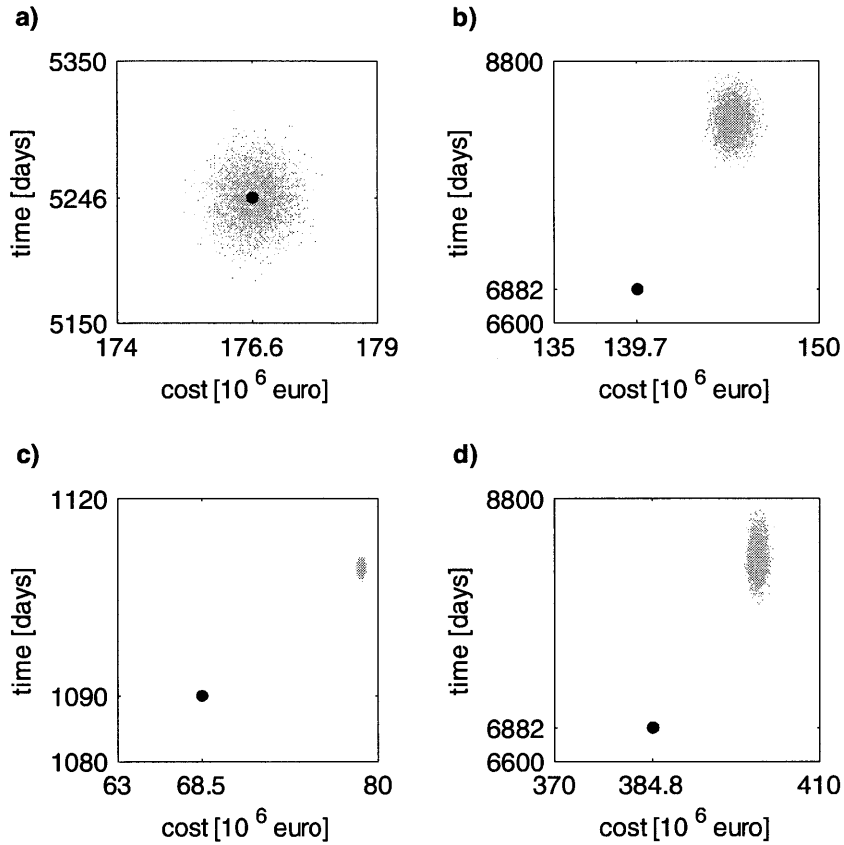


Figure D-3: Alignment B-b, deterministic total cost and total time (black) and cost-time scattergrams that represent the variability (gray) for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

D.2 Cost Correlations

In this section, the results comparing 1) the simulation modeling the variability in the construction process and 2) the simulation modeling the variability and the correlations between the costs of the activities are presented for alignments A-c, B-a, and B-b. As for alignment A-a, the range of the total cost distribution increases dramatically due to cost correlations. Further findings and conclusions are discussed in section 7.4.2.

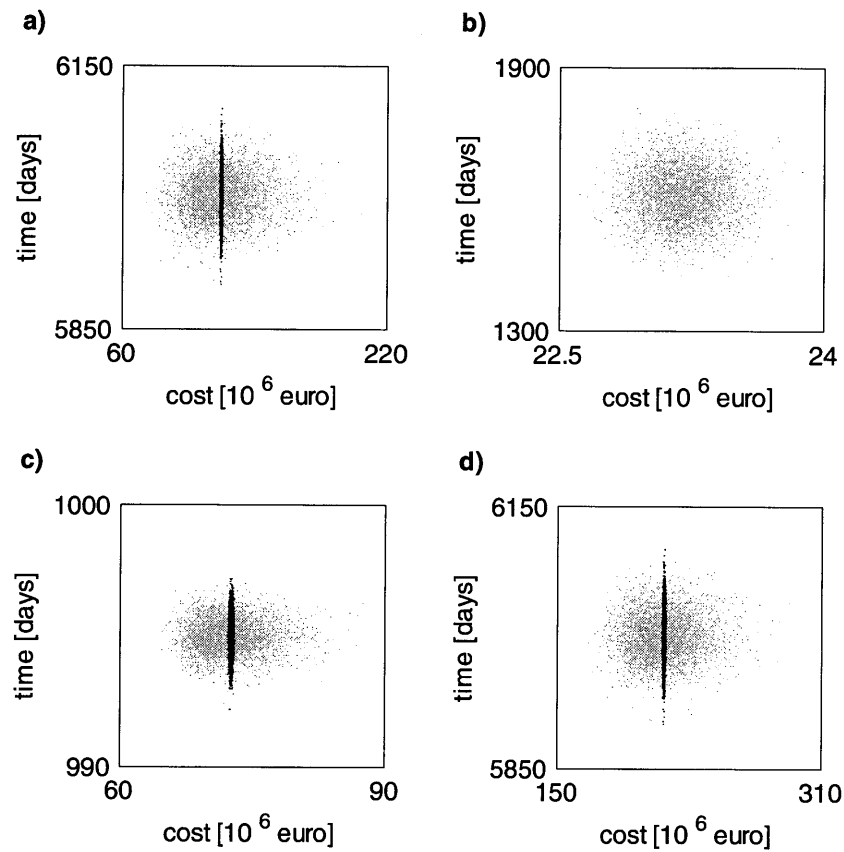


Figure D-4: Alignment A-c, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

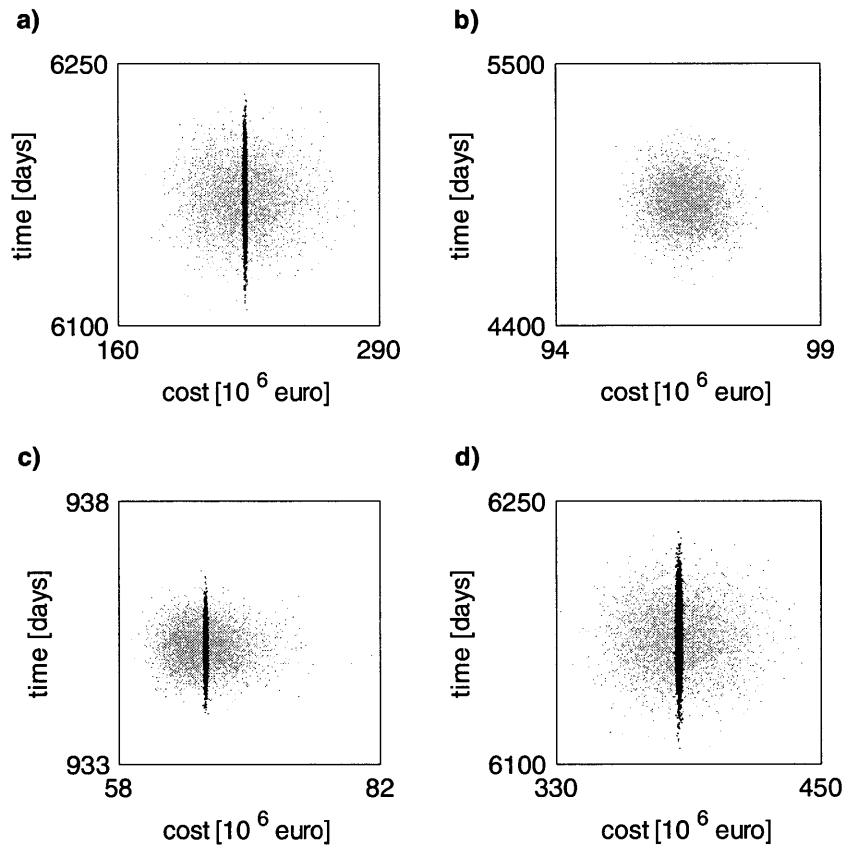


Figure D-5: Alignment B-a, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

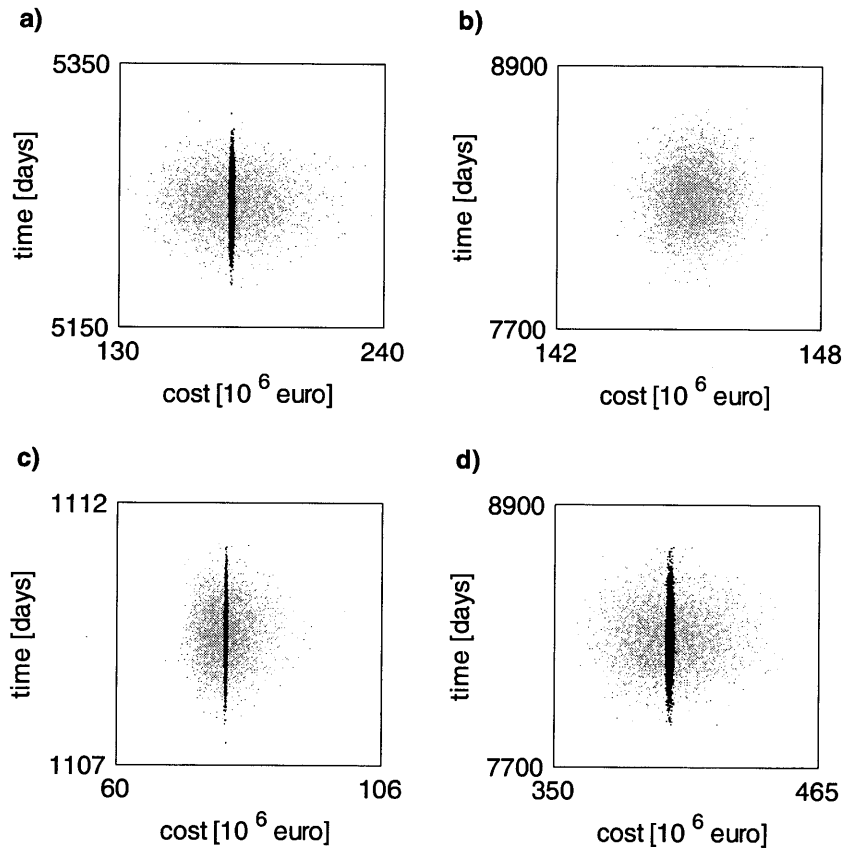


Figure D-6: Alignment B-b, uncertainty sources: variability (black) versus variability and cost correlations (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

D.3 Disruptive Events

In this section, the results comparing 1) the simulation modeling the variability in the construction process and the correlations between the costs of the activities and 2) the simulation modeling the variability, the cost correlations and the disruptive events are presented for alignments A-c, B-a, and B-b. As for alignment A-a, due to disruptive events the ranges of the total cost and total time distributions increase significantly albeit with different magnitudes and patterns depending on the structures analyzed. Further findings and the conclusions discussed in section 7.4.3 also apply to alignments A-c, B-a, and B-b with one exception. While in Alignments A-a, A-c, and B-a the scattergram for all structures shows similar patterns as for the tunnels, in Alignment B-b it shows a different pattern for the following reasons:

1. Viaducts are constructed in parallel with the tunnels and the sequence of cuts and embankments. Since the viaduct construction time is the longest (Figure D-9), the viaducts determine the construction time of all structures.
2. The total cost for all structures is equal to the sum of all costs. Thus, the scattergram of the total cost and total time of all structures (Figure D-9d) is the scattergram of the total cost of all structures and the total time of the viaducts.

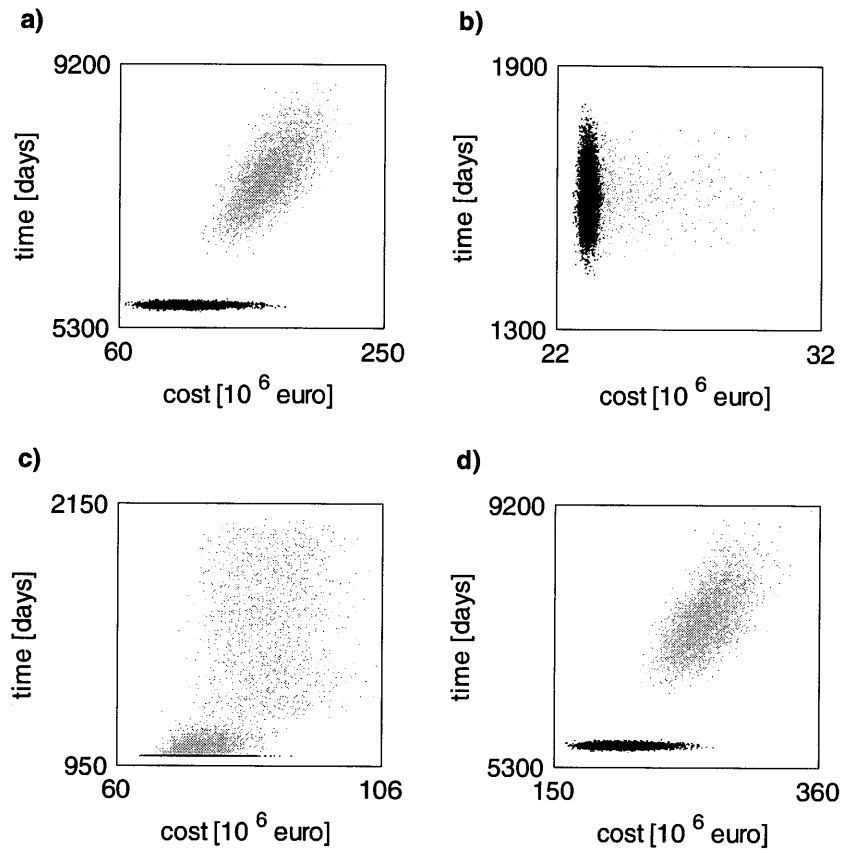


Figure D-7: Alignment A-c, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

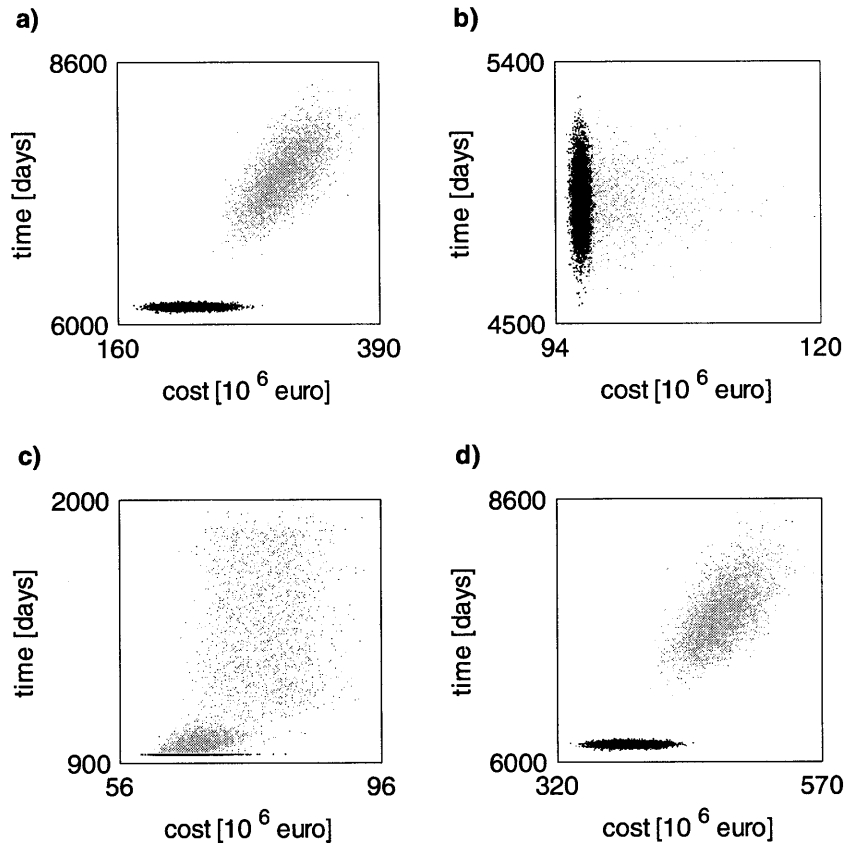


Figure D-8: Alignment B-a, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

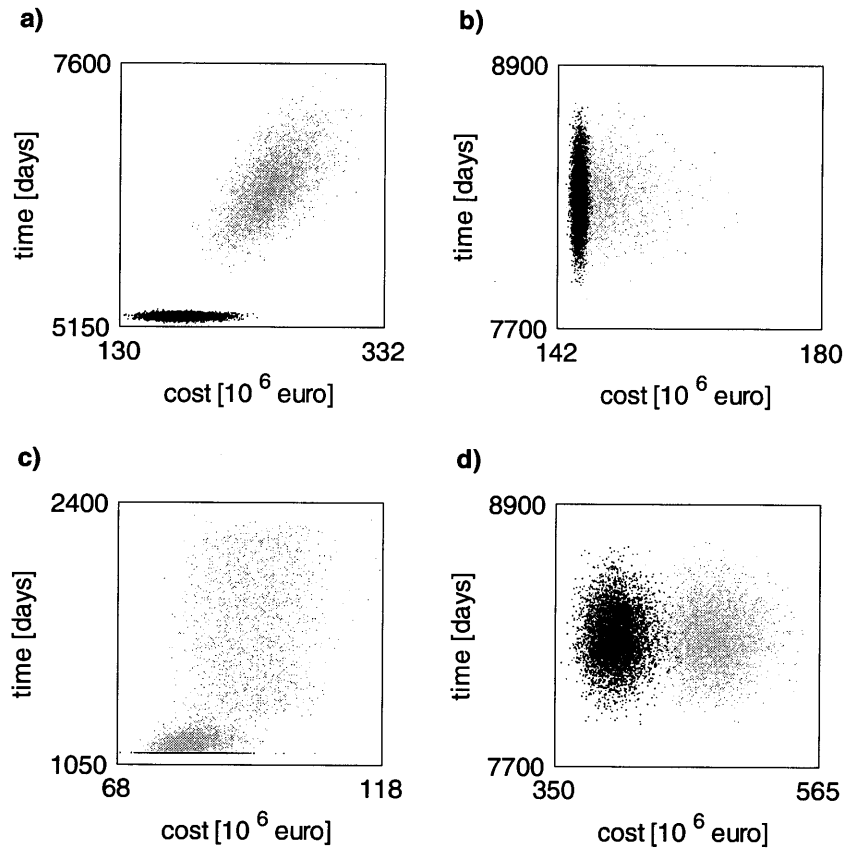


Figure D-9: Alignment B-b, uncertainty sources: variability and cost correlations (black) versus variability, cost correlations and disruptive events (gray). Cost-time scattergrams for the construction of a) tunnels, b) viaducts, c) cuts and embankments, and of d) all the structures.

D.4 Value of the Construction and Uncertainty Models

In this section, the deterministic total cost and total time and the 90th percentiles of the total cost and total time distributions of alignment A-c, B-a, and B-b are compared. Similarly to alignment A-a (section 7.6), the impacts of variability, cost correlations, and disruptive events on the total cost and total time distributions vary depending on the structure suggesting the use of differentiated strategies to contain cost overrun and delays for the stakeholders of the project. Most importantly, the cumulative impact of the three sources of uncertainty caused the construction cost and time to increase significantly. Again, the largest increases are observed in the tunnel construction cost and in the earthwork (cuts and embankments) construction time.

Similarly to section D.3, alignments A-a, A-c and B-a differ from alignment B-b in the increase of the total time for the construction of all structures. Since the sequence of tunnels is constructed in parallel with the sequence of viaducts and with the sequence of cuts and embankments, the structure type with the longest construction time determines the total time for the construction of all structures. In alignments A-a, A-c and B-a tunnels have the longest construction time, while in alignment B-b viaducts have the longest construction time. It follows that in alignments A-a, A-c and B-a the increases of the total time for the construction of all structures is equal to the increases of the total time for the construction of the tunnels, while in alignment B-b the increases of the total time for the construction of all structures is equal to the increases of the total time for the construction of the viaducts.

Table D.1: Alignment A-c. Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 56% in tunnel construction, while the largest total time increase is 90% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations.

Total cost	tunnels		viaducts		cuts and embankments		all structures	
	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]
deterministic	119.5	–	22.4	–	63.5	–	205.4	–
90 th percentile variability	119.9	0.3	23.4	4.6	72.7	14.5	215.8	5.5
variability and cost correlations	140.5	17.6	23.4	4.6	76.5	20.5	236.7	15.2
variability, cost correlations and disruptive events	186.8	56.3	24.1	7.4	91.4	43.9	293.7	43.0
Total time	tunnels		viaducts		cuts and embankments		all structures	
	[days]	[%]	[days]	[%]	[days]	[%]	[days]	[%]
deterministic	6,003	–	1,325	–	981	–	6,003	–
90 th percentile variability	6,035	0.5	1,680	26.8	996	1.5	6,035	0.5
variability and cost correlations	6,035	0.5	1,680	26.8	996	1.5	6,035	0.5
variability, cost correlations and disruptive events	8,065	34.3	1,681	26.9	1,868	90.4	8,065	34.3

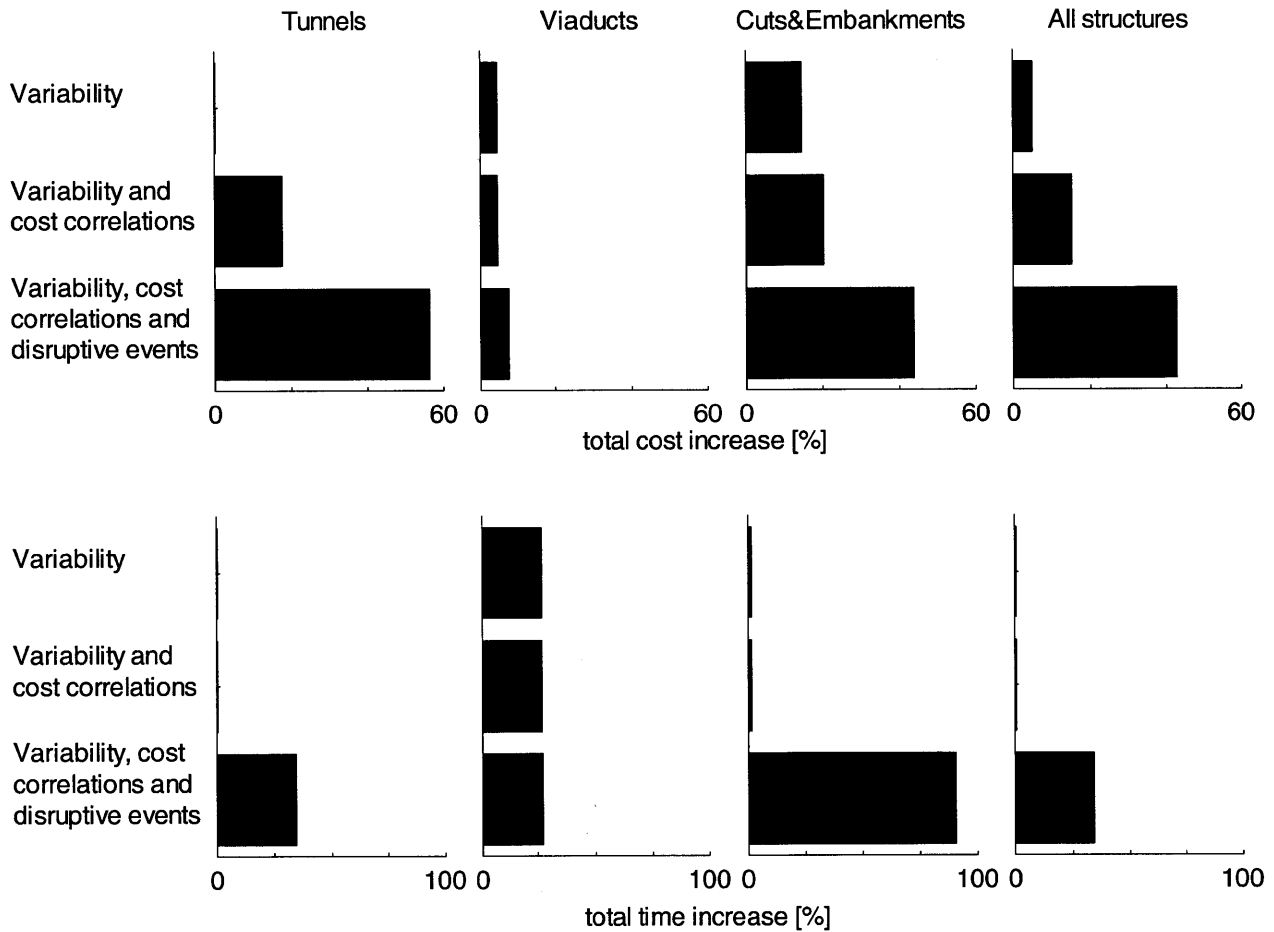


Figure D-10: Alignment A-c. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

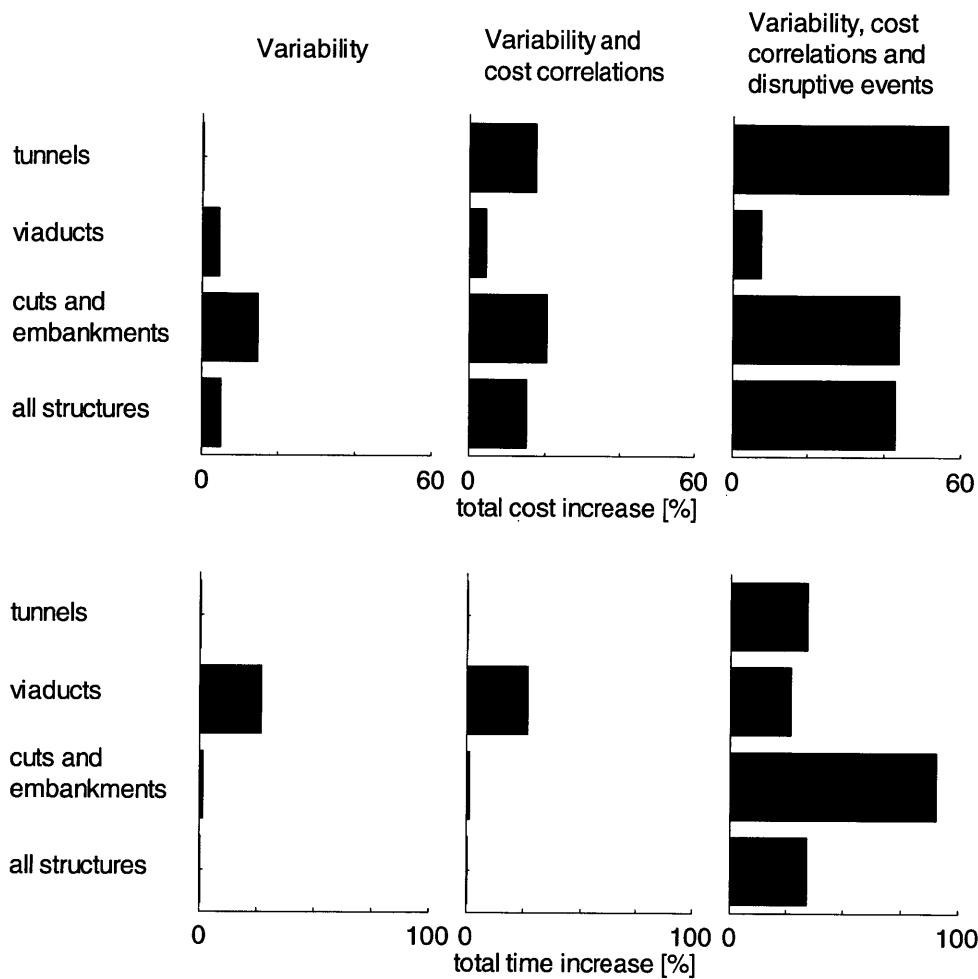


Figure D-11: Alignment A-c. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

Table D.2: Alignment B-a. Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 48% in tunnel construction, while the largest total time increase is 90% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations.

Total cost	tunnels		viaducts		cuts and embankments		all structures	
	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]
deterministic	223.1	–	93.0	–	57.3	–	373.4	–
90 th percentile								
variability	223.6	0.2	97.0	4.3	66.0	15.3	386.3	3.4
variability and cost correlations	243.0	8.9	97.0	4.3	69.0	20.5	405.5	8.6
variability, cost correlations and disruptive events	331.4	48.5	100.8	8.4	82.8	44.6	504.4	35.1
Total time	tunnels		viaducts		cuts and embankments		all structures	
	[days]	[%]	[days]	[%]	[days]	[%]	[days]	[%]
deterministic	6,174	–	4,080	–	918	–	6,174	–
90 th percentile								
variability	6,194	0.3	5,044	23.6	936	1.9	6,194	0.3
variability and cost correlations	6,193	0.3	5,044	23.6	936	1.9	6,193	0.3
variability, cost correlations and disruptive events	7,818	26.6	5,045	23.7	1,745	90.1	7,818	26.6

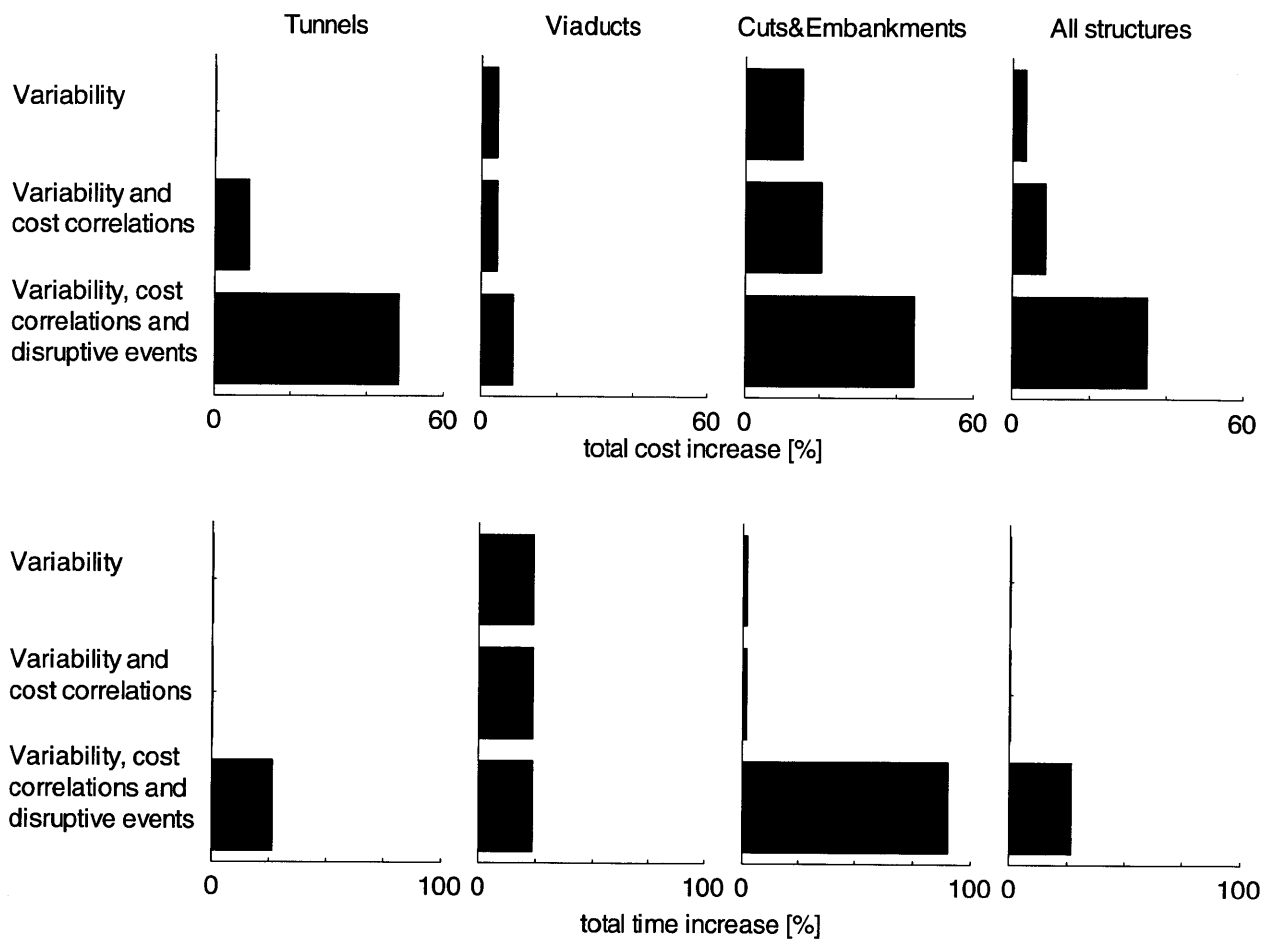


Figure D-12: Alignment B-a. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

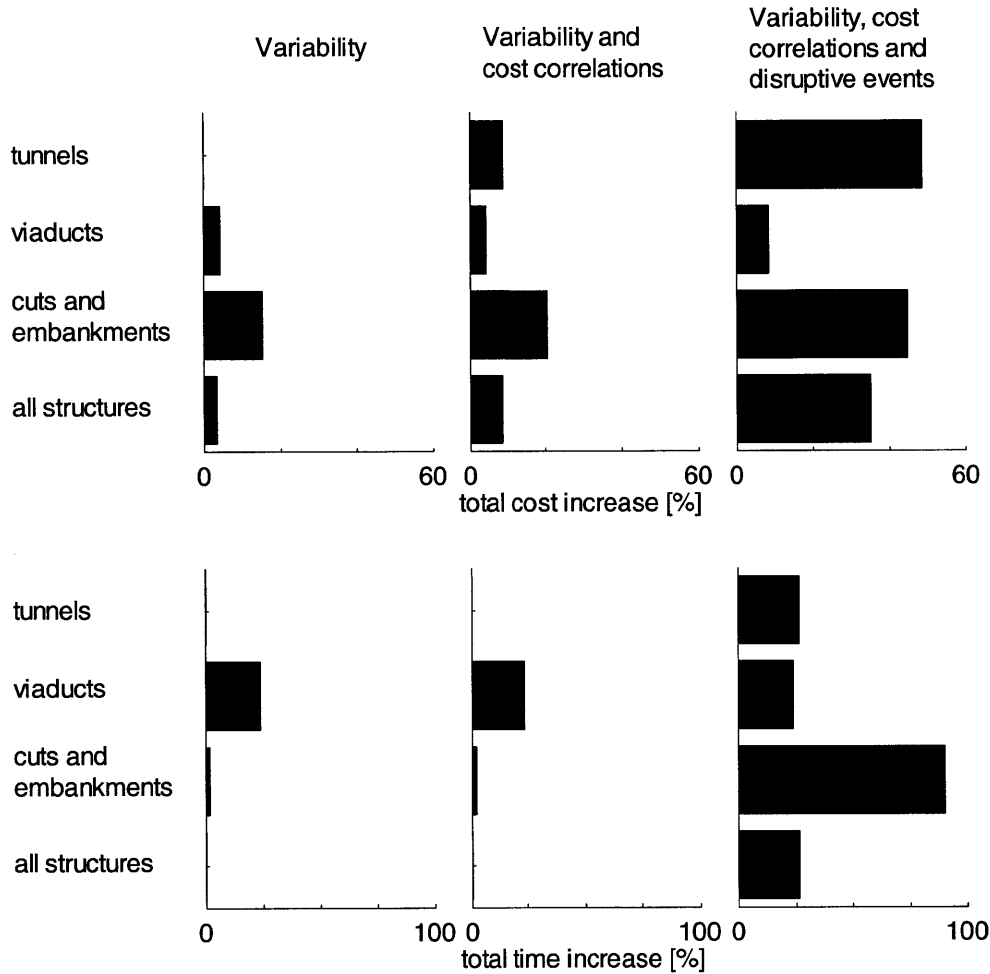


Figure D-13: Alignment B-a. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

Table D.3: Alignment B-b. Deterministic total cost and total time, 90th percentiles of the total cost and total time distributions, and increases (see expression 7.13) in total cost and total time for different structures (tunnels, viaducts, cuts and embankments, and all structures) depending on the sources of uncertainty (variability, variability and cost correlations, and variability, cost correlations and disruptive events). The largest total cost increase is 52% in tunnel construction, while the largest total time increase is 90% in earthwork (cuts and embankments) construction. Note that the increases [%] represent the differences between the 90th percentiles of the total cost/time distribution and the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the increase due to variability, cost correlations, and disruptive events and the increase due to variability and cost correlations.

Total cost	tunnels		viaducts		cuts and embankments		all structures	
	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]	[10 ⁶ euro]	[%]
deterministic	176.6	–	139.7	–	68.5	–	384.8	–
90 th percentile variability	177.1	0.3	145.8	4.4	79.0	15.4	401.5	4.3
variability and cost correlations	194.4	10.1	145.8	4.4	83.0	21.2	419.3	9.0
variability, cost correlations and disruptive events	268.2	51.9	151.3	8.3	99.1	44.6	505.2	31.3
Total time	tunnels		viaducts		cuts and embankments		all structures	
	[days]	[%]	[days]	[%]	[days]	[%]	[days]	[%]
deterministic	5,246	–	6,882	–	1,090	–	6,882	–
90 th percentile variability	5,268	0.4	8,453	22.8	1,110	1.8	8,453	22.8
variability and cost correlations	5,268	0.4	8,453	22.8	1,110	1.8	8,453	22.8
variability, cost correlations and disruptive events	6,794	29.5	8,457	22.9	2,067	89.6	8,457	22.8

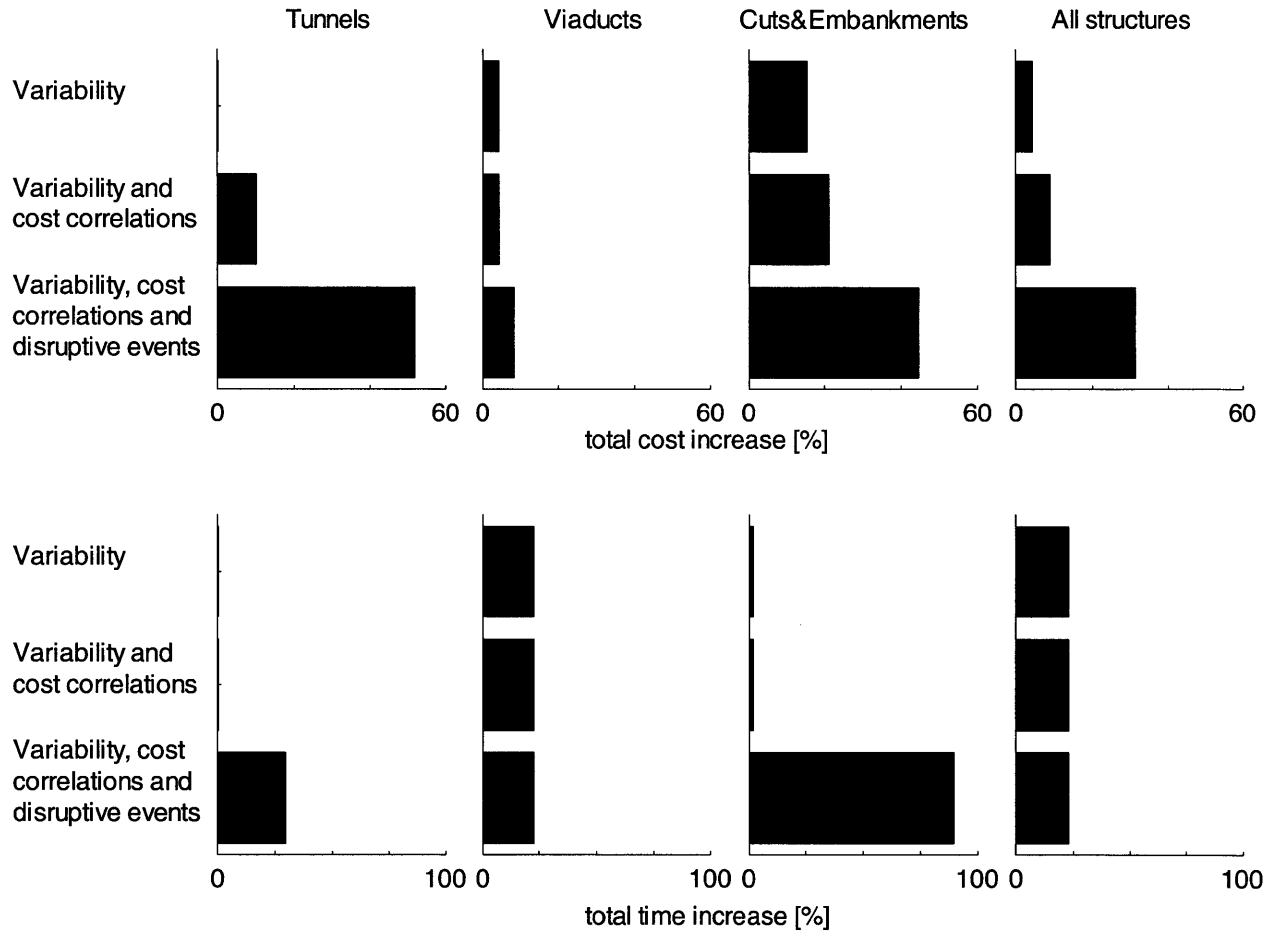


Figure D-14: Alignment B-b. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different structures. While for tunnels and for cuts and embankments the largest increases are caused by disruptive events, for viaducts they are caused by variability. Thus, depending on the structure the impact of the uncertainty sources varies. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

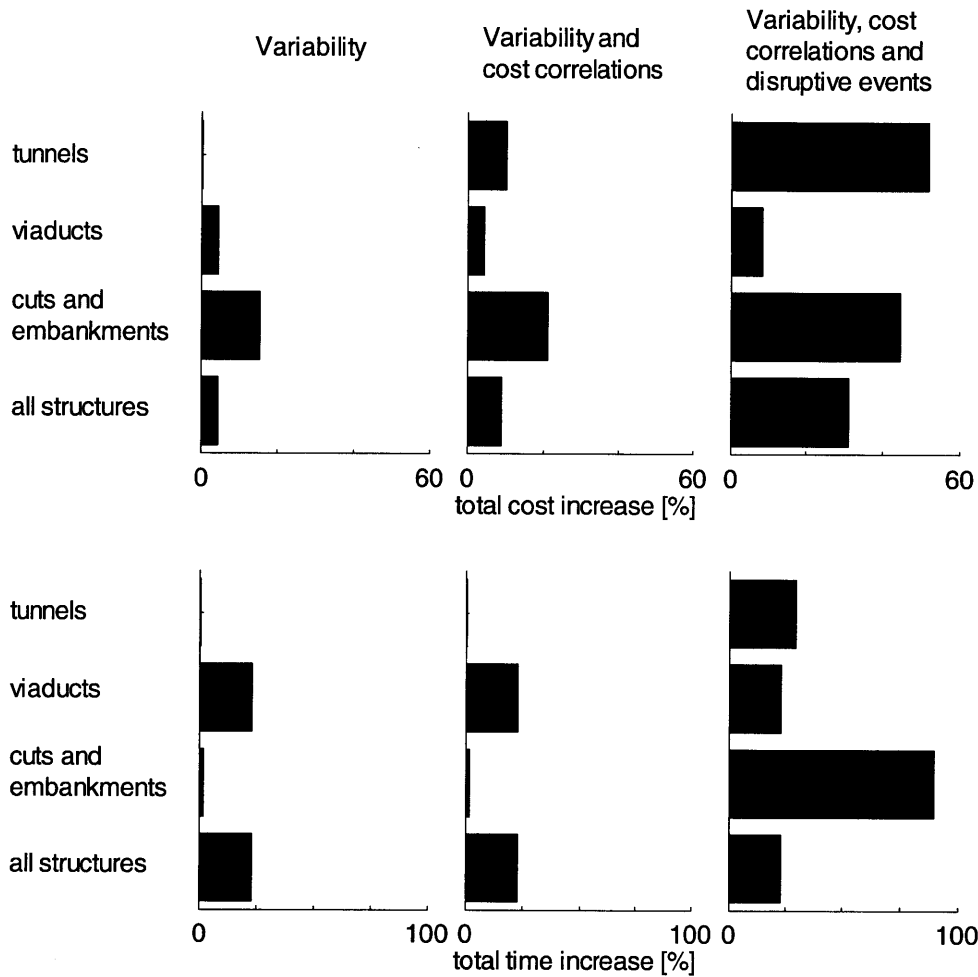


Figure D-15: Alignment B-b. Increases (from deterministic total cost/time to the 90th percentile of the total cost/time distribution; see expression 7.13) in total cost (above) and total time (below) for different sources of uncertainty. While variability has the largest impact on the total cost and total time of viaducts, cost correlations have a significant impact on tunnels. However, for all structures but viaducts the largest impact is caused by the disruptive events. Note that the bars represent the increase from the deterministic total cost/time (see expression 7.13). Thus, the increase due to e.g. disruptive events is the difference between the bar of the increase due to variability and cost correlations and the bar of the increase due to variability, cost correlations, and disruptive events.

Literature References

- AbouRizk, S. (2010). Role of simulation in construction engineering and management, *Journal of Construction Engineering and Management* **136**(10): 1140–1153.
- Anderson, S. D., Molenaar, K. R. & Schexnayder, C. J. (2007). *NCHRP 574: Guidance for cost estimation and management for highway projects during planning, programming, and preconstruction*, Transportation Research Board, Washington, D.C.
- Ang, A. H.-S. & Tang, W. H. (2007). *Probability concepts in engineering : emphasis on applications in civil and environmental engineering*, 2nd edn, Wiley, Hoboken, NJ.
- Ayyub, B. M. (2001). *Elicitation of expert opinions for uncertainty and risks*, CRC Press, Boca Raton, Fla.
- Back, W. E., Boles, W. W. & Fry, G. T. (2000). Defining triangular probability distributions from historical cost data, *Journal of Construction Engineering and Management* **126**(1): 29–37.
- Bertsekas, D. P. & Tsitsiklis, J. N. (2002). *Introduction to probability*, Athena Scientific, Belmont, Mass.
- Caltrans (2007). *Project risk management handbook*, California. Dept. of Transportation Office of Statewide Project Management Improvement, Sacramento, Calif.
- Cario, M. C. & Nelson, B. L. (1997). Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix, *Technical report*, Northwestern University, Evanston, IL.
- Chau, K. W. (1995). The validity of the triangular distribution assumption in monte carlo simulation of construction costs: empirical evidence from hong kong, *Construction Management and Economics* **13**(1): 15.
- Clemen, R. T., Fischer, G. W. & Winkler, R. L. (2000). Assessing dependence: Some experimental results, *Management Science* **46**(8): 1100.
- Clemen, R. T. & Reilly, T. (1999). Correlations and copulas for decision and risk analysis, *Management Science* **45**(2): 208–224.
- Cooke, R. (1991). *Experts in uncertainty : opinion and subjective probability in science*, Environmental ethics and science policy series, Oxford University Press, New York.
- Einstein, H. (2004). Decision aids for tunneling: Update, *Transportation Research Record: Journal of the Transportation Research Board* **1892**(-1): 199–207.

- Einstein, H. H. (2002). Risk assessment and management in geotechnical engineering, *Workshop on Risk Analysis of the Portuguese Congress for Geotechnique*, Portugal.
- Einstein, H. H., Indermitte, C., Sinfield, J., Descoedres, F. P. & Dudt, J.-P. (1999). Decision aids for tunneling, *Transportation Research Record* pp. 6–13.
- Einstein, H. H., Min, S. & Moret, Y. (2011). Discussion of "role of simulation in construction engineering and management", *Under Review* .
- Embrechts, P., Lindskog, F. & McNeil, A. (2003). Modelling dependence with copulas and applications to risk management, in T. R. Svetozar (ed.), *Handbook of Heavy Tailed Distributions in Finance*, North-Holland, Amsterdam, pp. 329–384.
- Embrechts, P., McNeil, A. J. & Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls, in M. A. H. Dempster (ed.), *Risk management : value at risk and beyond*, Cambridge University Press, Cambridge ; New York, pp. 176–223.
- Fell, R., Ho, K., Lacasse, S. & Leroi, E. (2005). A framework for landslide risk assessment and management, in O. Hungr, R. Fell, R. Couture & E. Eberhardt (eds), *International Conference on Landslide Risk Management*, Taylor and Francis, Vancouver, Canada.
- FHWA (2005). Construction management practices in canada and europe, *Technical Report FHWA-PL-05-010*, U.S. Department of Transportation, Federal Highway Administration (FHWA), Washington, DC.
- Flyvbjerg, B. (2006). From nobel prize to project management: Getting risks right, *Project Management Journal* **37**(3): 5.
- Flyvbjerg, B. (2007). Policy and planning for large-infrastructure projects: problems, causes, cures, *Environment and Planning B: Planning and Design* **34**(4): 578–597.
- Flyvbjerg, B. (2008). Public planning of mega-projects: overestimation of demand and overestimation of demand and underestimation of costs, in H. Premius, B. Flyvbjerg & B. v. Wee (eds), *Decision-Making on Mega-Projects*, Transportation economics, management and policy, Edward Elgar, Northampton, MA.
- Flyvbjerg, B., Bruzelius, N. & Rothengatter, W. (2003). *Megaprojects and risk : an anatomy of ambition*, Cambridge University Press, New York.
- Flyvbjerg, B. & Cowi (2004). Procedures for dealing with optimism bias in transport planning: Guidance document, *Technical report*, UK Department for Transport, London.
- Flyvbjerg, B., Holm, M. S. & Buhl, S. (2002). Underestimating costs in public works projects: *error or lie?*, *Journal of the American Planning Association* **68**(3): 279 – 295.
- Fouracre, P. R., Allport, R. J. & Thomson, J. M. (1990). The performance and impact of rail mass transit in developing countries, *Technical Report Research Report 278*, Transport and Road Research Laboratory, Crowthorne, UK.
- FTA (2004). Risk assessment methodologies and procedures, *Technical Report Report Nr DTFT60-98-D-41013*, Federal Transit Administration, Washington, DC.

- Genest, C. & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask, *Journal of Hydrologic Engineering* **12**(4): 347–368.
- Genest, C., Quessy, J.-F. & Remillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation, *Scandinavian Journal of Statistics* **33**(2): 337–366.
- Ghosh, S. & Henderson, S. G. (2002). Chessboard distributions and random vectors with specified marginals and covariance matrix, *Operations Research* **50**(5): 820–834.
- Ghosh, S. & Henderson, S. G. (2003). Behavior of the norta method for correlated random vector generation as the dimension increases, *ACM Trans. Model. Comput. Simul.* **13**(3): 276–294.
- Gilovich, T., Griffin, D. W. & Kahneman, D. (2002). *Heuristics and biases : the psychology of intuitive judgment*, Cambridge University Press, Cambridge, UK ;New York.
- Girmscheid, G. (2005). *Leistungsermittlungshandbuch fuer Baumaschinen und Bauprozessen*, 3rd edn, Springer, Berlin.
- Guglielmetti, V., Grasso, P., Mahtab, A. & Xu, S. (2008). *Mechanized tunnelling in urban areas : design methodology and construction control*, Taylor and Francis, London; New York.
- Haas, C. & Einstein, H. H. (2002). Updating the decision aids for tunneling, *Journal of Construction Engineering and Management* **128**(1): 40–48.
- Hall, P. G. (1980). *Great planning disasters*, Weidenfeld and Nicolson, London.
- Hallowell, M. R. & Gambatese, J. A. (2010). Qualitative research: Application of the delphi method to cem research, *Journal of Construction Engineering and Management* **136**(1): 99–107.
- Halpin, D. W. & Woodhead, R. W. (1976). *Design of Construction and Process Operations*, John Wiley & Sons, Inc.
- Iman, R. L. & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables, *Communications in Statistics - Simulation and Computation* **11**(3): 311 – 334.
- Institution of Civil Engineers and Faculty of Actuaries (2005). *RAMP : Risk Analysis and Management for Projects : A strategic framework for managing project risk and its financial implications*, Thomas Telford Ltd, London.
- ISSMGE (2004). Glossary of risk assessment terms, *Technical report*, Technical Committee on Risk Assessment and Management, International Society of Soil Mechanics and Geotechnical Engineering, London, United Kingdom.
- Johnson, M. & Ramberg, J. (1978). Transformations of the multivariate normal distribution with applications to simulation, *11th International Conference on System Sciences*, Honolulu, Hawaii.
- Kollarou, C. (2002). *Excavation materials handling in the Loetschberg Base Tunnel using Decision Aids for Tunneling (DAT)*, Master thesis, Massachusetts Institute of Technology.

- Kruskal, W. H. (1958). Ordinal measures of association, *Journal of the American Statistical Association* **53**(284): 814–861.
- Kurowicka, D. & Cooke, R. (2006). *Uncertainty analysis with high dimensional dependence modelling*, Wiley, Chichester, England ; Hoboken, NJ.
- Leavitt, D., Ennis, S. & McGovern, P. (1993). The cost escalation of rail projects: Using previous experience to re-evaluate the cal-speed estimates, *Technical Report Working Paper No. 567*, Institute of Urban and Regional Development, University of California, Berkeley.
- Leite, R. (2009). Personal communication on the construction costs of viaducts.
- Lewis, H. (1986). The metro report: The impact of metro and public transport integration in tyne and wear, *Technical report*, Tyne and Wear Passenger Transport Executive, Newcastle, UK.
- Li, S. T. & Hammond, J. L. (1975). Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients, *IEEE Transactions on Systems, Man, and Cybernetics* **5**: 557–561.
- Lovullo, D. & Kahneman, D. (2003). Delusions of success, *Harvard Business Review* **81**(7): 56–63.
- Lurie, P. M. & Goldberg, M. S. (1998). An approximate method for sampling correlated random variables from partially-specified distributions, *Management Science* **44**(2): 203–218.
- Marino, M. (2009). Personal communication on the construction of cuts and embankments.
- Martinez, J. C. & Ioannou, P. G. (1999). General-purpose systems for effective construction simulation, *Journal of Construction Engineering and Management* **125**(4): 265–276.
- Marzer, C. (2002). Decision aids for tunneling (dat): Development of the resource model, *Technical report*, Massachusetts Institute of Technology, Cambridge, MA, United States.
- Menezes, P. (2010). Personal communication on the estimation of the volumes of cuts and embankments.
- Merewitz, L. (1973). How do urban rapid transit projects compare in cost estimate experience?, *Technical Report Reprint No. 104*, Institute of Urban and Regional Development, University of California, Berkeley.
- Moavenzadeh, F., Einstein, H., Markow, M. & Minott, C. (1974). Tunnel cost model: A stochastic simulation of hard rock tunneling, *Technical Report MIT Reports to NSF-Rann*, Vols. 1-9.
- Molenaar, K. R. (2005). Programmatic cost risk analysis for highway megaprojects, *Journal of Construction Engineering and Management* **131**(3): 343–353.
- Molenaar, K. R., Anderson, S. D. & Schexnayder, C. J. (2010). *NCHRP 658: Guidebook on risk analysis tools and management practices to control transportation project costs*, Transportation Research Board, Washington, D.C.
- National Audit Office (1985). Expenditure on motorways and trunk roads, *Technical report*, National Audit Office, Department of Transport, London, UK.

- National Audit Office (1992). Contracting for roads, *Technical report*, National Audit Office, Department of Transport, London, UK.
- National Audit Office & Scottish Development Department (1988). Contracting for roads, *Technical report*, Her Majesty's Stationary Office, London, UK.
- Nelsen, R. B. (2006). *An introduction to copulas*, 2nd edn, Springer, New York.
- Newton, S. (1992). Methods of analysing risk exposure in the cost estimates of high quality offices, *Construction Management and Economics* **10**(5): 431 – 449.
- O'Donnell, L. (2008). Personal communication on the construction of viaducts.
- Peterson, C. R. & Einstein, H. H. (1993). Manufacturing underground space, *Proc. NSF Design and Manufacturing Systems Grantees Conference*.
- Pfisterer, T. (2005). Kosten, mehrkosten und politik in der neat, *Fachtagung fÄijr Untertagbau*, Luzern, Switzerland.
- Pickrell, D. H. (1990). Urban rail transit projects: Forecast versus actual ridership and cost, *Technical report*, U.S. Department of Transportation, Washington, DC.
- PMI (2008). *A guide to the project management body of knowledge (PMBOK Guide)*, 4th edn, Project Management Institute, Inc., Newtown Square, Pa.
- Pouliquen, L. Y. (1970). *Risk analysis in project appraisal*, World Bank staff occasional papers, The Johns Hopkins University Press, Baltimore.
- RAVE (2006a). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1a, troco aeroporto francisco sa carneiro - braga/barcelos, estudo previo, volume 03 - geologia e geotecnia, *Technical report*.
- RAVE (2006b). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1a, troco aeroporto francisco sa carneiro - braga/barcelos, estudo previo, volume 05 - terraplenagens e drenagem, *Technical report*.
- RAVE (2006c). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1a, troco aeroporto francisco sa carneiro - braga/barcelos, estudo previo, volume 08 - obras de arte: Pontes e viaductos, *Technical report*.
- RAVE (2006d). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1a, troco aeroporto francisco sa carneiro - braga/barcelos, estudo previo, volume 09 - obras de arte: Tuneis, *Technical report*.
- RAVE (2008a). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1b, braga - valenca, estudo previo, volume 03 - geologia e geotecnia, *Technical report*.
- RAVE (2008b). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1b, braga - valenca, estudo previo, volume 05 - terraplenagens e drenagem, *Technical report*.
- RAVE (2008c). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1b, braga - valenca, estudo previo, volume 08 - obras de arte: Pontes e viaductos, *Technical report*.

- RAVE (2008d). Ligacao ferroviaria de alta velocidade entre porto e vigo, lote1b, braga - valenca, estudo previo, volume 09 - obras de arte - tuneis, *Technical report*.
- Reilly, J. J. (2005). Cost estimating and risk - management for underground projects, *Proceedings of the International Tunneling Association Conference*, International Tunneling Association, Istanbul, Turkey.
- Reilly, J. J., McBride, M., Sangrey, D., MacDonald, D. & Brown, J. (2004). The development of cevp - wsdot's cost-risk estimating process, *Technical report*, Washington State Department of Transportation, Olympia, WA.
- Riksrevisionsverket (1994). Infrastrukturinvesteringar: En kostnadsjamforelse mellan plan och utfall i 15 storre projekt inom vagverket och banverket, *Technical report*, Riksrevisionsverket, Stockholm.
- Robert, C. P. & Casella, G. (2004). *Monte Carlo statistical methods*, 2nd edn, Springer, New York.
- Salvucci, F. (2007). Personal communication on the construction of the big dig.
- Salvucci, F. P. (2003). The "big dig" of boston, massachusetts: lessons to learn, in J. Saveur (ed.), *(Re)Claiming the underground space : proceedings of the ITA world tunnelling congress, 12-17 april 2003, Amsterdam, The Netherlands*, Vol. 1, Balkema, Lisse.
- Schmeiser, B. W. & Lal, R. (1982). Bivariate gamma random vectors, *Operations Research* **30**(2): 355–374.
- Schoelzel, C. & Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research - introduction to the copula approach, *Nonlin. Processes Geophys.* **15**(5): 761–772.
- Sinfield, J. V. & Einstein, H. H. (1996). Evaluation of tunneling technology using the "decision aids for tunneling", *Tunneling and Underground Space Technology* **11**: 491–504.
- Slaughter, E. S. (1999). Assessment of construction processes and innovations through simulation, *Construction Management and Economics* **17**(3): 341–350.
- Strang, G. (2006). *Linear algebra and its applications*, 4th edn, Thomson, Brooks/Cole, Belmont, CA.
- Teuscher, P. (2007). Ein kurzer rÄijckblick und erste erkenntnisse, *Fachtagung fÄijr Untertagbau*, Luzern, Switzerland.
- tHR (2007). RelÄatorio de custos.
- Touran, A. (1993). Probabilistic cost estimating with subjective correlations, *Journal of Construction Engineering and Management* **119**(1): 58–71.
- Touran, A. & Suphot, L. (1997). Rank correlations in simulating construction costs, *Journal of Construction Engineering and Management* **123**(3): 297–301.
- Touran, A. & Wiser, E. P. (1992). Monte carlo technique with correlated random variables, *Journal of Construction Engineering and Management* **118**(2): 258–272.

- Vejdirektoratet (1995). Notat om anlægsregnskaber, *Technical report*, Danish Road Directorate, Copenhagen.
- Walmsley, D. A. & Pickett, M. W. (1992). The cost and patronage of rapid transit systems compared with forecasts, *Technical Report Research Report 352*, Transport Research Laboratory, Crowthorne, UK.
- WSDOT (2005). A policy for cost risk assessment, *Technical report*, Washington State Department of Transportation (WSDOT), Olympia, WA.
- WSDOT (2008). Wsdot guidelines for cra-cevp workshops, *Technical report*, Washington State Department of Transportation (WSDOT), Olympia, WA.
- Yang, I. (2006). Using gaussian copula to simulate repetitive projects, *Construction Management and Economics* **24**(9): 901–909.
- Zbinden, P. (2007). Halbzeit an der gotthardachse - ein bilck zurÄijck auf die vergangenen 15 jahren, *Fachtagung fÄijr Untertagbau*, Luzern, Switzerland.