A probabilistic (max,+) approach for determining railway infrastructure capacity

Document status and date:
Published: 01/01/2000

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

A Probabilistic (max,+) Approach for Determining Railway Infrastructure Capacity

A.F. de Kort [*]   B. Heidergott [†]   H. Ayhan [‡]

Abstract

We consider the problem of determining the capacity of a planned railway infrastructure layout under uncertainties. In order to address the long–term nature of the problem, in which the exact (future) demand of service is unknown, we develop a "timetable"–free approach to avoid the specification of a particular timetable. We consider a generic infra–element that allows a concise representation of many different combinations of infrastructure, safety systems and traffic regimes, such as mixed double and single track lines (e.g. a double track line including a single tunnel tube), and train operations on partly overlapping routes at station yards. We translate the capacity assessment problem for such a generic infra–element into an optimization problem and provide a solution procedure. We illustrate our approach with a capacity assessment for the newly built high–speed railway line in the Netherlands.

Keywords: Transportation, Stochastic Processes, (Max,+) Semi–ring

[*]Holland Railconsult, Plan Development Department, P.O. BOX 2855, 3500 GW Utrecht, the Netherlands, email: afdekort@hr.nl

[†]EURANDOM, P.O. BOX 513, 5600 MB Eindhoven, the Netherlands, email: heidergott@eurandom.tue.nl .

[‡]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0205, U.S.A, email: hayhan@isye.gatech.edu

1 Introduction

Capacity assessment for railway infrastructure plays a key role in the design of layouts of infra–elements, such as tunnels, bridges, lines and complex railway yards, such as stations. We define capacity of a railway infra–element (denoted by N_{sys}) as the maximal number of train movements that can be executed on the particular infra–element in T time units (e.g. $T = 60$ minutes) with probability greater than or equal to p where p is a predefined reliability threshold. During railway capacity assessments it is examined whether the proposed infrastructure layout can handle the intended traffic load (denoted by N_{intended}) within a pre-defined level of quality. If capacity assessments yield negative results, that is, if $N_{\text{sys}} < N_{\text{intended}}$ with probability greater than p, alternative track layouts have to be considered. In this way, railway planners can adequately weigh involved construction costs against expected revenues (in terms of quality of service offered). In fact, changes in the European railway market show that this cost awareness argument for performing capacity studies is of growing importance; see [6].

The capacity of an infra–element is determined by (a) *structural aspects*, like the proposed track layout and the underlying safety system, (b) *timing aspects*, such as running times and dwelling times of trains as well as the amount of time required for boarding and alighting, and (c) the timetable (precise arrangement of train arrivals and departures in time and space). Structural aspects remain constant for many years ahead, and they are usually known during the planning phase. This is in contrast to the timing aspects as well as the timetable. As for the timing aspects, they are typically unknown during planning stages for two reasons: (1) the future service demand (like the expected numbers of passengers traveling by train, and the types of rolling stock to be deployed during operation) is unknown, and (2) external influences (like weather conditions and malfunctioning of material) may cause fluctuations in the (actual) process times. On the other hand, the timetable is usually altered at least once a year, and, hence, it is not practical to use a particular timetable to assess the infrastructure capacity.

The uncertainty about the timing aspects is dealt with by letting the correspond-

ing variables be stochastic, and in order to obtain a capacity measurement that is insensitive to a particular timetable, we apply "Wakob's razor": (1) we let trains arrive to the infra–element with interarrival time 0, and (2) we assume that initially there are infinitely many trains waiting in *reservoirs* to enter the infra-element. This saturation type of approach was introduced in [10], while its effectiveness in practice was demonstrated in De Kort et al. [7]. Apart from avoiding the specification of a particular timetable, Wakob's razor has the benefit that the measured capacity is independent of stochastic perturbations of the "outer" system, that is, the cause of any observed delay has to be the layout of the infra–element.

In order to model the infra–element according to steps (1) and (2) above, we determine N_{sys}, the maximal number of train movements that can be executed in T time units (e.g. $T = 60$ minutes) with probability p . Hence, we consider the capacity to be sufficient if and only if $N_{\mathrm{sys}} \geq N_{\mathrm{intended}}$. A mathematical framework most suitable for the analysis of transportation systems, such as train networks, is the (max,+) semi–ring (to be introduced presently), and we model dynamics of the train system (including the impact of the underlying safety system) by a set of stochastic difference equations that are linear in the (max, +) semi–ring. Heidergott & De Vries [5] provide a state–of–the–art overview of the applications of (max,+) techniques to the control of train networks.

The paper is organized as follows. Section 2 provides a brief introduction to (max,+) algebra. In Section 3, we introduce our generic building block and we derive the difference equations that describe the temporal dynamic behavior of the building block. In Section 4, we formulate the optimization problem and provide a solution procedure. In Section 5, we apply our approach to a real–life situation: a capacity assessment for the new high–speed railway line in the Netherlands. Section 6 concludes the paper.

2   The (Max,+) Algebra

In this section we introduce the so-called (max,+) algebra which will be the basic reference algebra throughout this paper. Let $\epsilon = -\infty$ and let us denote by $\mathbb{R}_\epsilon$ the set $\mathbb{R} \cup \{\epsilon\}$. For elements $a, b \in \mathbb{R}_\epsilon$ we define the operations $\oplus$ and $\otimes$ by

$$a \oplus b = \max(a, b) \quad \text{and} \quad a \otimes b = a + b \,,$$

where we adopt the convention that for all $a \in \mathbb{R}$ $\max(a, -\infty) = \max(-\infty, a) = a$ and $a + (-\infty) = -\infty + a = -\infty$. The set $\mathbb{R}_\epsilon$ together with the operations $\oplus$ and $\otimes$ is called the (max,+)–algebra and is denoted by $\mathbb{R}_{\max}$. In particular, ϵ is the neutral element for the operation $\oplus$ and absorbing for $\otimes$, that is, for all $a \in \mathbb{R}_\epsilon$ $a \otimes \epsilon = \epsilon$. The neutral element for $\otimes$ is $e = 0$.

The name "(max,+)–algebra" is only historically justified since $\mathbb{R}_{\max}$ is by no means an algebra in the classical sense. Structures like $\mathbb{R}_{\max}$ are referred to as *semi–rings*[1] in the literature. Moreover, $\mathbb{R}_\epsilon$ is *idempotent*, that is, for all $a \in \mathbb{R}_\epsilon$ $a \oplus a = a$. Idempotent semi–rings are called *dioids* in [2]. Hence, the correct name for $\mathbb{R}_{\max}$ would be "idempotent semi–ring" or "dioid" (which might explain why the name "(max,+)–algebra" is still predominant in the literature). The structure $\mathbb{R}_{\max}$ is richer than that of a dioid since $\otimes$ is commutative and has an inverse. However, in what follows we will work with matrices over $\mathbb{R}_{\max}$ and thereby lose, like in conventional algebra, commutativity and general invertability of the product.

Observe that the idempotency of $\oplus$ implies that $\oplus$ has no inverse (which explains why $\mathbb{R}_{\max}$ is not an algebra). Indeed, if $a \neq \epsilon$ had an inverse element, say b, w.r.t. $\oplus$, then $a \oplus b = \epsilon$ would imply $a \oplus a \oplus b = a \oplus \epsilon$. By idempotency, the left–hand side equals $a \oplus b$, whereas the right–hand side is equal to a. Hence, we have $a \oplus b = a$, which contradicts $a \oplus b = \epsilon$.

We extend the (max,+)–algebra operations to matrices in the following way. For

---

[1] A *semi–ring* is a set R endowed with two binary operations, $\oplus$ and $\otimes$, so that $\oplus$ is associative and commutative with zero–element ϵ, $\otimes$ is associative and has zero–element e, $\otimes$ distributes over $\oplus$ and ϵ is absorbing for $\otimes$.

$A, B \in \mathbb{R}_\epsilon^{m \times n}$, we define $A \oplus B$ as follows

$$(A \oplus B)_{ij} = A_{ij} \oplus B_{ij}, \qquad i = 1, \ldots, m; j = 1, \ldots, n.$$

For $A \in \mathbb{R}_\epsilon^{m \times l}$ and $B \in \mathbb{R}_\epsilon^{l \times n}$, we define $A \otimes B$ by

$$(A \otimes B)_{ij} = \bigoplus_{k=1}^{l} A_{ik} \otimes B_{kj} = \max_k (A_{ik} + B_{kj}) . \qquad (1)$$

3   The (max,+) model of a generic infrastructure element

This section provides the mathematical analysis of a generic infrastructure element, called 'generic building block'. Here the term 'generic' means that the infrastructure element contains all basic elements of railway infrastructure. Particularly, the generic building block allows a concise representation of many different combinations of infrastructure, safety systems and traffic regimes, such as mixed double and single track lines (e.g. a double track line including a single tunnel tube), and train operations on partly overlapping routes at station yards. The term 'building block' stems from the fact that a complex railway network can easily be represented by linking as many of these generic building blocks as desired.

The overall capacity of a complex network can be approximated by the capacities of its individual generic building blocks in the following way. First, the capacities of the 'building blocks' are determined in order to identify the potential bottleneck 'building block' and then the capacity of this building block is taken as an indicator for the capacity of the complex network. This is a common approach in the railway planning business, see e.g. Schwanhäußer [9].

This section is organized as follows. Section 3.1 introduces the generic building block. In Section 3.2 we determine conditions for the departure times of trains at specific locations of the generic building block. Based on these conditions, we derive in Section 3.3 a set of difference equations for the dynamic behavior of the generic building block that are linear in the (max,+) semi–ring. Finally,

in Section 3.4 we explain the stochasticity that is incorporated in this (max,+) model.

3.1 Layout of the generic building block

Throughout this section we consider the generic building block given in Figure 1. In fact, the layout of this building block resembles that of an elementary bottleneck in any complex railway infrastructure network. Our generic building block contains two parts of double track (i.e. separate tracks for opposite running directions), linked together by a single railway track. Each track lies between two adjacent nodes, numbered from 1 to 10. Altogether, the building block thus represents a network of seven railway tracks. These tracks are numbered $\tau_1 \ldots, \tau_7$ as shown in Figure 1.

In line with Wakob's razor, we let an infinite number of trains reside in *reservoirs* in front of the isolated building block. For our model, we require two reservoirs, one for each direction, which in turn are represented by nodes 1 and 6, respectively. Hence, trains are generated at either reservoir and move to the downstream node (either node 2 or 7) where they enter the building block. From these points, trains run on consecutive tracks according to the arrows. So, trains running from left to right will visit tracks τ_1, τ_3, τ_4 and τ_6 (in consecutive order), whereas trains from right to left will do so via tracks τ_7, τ_5, τ_4 and τ_2. Trains leave the building block at nodes 5 and 10, respectively. Therefore, we call nodes 5 and 10 *sinks*.

As a final remark, observe that τ_4 is used by all trains from both directions. Consequently, this track should only be occupied by one train at a time to prevent deadlocks and train collisions. Because of this, τ_4 acts as the bottleneck of our building block.

insert Figure 1 here

6

3.2 Departure conditions on the generic building block

In what follows, we provide conditions for departure times of a train from each node in the generic building block, as a function of departure times of previous trains. To this end, we define $x_i(k)$ as the departure time of the kth train from node i, for $i = 1, \ldots, 10$.

For the internal nodes of the system (i.e. all nodes except the reservoirs and the sinks), the departure time of the kth train depends on two conditions: (1) the train must be ready to leave the node, (2) the safety system should have authorized the train to enter the downstream track.

Clearly, condition (1) means that the departure time of the kth train from node i can not be earlier than its arrival time at node i which is equal to the sum of its departure time from the node upstream of i (according to its running direction), say $\pi(i)$, and the travel time on the track between $\pi(i)$ and i. Let $a_{i\,\pi(i)}(k)$ represent the required travel time between $\pi(i)$ and i for the kth train (including all relevant time components for running, dwelling, etc.). Then, the first condition becomes

$$x_i(k) \geq x_{\pi(i)}(k) + a_{i\,\pi(i)}(k) \text{ for } k \geq 1 , \tag{2}$$

where $a_{ij} \geq 0$ if $j = \pi(i)$ and $a_{ij} = -\infty$ otherwise for all $1 \leq i \leq 10$ and $k \in \mathbb{N}$.

Condition (2) is satisfied if the *rear end* of the $(k-1)^{\text{th}}$ train has reached a safe position in front of the kth train. For simplicity, assume that this 'track release' is achieved some time after the front of the $(k-1)^{\text{th}}$ train has departed from the node downstream of i, say $\sigma(i)$. Then, the 'track access condition' reads

$$x_i(k) \geq x_{\sigma(i)}(k-1) + r_{i\,\sigma(i)}(k-1) \text{ for } k \geq 1 , \tag{3}$$

where $r_{ij}(k)$ denotes the time elapsed before the track between i and j is released by the kth train (emphasizing that train length must not be neglected). We call this type of time variable the *release* time of the associated track. Furthermore, we have $r_{ij} \geq 0$ if $j = \sigma(i)$ and $r_{ij} = -\infty$ otherwise for all $1 \leq i \leq 10$ and $k \in \mathbb{N}$.

Remark: *Condition (3) represents a wide variety of safety regimes. More precisely, we can make the condition suitable for any specific safety principle by choosing the*

appropriate value for the release time. For example, setting $r_{ij} \geq 0$, no more than one train will occupy the track between nodes i and j at a time; whereas setting $r_{ij} < 0$ enables track occupation by several trains at the same time (e.g. with a moving block safety system.) Bailey [3] provides a detailed overview of principles and properties of safety and signaling systems in Europe.

At track τ_4, we actually have two track access conditions, namely one condition for consecutive trains running in the same direction (see above) and another condition for trains running in the opposite direction. In the latter case, we have to deal with the order of succession in which trains enter track τ_4. For our analysis, we assume that trains from either direction visit track τ_4 *alternately*, starting with a train running from left to right. This agrees with common railway practice where dispatchers will only deviate from this alternate passing of trains on single track parts of infrastructure to prevent large disturbances. As for our formal analysis, the fixed order assumption allows us to model the generic building block with the (max,+) semi–ring (see Heidergott [4]).

The above fixed order assumption implies that the kth train running from right to left will occupy τ_4 after the kth train in the opposite direction has done so. Accordingly, the kth departure time from node 8 depends on the kth departure from node 4 according to

$$x_8(k) \geq r_{84}(k) + x_4(k) \ , \ \ k = 1, 2, \dots \ ,$$

where $r_{84}(k)$ is again a release time, like $r_{34}(k)$. However, here $r_{84}(k)$ refers to the release time between *opposite* train movements, whereas $r_{34}(k)$ applies to trains running in the same direction. Conversely, the kth departure from node 3 depends on the $(k-1)^{\text{th}}$ departure from node 9. Hence,

$$x_3(k) \geq r_{39}(k-1) + x_9(k-1) \ , \ \ k = 2, 3, \dots \ \ .$$

The departure conditions for the reservoirs are

$$x_1(k) \geq z_1(k) + x_1(k-1) \ , \ \ k \geq 2$$

and

$$x_6(k) \geq z_6(k) + x_6(k-1) \ , \ \ k \geq 2 \ ,$$

8

where $z_i(k)$ denotes the time between generation of the $(k-1)^{\text{th}}$ train and the kth train (corresponding interarrival times at the entrance nodes 2 and 7 of the system). However, we assume that initially, an infinite number of trains are waiting at both sources. Moreover, we assume that there is a train ready to enter the system at any time. Accordingly, we may set $z_1(k) = z_6(k) = 0$ (deterministic). Consequently, we have

$$x_1(k) \geq x_1(k-1) \text{ and } x_6(k) \geq x_6(k-1) \ .$$

In addition, trains cannot leave either reservoir unless the respective downstream track is released. Thus, track access conditions given in (3) should also be fulfilled at the reservoirs.

Finally, trains leave the system immediately, once they have arrived at the sinks. Consequently, train departure times from the sinks only depend on the required travel time along the respective upstream tracks, that is, track access conditions can be omitted in this case. Thus, the departure conditions from the sinks are expressed as

$$x_5(k) \geq x_4(k) + a_{5\,4}(k) \ , \ \ k \geq 1$$
$$x_{10}(k) \geq x_9(k) + a_{10\,9}(k) , \ \ k \geq 1 \ .$$

3.3 Difference equations in (max,+) algebra

We assume that each train departure from the respective nodes takes place immediately after all conditions derived in Section 3.2 are satisfied. We then obtain

9

the following difference equations in (max,+) notation

$$
\begin{cases}
x_1(k) = x_1(k-1) \oplus (r_{1\,2}(k-1) \otimes x_2(k-1)) \\
x_2(k) = (a_{2\,1}(k) \otimes x_1(k)) \oplus (r_{23}(k-1) \otimes x_3(k-1)) \\
x_3(k) = (a_{3\,2}(k) \otimes x_2(k)) \oplus (r_{3\,4}(k-1) \otimes x_4(k-1)) \oplus (r_{39}(k-1) \otimes x_9(k-1)) \\
x_4(k) = (a_{4\,3}(k) \otimes x_3(k)) \oplus (r_{45}(k-1) \otimes x_5(k-1)) \\
x_5(k) = a_{5\,4}(k) \otimes x_4(k) \\
x_6(k) = x_6(k-1) \oplus (r_{6\,7}(k-1) \otimes x_7(k-1)) \\
x_7(k) = (a_{7\,6}(k) \otimes x_6(k)) \oplus (r_{78}(k-1) \otimes x_8(k-1)) \\
x_8(k) = (r_{8\,4}(k) \otimes x_4(k)) \oplus (a_{87}(k) \otimes x_7(k)) \oplus (r_{89}(k-1) \otimes x_9(k-1)) \\
x_9(k) = (a_{9\,8}(k) \otimes x_8(k)) \oplus (r_{9\,10}(k-1) \otimes x_{10}(k-1)) \\
x_{10}(k) = a_{10\,9}(k) \otimes x_9(k)
\end{cases}
$$

for $k = 2, 3, \ldots$ with the initial condition

$$
x(1) = \Big[x_1(1) \ldots x_{10}(1) \Big]^T = \Big[0 \; \epsilon \; \epsilon \; \epsilon \; 0 \; \epsilon \; \epsilon \; \epsilon \Big]^T \quad .
$$

We can simplify the above expressions for $x_3(k)$ and $x_8(k)$ as follows. We have assumed that, independent of the release times r_{ij}, the kth train running from left to right always passes track τ_4 before the kth train from right to left. Consequently, the $(k-1)^{\text{th}}$ departure time from node 9 dominates over the $(k-1)^{\text{th}}$ departure time from node 4 in the release of τ_4, that is

$$
r_{3\,4}(k-1) \otimes x_4(k-1) \le r_{3\,9}(k-1) \otimes x_9(k-1) \quad .
$$

Hence, we conclude that the term $r_{3\,4}(k-1) \otimes x_4(k-1)$ is superfluous in the expression for $x_3(k)$ and

$$
x_3(k) = (a_{3\,2}(k) \otimes x_2(k)) \oplus (r_{39}(k-1) \otimes x_9(k-1)) \quad .
$$

Likewise, we have

$$
r_{89}(k-1) \otimes x_9(k-1) \le r_{8\,4}(k) \otimes x_4(k) \quad .
$$

Consequently, we can leave out the term $r_{89}(k-1) \otimes x_9(k-1)$ in the expression for $x_8(k)$, which in turn results in

$$
x_8(k) = ((r_{84}(k) \otimes x_4(k)) \oplus (a_{87}(k) \otimes x_7(k)) \quad .
$$

10

In matrix form, the above recursion then reads

$$x(k) = A_0(k) \otimes x(k) \oplus A_1(k) \otimes x(k-1) \; , \tag{4}$$

with

$$A_0(k) = \begin{bmatrix} \epsilon & & & & & & & & & \\ a_{21}(k) & \epsilon & & & & & & & & \\ \epsilon & a_{32}(k) & \epsilon & & & & & & & \\ & \epsilon & a_{43}(k) & \epsilon & & & & & & \\ & & \epsilon & a_{54}(k) & \epsilon & & & & & \\ & & & \epsilon & \epsilon & \epsilon & & & & \\ & & & & \epsilon & a_{76}(k) & \epsilon & & & \\ & & r_{84}(k)\,\epsilon & & \epsilon & a_{87}(k) & \epsilon & & \\ & & & & & \epsilon & a_{98}(k) & \epsilon & \\ & & & & & & \epsilon & a_{109}(k)\,\epsilon \end{bmatrix}$$

and

$$A_1(k+1) = \begin{bmatrix} e\; r_{12}(k) & \epsilon & & & & & & & \\ & \epsilon & r_{23}(k)\,\epsilon & & & & & & \\ & & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & r_{39}(k) & \epsilon \\ & & \epsilon r_{45}(k)\,\epsilon & & & & & & \\ & & & \epsilon & \epsilon & \epsilon & & & \\ & & & e\; r_{67}(k) & \epsilon & & & & \\ & & & & \epsilon & r_{78}(k) & \epsilon & & \\ & & & & & \epsilon & \epsilon & \epsilon & \\ & & & & & & \epsilon & r_{910}(k) & \\ & & & & & & & \epsilon & \end{bmatrix} \; .$$

In recursion (4), the term $x(k)$ occurs on both sides of the equation. Using basic results from the theory of the (max,+) semi–ring, we transform (4) into a recursion of type $x(k+1) = A(k) \otimes x(k)$. To this end, we set

$$b(k) = A_1(k) \otimes x(k-1)$$

then (4) reduces to

$$x(k) = A_0(k) \otimes x(k) \oplus b(k) \; . \tag{5}$$

11

For fixed k, the above equation can be written as $x = A_0 \otimes x \oplus b$. It is well–known that $x = A_0^\star \otimes b$ solves this equation, see Theorem 3.17 in [2], where

$$A_0^\star = \bigoplus_{i=0}^{\infty} A_0^i$$

and

$$A_0^i = \underbrace{A_0 \otimes \ldots \otimes A_0}_{i \text{ times}}$$

denotes the ith power of A_0; in particular, $A_0^0 = E$ (a matrix with its diagonal entries equal to e and the off–diagonal entries equal to ϵ, that is, E is the unit element of matrix multiplication in the (max,+) semi–ring). If A_0 is a lower triangular matrix, which is true in our case, then a finite integer n exists such that

$$A_0^\star = \bigoplus_{i=0}^{n} A_0^i \quad , \tag{6}$$

and it is easily checked that in our case $n = 6$ (by determining the longest path in the precedence graph associated with $A_0(k)$, i.e. containing the largest number of edges, see Figure 2).

insert Figure 2 here

Thus, (5) reads as

$$x(k) = A_0^\star(k) \otimes b(k) ,$$

or, more explicitly,

$$x(k) = A_0^\star(k) \otimes A_1(k) \otimes x(k-1) . \tag{7}$$

As a final step, we set

$$A(k) = A_0^\star(k+1) \otimes A_1(k+1) ,$$

and obtain

$$x(k+1) = A(k) \otimes x(k) \tag{8}$$

as our (max,+) model for the dynamic behavior of the generic building block. Elements of the $A(k)$ matrix are given in the appendix.

3.4 The stochastic properties of the model

As explained in Section 1, we assume that (a) the 'internal' time variables, like travel times, are stochastically independent, (b) the travel times on a particular track are identically distributed, and (c) all random variables are defined on a common probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Note that, even though the travel and release times are stochastically independent, the entries of $A(k)$ are in general dependent.

Remark: Observe that (max,+) models like in (8) do not distinguish between classes of items (like trains in our case). This implies that all time variables contained in $A(k)$ apply to all respective trains. However, in practice, each train may in fact be of different type, that is requiring different travel times and release times. Still, we can incorporate class–dependent travel times and release times into our (max,+)–linear model by constructing distributions that are weighted mixtures of the distributions corresponding to the present distinct train types.

4   Formulation of the capacity assessment problem

Our objective is to find the maximum number of train movements the system can handle within a predefined period of time, denoted by T. In railway practice, T is usually set equal to 60 minutes as the intended operation involves the same arrival and departure times during every operating hour.

Observe that $x(k)$ contains the kth departure times from all nodes. In particular, $x_5(k)$ refers to the departure time of the kth train running from left to right, and $x_{10}(k)$ refers to the departure time of the kth train running from right to left at the respective sinks. Thus, at time $\max(x_5(k), x_{10}(k))$, $2 \cdot k$ trains have left the system. This is tantamount to saying that the matrix $A(k)$ describes the transitions of the system for the kth train running from right to left and also for the kth train running from left to right, simultaneously. Hence, for the rest of the paper we will consider the kth *pair* of trains.

The fixed train order assumption for track τ_4 implies that $x_{10}(k) \geq x_5(k)$ (since

the kth train from left to right will be the first one to occupy τ_4). In other words, at time $x_{10}(k)$ the kth pair of trains has left the system. Consequently, for any time period T, the number of train movements handled by the system is determined by the last train from right to left which departed from node 10 (just) before time T.

Since the entries of $A(k)$ in (8) are stochastic, $\{x(k)\}$ becomes a random sequence. As a consequence, the state dynamics which we observe are those along a particular sample path $\omega \in \Omega$, denoted by $\{x(k;\omega)\}$. Moreover, since the number of train movements is directly related to departure times at node 10, the number of trains that can be processed in a time span T, i.e. the capacity, also becomes random. In order to evaluate the system's capacity, we introduce a pre-specified probability value $p \in (0,1)$, which can be regarded as a measure for reliability of train operations, and we evaluate the capacity N_{sys} of the generic building block by solving the following optimization problem

$$\max \ k$$

$$\text{s.t.}$$

$$P\{x_{10}(k) \leq T\} \geq p$$

The optimal solution k^* of the above optimization problem is equal to N_{sys} and it can be computed as

$$k^* = \inf\{\, k \mid P\{x_{10}(k) \leq T\} < p \,\} - 1.$$

Hence, one needs to compute $P\{x_{10}(k) \leq T\}$. Let κ_k be the set of travel times and release times that exist in the expression of $x_{10}(k)$. Then

$$P\{x_{10}(k) \leq T\} = \int_{S_k} F(ds_1, \cdots, ds_{\kappa_k})$$

where F is the joint distribution of the random variables in the set κ_k and $S_k \subset \mathbb{R}^{|\kappa_k|}$ is the image of the event $\{x_k(10) \leq T\}$. Thus, computation of $P\{x_{10}(k) \leq T\}$ requires the evaluation of a multi-dimensional integral. Exploiting the property that S_k is a convex polytope, Ayhan and Wortman [1] convert this cumbersome integral to a more straightforward optimization problem. Since

14

their technique also provides bounds for $P\{x_{10}(k) \leq T\}$, in some cases the feasibility of the constraint in the above optimization problem can be checked without actually computing $P\{x_{10}(k) \leq T\}$. The interested reader could refer to [1] for the details of this algorithm. Note that for each value of p we need to solve this optimization problem only once in order to determine $k^\star$ ($= N_{\text{sys}}$).

5 Application: capacity assessment for HSL South

5.1 Problem statement

As an application, we consider the capacity assessment problem for the Netherlands portion of HSL South, the new high–speed railway line which is being built in the Netherlands and will connect Amsterdam with Brussels and Paris via Rotterdam and vice versa. The Dutch part of the line will be operational in 2005 and from then on, the expected traffic load will be 8 trains per hour (in both directions), increasing up to 16 trains per hour by 2015.

The Dutch part of the line includes three special tunnels, each with separated tunnel tubes for both running directions. Figure 3 shows the line schematically, with the tunnels under consideration represented by boxes. From North to South (left to right), the tunnels are called "Groene Hart" tunnel, "Oude Maas" tunnel and "Dordtsche Kil" tunnel, respectively. The nodes are fictitious points distinguishing the track parts for which we provide time values later on in this section. The corresponding track distances are depicted in kilometers.

insert Figure 3 here

Due to the absence of emergency exits, at each tunnel only one tube may be used at a time, in order to guarantee passengers a safe escape route (to the opposite tunnel tube) in case of an emergency. In other words, if a train occupies one of the tunnel tubes, the opposite tube is immediately blocked for other trains. Consequently, even though HSL South is an entirely double track line, all three tunnels

15

behave like single track parts and as such, the capacity of the entire line may be restricted too much to achieve the expected traffic loads as indicated above. An alternative would then be to build an extra tube (e.g. a service channel). However, drilling a railway–tunnel tube costs about \$50,000 per running meter (including the installation of all equipment), which explains the importance of carefully examining the capacity of the line.

5.2 Capacity assessment procedure

In order to determine whether the two–tube tunnel layout for HSL South offers sufficient capacity to process the expected traffic load, we proceed as follows. Since each tunnel behaves as a single track part, we may split up the line into three tunnel elements, each of which can be considered as a generic building block. Following the line of argument in Section 3.3, we obtain for each of the three tunnel elements a (max,+) model $x^{(i)}(k+1) = A^{(i)}(k) \otimes x^{(i)}(k)$, $i = 1, \ldots, 3$. As a first step, we evaluate the capacity of each tunnel element, denoted by $N_{\mathrm{sys}}^{(i)}$, for a fixed value of $p \in (0, 1)$. Subsequently, we take

$$N_{\mathrm{sys}} = \min \{ N_{\mathrm{sys}}^{(i)} : 1 \leq i \leq 3 \}$$

as an estimator for the capacity of the entire HSL South. Indeed, using the minimum of these capacity indicators as an indication for the overall capacity is a commonly adopted approach in railway planning; (see for example [8], [9] for Schwanhäußer's funnel argument and also [10]).

Finally, we check if this capacity is sufficient to process the expected traffic load N_{intended}. Recall that $8 \leq N_{\mathrm{intended}} \leq 16$. Thus, the capacity is sufficient if

$$N_{\mathrm{sys}} \geq N_{\mathrm{intended}} \geq 8 \, .$$

5.3 Values for all time variables

According to the above assessment procedure, we have to evaluate $A^{(i)}(k)$ for each tunnel element. To this end, we have to define the type of distributions that

apply to every relevant time variable. More specifically, the following assumptions are made.

- All non-negligible travel times consist of both a deterministic and a stochastic component. Hence, $a_{ij}(k) = a_{ij} + \delta_{ij}(k)$, with a_{ij} denoting the deterministic travel time and $\delta_{ij}(k)$ denoting the delay for the kth train $(k = 1, 2, \ldots)$, the distribution of which is given by

$$\delta_{ij}(k) \sim \text{Unif}([0, 3]) \text{ with probability } 0.95 \text{ and}$$
$$\delta_{ij}(k) \sim \text{Exp}(\lambda) \text{ with probability } 0.05 \text{ ,}$$

where $1/\lambda = 3 + \frac{a_{ij}}{2}$, while a_{ij} is expressed in minutes. So, for example, if a train with a deterministic travel time of 30 minutes is delayed, then the expected (large) delay is 15 minutes. Although estimation of delay distributions is still an ongoing research topic, historical data indicates that the majority of trains on the Netherlands railways are less than three minutes late on every trip, whereas large deviations are proportional to the scheduled travel times.

- Once a train has left either reservoir, it immediately enters the generic building block. Consequently, we have $a_{21}(k) = a_{76}(k) \equiv 0$, for $k \geq 1$.

- All release times $r_{ij}(k)$ are deterministic and fixed for all $k \geq 1$.

- The single track part (τ_4) is 'released' for the next train immediately after the front of the previous train has reached the exit node at either side of the tunnel (i.e. node 4 or 8). Accordingly, we set $r_{39} = r_{84} = 0$.

Tables 2– 3 contain values for all relevant time variables that apply to HSL South, from North to South and vice versa, respectively. All values are expressed in minutes and determined for one specific type of rolling stock, namely two coupled TGV trains, denoted TGV2, the main characteristics of which are given in Table 1. Among all types of rolling stock that will actually run on this line, this particular type requires the smallest travel times. In this respect, we thus obtain a best–case indication of the capacity. Furthermore, we assume that the

17

line is equipped with a moving block safety system, while allowing a maximum speed of 300 km/h (where possible). That is why negative release times appear in Tables 2– 3.

insert Table 1 here

Finally, all trains are supposed to dwell two minutes at Rotterdam Central Station for an intermediate stop, which is located between the Groene Hart tunnel and Oude Maas tunnel. We refer to [6] for more technical details behind these values. One should be aware that, in Tables 2 and 3, the indices i and j refer to the node defined in Figure 3, rather than to the nodes belonging to a generic building block. For example, $a_{43}(k)$ in Table 2 corresponds to the travel time on track τ_3 of the generic building block for the Oude Maas tunnel, whereas it applies to the travel time on τ_6 with respect to the generic building block for the Groene Hart tunnel.

insert Table 2 here

insert Table 3 here

5.4 Numerical results

We present our numerical results in Table 4. For each reliability measure p, we provide the capacity of each tunnel as well as the capacity of the whole line. The amount of time it takes to compute $N_{sys}^{(i)}$ for a given value of p and i ($i = 1, 2, 3$) takes only a few minutes (on a 133 MHZ PC). Based on the outcomes for the whole line (third column), it is immediately clear that the expected traffic load lies significantly higher than the capacity of the line for all values of p. That is, $N_{\text{sys}} \leq 4$, whereas $N_{\text{intended}} \geq 8$. The Oude Maas tunnel acts as bottleneck for HSL South. Moreover, the punctuality objectives of the Netherlands Railways require a reliability measure $p \geq 0.92$, which in turn implies a capacity of at most three trains per hour per direction for the entire line according to Table 4. Therefore, we conclude that the original layout with two–tube tunnels while per-

mitting only one train at a time in each tunnel does not provide sufficient capacity and hence either extra tunnel tubes are needed or the strategy that prevents the simultaneous occupation of both tunnel tubes has to be abandoned (which of course requires additional measures to guarantee that passengers can still safely escape to the opposite tunnel tube in case of an emergency). As a result, studies on both options are now being performed to increase capacity of HSL South.

insert Table 4 here

6 Conclusions

In this paper we illustrate that the (max,+) semi–ring is a suitable mathematical framework for analyzing the impact of infrastructure constraints, the underlying safety system and special traffic regimes, on the capacity of a given railway track layout. Moreover, the generic building block concept allows a similar assessment of infrastructure elements, lines and complex junctions since it focuses on the potential bottlenecks of the layout under consideration. By adopting the principle of Wakob's razor (isolating the infrastructure while letting an infinite number of trains reside at the boundaries of the building block), any observed delay can be attributed to the track layout. In order to account for the stochastic nature of all time variables, we consider the maximum number of trains per direction that can be processed in T time units with a predefined probability p as our performance measure.

An assessment for HSL South, concerning the capacity offered by the proposed two–tube tunnel layout, shows that this probabilistic (max,+) approach can be effectively adopted to obtain insight into the long–term perspectives of railway infrastructure. Since the computational efforts are very small, it may serve as a useful substitute to simulation in similar cases.

Acknowledgements

References

[1] Ayhan, H., & M.A. Wortman, Job Flow Control in Assembly Operations. *IEEE Trans. on Automatic Control*, **44**, 864-868, 1999.

[2] Baccelli, F., G. Cohen, G.J. Olsder & J.P. Quadrat, *Synchronization and Linearity: An Algebra for Discrete Event Systems.* John Wiley & Sons, New York, 1992.

[3] Bailey, C. (general editor), *European Railway Signaling.* Institution of Railway Signal Engineers, A&C Black, London, 1995.

[4] Heidergott, B. A Characterization for (max,+)−linear Queueing Systems. *QUESTA*, **35**, 237-262, 2000.

[5] Heidergott, B., & R. de Vries, Towards a Control Theory for Transportation Networks. *DEDS*, (to appear).

[6] Heidergott, B., & A.F. de Kort, Dimensioning Tunnels of a Railway Line: How Many Tubes are Needed? (submitted for publication).

[7] Kort, A.F. de, B. Heidergott, R.J. van Egmond, & G Hooghiemstra, *Train Movement Analysis at Railway Stations: Procedures & Evaluation of Wakob's Approach.* TRAIL Studies in Transportation Science, **S99/1** Delft University Press/The Netherlands TRAIL Research School, Delft, 1999.

[8] Schwanhäußer W., Wahrscheinlichkeitsmodell zur Ermittlung von Wartezeiten bei Zugfolgefällen. *Archiv für Eisenbahntechnik* (AET), **36**, pp. 15–22, 1981 (in German).

[9] Schwanhäußer W., Leistungsfähigkeit und Kapazität. in: *Kapazitätreserven der Verkehrsträger* DVWG – Schrift no. **B 178**, pp. 3–17, 1994 (in German).

[10] Wakob, H., *Ableitung eines generellen Wartemodells zur Ermittlung der planmäßigen Wartezeiten im Eisenbahnbetrieb unter besonderer Berücksichtigung der Aspekte Leistungsfähigkeit und Anlagebelastung.* Veröffentlichungen des verkehrwissenschaftlichen Institutes der RWTH Aachen, **36**, 1985 (in German).

Appendix

Based on $A(k) = A_0^*(k+1) \otimes A_1(k+1)$ (see Section 3.3), we get

$$A(k) = [A_{.1}(k) \ \cdots \ A_{.10}(k)] \ ,$$

where

$$A_{.1}(k) = \begin{bmatrix} e \\ a_{21}(k+1) \\ a_{32}(k+1) \otimes a_{21}(k+1) \\ a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \\ a_{54}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \\ \epsilon \\ \epsilon \\ r_{84}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \\ a_{98}(k+1) \otimes r_{84}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \\ a_{109}(k+1) \otimes a_{98}(k+1) \otimes r_{84}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \end{bmatrix} \ ,$$

$$A_{.2}(k) = \begin{bmatrix} r_{12}(k) \\ a_{21}(k+1) \otimes r_{12}(k) \\ a_{32}(k+1) \otimes a_{21}(k+1) \otimes r_{12}(k) \\ a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \otimes r_{12}(k) \\ a_{54}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \otimes r_{12}(k) \\ \epsilon \\ \epsilon \\ r_{84}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \otimes r_{12}(k) \\ a_{98}(k+1) \otimes r_{84}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \otimes r_{12}(k) \\ a_{109}(k+1) \otimes a_{98}(k+1) \otimes r_{84}(k+1) \otimes a_{43}(k+1) \otimes a_{32}(k+1) \otimes a_{21}(k+1) \otimes r_{12}(k) \end{bmatrix} \ ,$$

$$A_{.\,3}(k) = \begin{bmatrix} \epsilon \\ r_{2\,3}(k) \\ a_{3\,2}(k+1) \otimes r_{2\,3}(k) \\ a_{4\,3}(k+1) \otimes a_{3\,2}(k+1) \otimes r_{2\,3}(k) \\ a_{5\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes a_{3\,2}(k+1) \otimes r_{2\,3}(k) \\ \epsilon \\ \epsilon \\ r_{8\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes a_{3\,2}(k+1) \otimes r_{2\,3}(k) \\ a_{9\,8}(k+1) \otimes r_{8\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes a_{3\,2}(k+1) \otimes r_{2\,3}(k) \\ a_{10\,9}(k+1) \otimes a_{9\,8}(k+1) \otimes r_{8\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes a_{3\,2}(k+1) \otimes r_{2\,3}(k) \end{bmatrix} \quad ,$$

$$A_{.\,4}(k) = \begin{bmatrix} \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon \end{bmatrix}^{T} \quad ,$$

$$A_{.\,5}(k) = \begin{bmatrix} \epsilon \\ \epsilon \\ \epsilon \\ r_{4\,5}(k) \\ a_{5\,4}(k+1) \otimes r_{4\,5}(k) \\ \epsilon \\ \epsilon \\ r_{8\,4}(k+1) \otimes r_{4\,5}(k) \\ a_{9\,8}(k+1) \otimes r_{8\,4}(k+1) \otimes r_{4\,5}(k) \\ a_{10\,9}(k+1) \otimes a_{9\,8}(k+1) \otimes r_{8\,4}(k+1) \otimes r_{4\,5}(k) \end{bmatrix} \quad ,$$

$$A_{.\,6}(k) = \begin{bmatrix} \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ e \\ a_{7\,6}(k+1) \\ a_{8\,7}(k+1) \otimes a_{7\,6}(k+1) \\ a_{9\,8}(k+1) \otimes a_{8\,7}(k+1) \otimes a_{7\,6}(k+1) \\ a_{10\,9}(k+1) \otimes a_{9\,8}(k+1) \otimes a_{8\,7}(k+1) \otimes a_{7\,6}(k+1) \end{bmatrix},$$

$$A_{.\,7}(k) = \begin{bmatrix} \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ r_{6\,7}(k) \\ a_{7\,6}(k+1) \otimes r_{6\,7}(k) \\ a_{8\,7}(k+1) \otimes a_{7\,6}(k+1) \otimes r_{6\,7}(k) \\ a_{9\,8}(k+1) \otimes a_{8\,7}(k+1) \otimes a_{7\,6}(k+1) \otimes r_{6\,7}(k) \\ a_{10\,9}(k+1) \otimes a_{9\,8}(k+1) \otimes a_{8\,7}(k+1) \otimes a_{7\,6}(k+1) \otimes r_{6\,7}(k) \end{bmatrix},$$
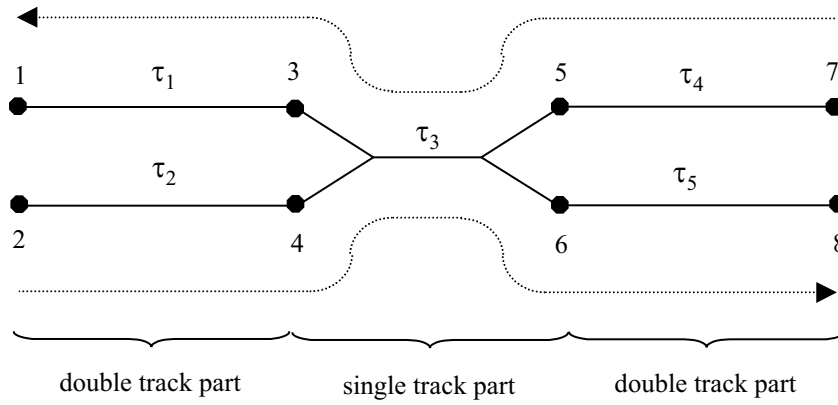
$$A_{.\,8}(k) = \begin{bmatrix} \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ r_{7\,8}(k) \\ a_{8\,7}(k+1) \otimes r_{7\,8}(k) \\ a_{9\,8}(k+1) \otimes a_{8\,7}(k+1) \otimes r_{7\,8}(k) \\ a_{10\,9}(k+1) \otimes a_{9\,8}(k+1) \otimes a_{8\,7}(k+1) \otimes r_{7\,8}(k) \end{bmatrix},$$

$$A_{.\,9}(k) = \begin{bmatrix} \epsilon \\ \epsilon \\ r_{3\,9}(k) \\ a_{4\,3}(k+1) \otimes r_{3\,9}(k) \\ a_{5\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes r_{3\,9}(k) \\ \epsilon \\ \epsilon \\ r_{8\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes r_{3\,9}(k) \\ a_{9\,8}(k+1) \otimes r_{8\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes r_{3\,9}(k) \\ a_{10\,9}(k+1) \otimes a_{9\,8}(k+1) \otimes r_{8\,4}(k+1) \otimes a_{4\,3}(k+1) \otimes r_{3\,9}(k) \end{bmatrix}$$

and

$$A_{.\,10}(k) = \begin{bmatrix} \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ \epsilon \\ r_{9\,10}(k) \\ a_{10\,9}(k+1) \otimes r_{9\,10}(k) \end{bmatrix}.$$
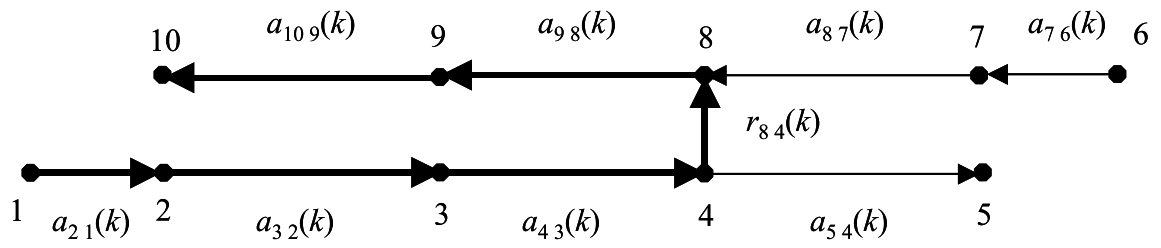
Figure 1: Layout of the generic building block

Figure 2: Precedence graph of $A_0(k)$ (with its longest path shown bold)
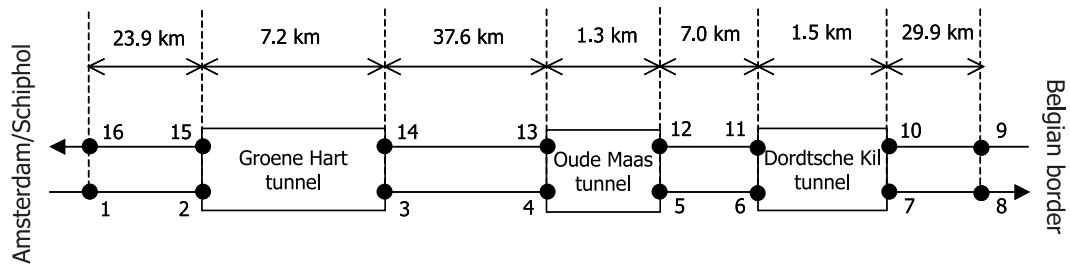
Figure 3: Distances along HSL South and its three tunnels

Table 1: Main characteristics of TGV2 rolling stock

length	$400\ m$
maximum speed	$300\ km/h$
emergency break	$\begin{cases} 1.0\ m/s^2, \text{if speed } > 215\ km/h \\ 1.3\ m/s^2, \text{if speed } \leq 215\ km/h \end{cases}$

Table 2: Relevant time values for HSL South from North to South [min] (see Figure 3)

travel time $a_{ij}(k)$	deterministic part a_{ij}	delay $\delta_{ij}(k)$		release time $r_{ij}(k)$	deterministic part r_{ij}
		with $p = 0.95$	with $p = 0.05$		
$a_{21}(k)$	8.07	Unif([0,3])	Exp(0.14)	$r_{12}(k)$	-7.30
$a_{32}(k)$	2.13	Unif([0,3])	Exp(0.25)	$r_{23}(k)$	0.53
$a_{43}(k)$	19.60	Unif([0,3])	Exp(0.08)	$r_{34}(k)$	-19.02
$a_{54}(k)$	0.87	Unif([0,3])	Exp(0.29)	$r_{45}(k)$	0.30
$a_{65}(k)$	1.05	Unif([0,3])	Exp(0.28)	$r_{56}(k)$	-0.32
$a_{76}(k)$	0.98	Unif([0,3])	Exp(0.29)	$r_{67}(k)$	0.47
$a_{87}(k)$	5.72	Unif([0,3])	Exp(0.17)	$r_{78}(k)$	-5.48

Table 3: Relevant time values for HSL South from South to North [min] (see Figure 3)

travel time $a_{ij}(k)$	deterministic part a_{ij}	delay $\delta_{ij}(k)$		release time $r_{ij}(k)$	deterministic part r_{ij}
		with $p = 0.95$	with $p = 0.05$		
$a_{10\,9}(k)$	5.98	Unif($[0,3]$)	Exp(0.17)	$r_{9\,10}(k)$	-5.22
$a_{11\,10}(k)$	0.98	Unif($[0,3]$)	Exp(0.29)	$r_{10\,11}(k)$	0.47
$a_{12\,11}(k)$	0.93	Unif($[0,3]$)	Exp(0.29)	$r_{11\,12}(k)$	-0.27
$a_{13\,12}(k)$	0.90	Unif($[0,3]$)	Exp(0.29)	$r_{12\,13}(k)$	0.38
$a_{14\,13}(k)$	19.05	Unif($[0,3]$)	Exp(0.08)	$r_{13\,14}(k)$	-18.28
$a_{15\,14}(k)$	2.15	Unif($[0,3]$)	Exp(0.25)	$r_{14\,15}(k)$	0.53
$a_{16\,15}(k)$	8.32	Unif($[0,3]$)	Exp(0.14)	$r_{15\,16}(k)$	-7.55

Table 4: Capacity for various values of reliability measure p ($T = 60$ min.)

p	$N_{\text{sys}}^{(i)}$			$N_{\text{sys}} = \min\{N_{\text{sys}}^{(i)} : 1 \le i \le 3\}$
	$i = 1$	$i = 2$	$i = 3$	
	Groene Hart	Oude Maas	Dordtsche Kil	
0.70	5	4	10	4
0.75	5	4	10	4
0.80	5	4	10	4
0.85	4	4	10	4
0.90	4	3	10	3
0.95	4	2	10	2
0.99	2	1	8	1