STATISTICAL MODELLING OF DELAYS IN A RAIL FREIGHT TRANSPORTATION NETWORK

János Barta

Andrea Emilio Rizzoli

Matteo Salani

Luca Maria Gambardella

Istituto Dalle Molle di Studi

sull'Intelligenza Artificiale (IDSIA - USI/SUPSI)

Galleria 2

Manno-Lugano, 6928, SWITZERLAND

ABSTRACT

This study analyzes the transportation network of a major rail freight operator in order to obtain a model of delay propagation of trains connecting intermodal terminals. Operational management of a rail freight operator needs to take into account deviations due to unexpected events such as unplanned maintenance, strikes, railroad works, traffic congestion. The dispatcher makes train assignment decisions based on a number of performance indicators and also on the expectancy that a given train, currently delayed, could recover or limit the amount of delay in the future. We have developed a Markov-chain based model in order to evaluate the evolution of train delays as a train visits successive terminals. Our model is based on the examination of a large set of historical data and we show how we can classify different terminals according to their ability either to absorb or to amplify delays.

1 INTRODUCTION

Rail freight transport has been constantly declining in Europe, despite all the efforts apparently made by political institutions such as the European Community to favour a modal switch from road to rail. In 2000 railways were transporting 19.7% of all freight, and road 73.7% while in 2009 the railways percentage went down to 16.5% and road went up to 77.5% (European Commission Eurostat 2012)

While there are a number of political causes in such an unexpected and undesirable decline, ranging from insufficient and badly aimed subsidies to strong lobbying by road transport operators, we wanted to shift our attention to the inefficiencies in rail freight transport, which can be divided into structural, bureaucratic, management and operational inefficiencies. In particular, given our scientific and research background, we decided to tackle the problem of the design and implementation of algorithms to improve the management and operational performance of rail-based freight transport.

In order to do so, we developed a joint project with Hupac Intermodal SA, a Swiss based company, which is one of the leading operators for combined transport in Europe. Hupac operates a network of 100 trains each day between Europe's main economic areas. In 2011, the traffic volume was around 724,000 road consignments (Hupac 2012). The project, codenamed WHATIF, had the objective of developing methodologies, algorithms and tools to assist a rail freight company, such as Hupac, in the planning and the management of its operations.

Combined transport operators offer their customers access to a network of intermodal terminals, where the customers can deliver and pick up transport units which are then transported by the combined transport operator over a rail network. The advantage of combined transport is that trains are used to cover the longest parts of the transport, while trucks are used only for the pick-up from origin and the deliver to

final destination, covering a small part of the overall transport. This approach combines the flexibility of trucks with the performance and efficiency of trains.

Operating an intermodal transport company requires solving a number of complex management problems, over different time horizons: from forecasting demand and structuring the network of connections among terminals, to dimensioning the size and capacity of intermodal terminals; from the operational management of terminals to planning the timetables of trains serving the network of terminals, down to dispatching trains from terminals according to the current state of the rail network. Such problems have attracted the attention of researchers for many years. Even limiting a search to papers appeared in previous Winter Simulation Conferences, the scope and breadth of research is considerable. As early as in 1969 the problem of designing and implementing a simulation model of the rail network of the United States was introduced (Minger and Cetinich 1969). Simulation of the rail network has always been an intensely investigated topic, as shown by later works (Leilich 1998; Lewellen and Tumay 1998; Huang, Seck, and Verbraeck 2010). Simulation has more recently also focused on the description of what happens within the rail terminals, which are the main hubs of interconnection between various modes of transport, from sea, to road, to rail (Rizzoli, Fornara, and Gambardella 2002; Benna and Gronalt 2008; Lin and Cheng 2009).

Simulation plays therefore a major role in the exploration of the possible future evolution of a system in order to devise the best strategies at various levels, from strategic down to planning and operational management. For this reason we focus our attention to the use of simulation techniques to support train dispatching at the terminal level. Given the train timetable, the role of the dispatcher is to assign incoming trains in a terminal to outgoing trains, respecting the original plan. Unfortunately delays, disruptions, unexpected maintenance interventions, strikes, can always alter the normal state of circulation of trains on a complex network thus making the planned assignments impossible. The dispatcher can therefore come up with contingency solutions, that require choosing different trains than those originally planned to serve some soon to depart connections. Such reassignments require the evaluation of a number of performance indicators and also an estimate of the possible evolution of the delay that has been accumulated. The ability to forecast the propagation of delays is essential for operational decision making, as the manager needs to know what could be the possible evolution of a delay which has been accumulated by an incoming train in order to make the best decision in order to re-route such train on a next connection.

Propagating delays over a rail freight network might seem like a trivial exercise, as it might be sufficient to add up the travel time on the different legs of the rail transport, and to compute the amount of time required to shunt, unload and load the train at each terminal stop. Actually, such times are very rarely deterministically known, and sometimes, especially for terminals which are not directly operated by the intermodal transport company (which are the majority for a company such as Hupac) are not available at all. For this reason we have developed a statistical model to estimate the propagation of delays on a rail network.

In the remainder of this paper in Section 2 we first describe this problem in detail, then we introduce the mathematical model of the problem in Section 3, and in Section 4 we analyze the results of the performance of our algorithm in the case of the Hupac network.

2 THE DELAY PROPAGATION PROBLEM

Combined transport operators offer their customers various transport services from terminal to terminal. Such transports are structured on a timetable that specifies the *closing time*, the time at which the departing train stops accepting new loads, and the *ready time*, the time when loading units such as semi-trailers and containers are ready to be picked up at destination.

Various authors have been studying rail networks since the 60s in the past century, and notably (Assad 1979) presented a first classification of the various Operations Research problems which could be identified in the rail freight context. Among the various problems, which range from finding the best location for terminals down to yard management operations within each terminal, particular attention has raised the problem of the definition of timetables (Caprara, Fischetti, and Toth 2002).

In our model, given a timetable, the set of trains in it is given by the *circulation plan*, which defines the *rotations* of *train compositions* over the network. A train composition is a specific set of wagons with given characteristics (length, height, maximum load, etc.) to accommodate specific loading units. We define the *relation* as a service between two terminals. The above mentioned *rotation* can now be defined as the sequence of connections between relations performed by the same train composition. A rotation is therefore also identified by departure and arrival times at the terminals. Once the train composition comes back to the origin terminal, it will start again following the same rotation pattern. The same composition can assume different train numbers according to the time of departure. For instance, in the Hupac network, the relation between the BUS2 terminal (Busto Arsizio, Italy) and the ROTT terminal (Rotterdam, The Netherlands) is a simple and direct link, and a rotation is a sequence of return trips between the two terminals performed a number of times over a specified period. If we analyze a typical week, the ROTT-BUS2 link is served by three rotations, the first two rotations use three train compositions, while the last one needs two trains. In other words, a rotation is a composition that travels from a terminal T_i to a terminal T_j through a number of intermediate terminals. The simplest case is a sequence $\{T_i, T_k, T_i\}$, but there are more complex situations, with trains being split at specific terminals and routed to different destinations, and then being recomposed on their way back.

A transport can be therefore seen as a sequence of activities which are: loading of units on rolling stock, train formation (traction of loading stock with a locomotor), train travel over the rail network, arrival to terminal and train shunting, unloading of units. Along any one of these activity problems might occur; for instance, a wagon might be discarded because of a breakdown, the terminal might be congested, and loading/unloading operations can be delayed because of lack of resources, trains are delayed on the rail network because of congestion or works on the line.

Any type of delay, independently on its cause, may disrupt the circulation plan: an incoming composition which was due to depart for the next link does not arrive on time, and the dispatcher must decide whether to wait for the late composition or reassign another available composition to serve the outgoing link.

Many authors have been investigating the problem of delays in rail networks. Higgins and Kozan (1998) adopted an analytic approach to compute the propagation of delays, while Özekici and Sengör (1994) used Markov chain techniques to compute cascade delays for urban passenger trains. More recent efforts, such as the one by (Acuna-Agost, Michelon, Feillet, and Gueye 2011) have focused on the use of the information produced by delay propagation techniques for an efficient rescheduling of operations. (D'Ariano, Pacciarelli, and Pranzo 2007; D'Ariano and Pranzo 2009) considered optimal train rescheduling in the case of disruptions. In the context of the WHATIF project we are also trying to provide operational tools to assist dispatchers in the minimization of the adverse effects of unexpected delays.

We have observed that dispatchers, in case of delays, choose the assignment of an incoming train to an outgoing composition based on an estimate of the ability of a given rotation to "absorb" the accumulated delay. This estimate is based on the personal experience of the dispatcher and it depends on a number of factors, such as the sequence of terminals to be visited and the route followed by the train. Terminals are important, as some terminals are better equipped to deal with delays, and they can recover or mitigate them; the route is also important, as some routes are chronically congested and once a train is delayed is its likely that things will get worse. We have therefore decided to develop a model based on the data obtained from the train circulation on the rail network, to discover the probability of absorbing or propagating delays over the rotations travelled by the compositions.

3 A MARKOV CHAIN BASED MODEL

We consider the delay propagation problem introduced in Section 2. Let t_0 be an arriving train with a delay δ_0 and $\{t_1, t_2, \ldots, t_n\}$ the rotation that follows train t_0. The propagation of the delay δ_0 on the train sequence $\{t_k\}_{k=1,\ldots,n}$ can be modeled by a stochastic sequence $\{X_k\}_{k=0,\ldots,n}$, where each X_k is a random variable describing the state of the delay at the arrival of train t_k.

In order to deal with a finite number of delay states we define a set of delay clusters $\mathcal{X} = \{C_0, \ldots, C_N\}$, so that for any delay $\delta \in \mathbb{R}^+$ there is exactly one cluster $C_l \in \mathcal{X}$ such that $\delta \in C_l$. In the remainder of the paper we will use for the state space $\mathcal{X} = \{C_0, \ldots, C_N\}$ the simplified notation $\mathcal{X} = \{0, \ldots, N\}$. Each random variable X_k of the stochastic sequence can assume any of the delay states $\{0, \ldots, N\}$.

In this context it is reasonable to assume that the delay state at the next time instant $(k+1)$ depends only on the current delay state at time instant k and not on previous delay states. In other words we assume that the memoryless property holds in our model. As a consequence, the stochastic sequence $\{X_k\}_{k=0,\ldots,n}$ can be classified as a discrete-time Markov chain.

A crucial issue that has to be addressed at this point is the definition of the *transition probabilities* from a given delay state i at time instant k to a delay state j at the next time instant $(k+1)$, that is

$$p_{ij}(k) = \mathbf{P}[X_{k+1} = j \mid X_k = i]. \tag{1}$$

The key idea of our approach is to extrapolate the transition probabilities of the delay propagation Markov chains from the historical delay data set. In other terms we define the transition probabilities in such a way, that the Markov chains generated are consistent with the real delay propagation patterns observed in the past. In order to generate the transition matrices $P(k) = [p_{ij}(k)]$ we focus on the delay propagation pattern of each couple of consecutive trains $\{t_k, t_{k+1}\}$ in the chain, where $k = 0, \ldots, n-1$.

With each couple of consecutive trains $\{t_k, t_{k+1}\}$ we can associate the corresponding couple of consecutive relations $\{r(t_k), r(t_{k+1})\}$. Usually trains transiting on the same relation have similar delay patterns. Therefore it is reasonable, from a statistical point of view, to aggregate the historical delay data per couples of consecutive relations.

More precisely, for each couple of relations $\{r_1, r_2\}$ that has occurred in this sequence in the past we count how many times a delay state i on relation r_1 has been followed by a delay state j on relation r_2.

Definition 1 Given a couple of relations $\{r_1, r_2\}$ we denote by $h_{ij}(r_1, r_2)$ the number of occurrences in the historical data of a delay transition from state i to state j on the couple of relations $\{r_1, r_2\}$. In the sequel we refer to the matrix $H(r_1, r_2) = [h_{ij}(r_1, r_2)]$ as the *histogram* of the couple of relations $\{r_1, r_2\}$.

Since the components $h_{ij}(r_1, r_2)$ of a histogram return the number of samples observed in the historical data, the corresponding transition probabilities $p_{ij}(r_1, r_2)$ can be obtained as

$$p_{ij}(r_1, r_2) = \frac{h_{ij}(r_1, r_2)}{\sum_{j=0}^{N} h_{ij}(r_1, r_2)} . \tag{2}$$

Definition 2 The *transition probability* $p_{ij}(r_1, r_2)$ expresses the probability that a delay state i on relation r_1 is followed by a delay state j on relation r_2. We refer to the matrix $P(r_1, r_2) = [p_{ij}(r_1, r_2)]$ as the *transition matrix* of the couple of relations $\{r_1, r_2\}$. The transition matrix of a couple of consecutive relations $\{r(t_k), r(t_{k+1})\}$ in the Markov chain will be denoted simply by $P(k)$.

It has to be remarked that in equation (2) we assume that none of the rows of the histogram $H(r_1, r_2)$ is zero. In this case each row of the transition matrix $P(r_1, r_2)$ satisfies the property

$$\sum_{j=0}^{N} p_{ij}(r_1, r_2) = 1 \tag{3}$$

and thus represents a probability distribution. In the Section 4, devoted to the application of the Markov model to the Hupac network, we will address the issue of the empty rows in the histograms.

3.1 A Model for Delay Propagation

Given an initial delay δ_0 on a train t_0, the Markov model described above enables us to calculate a forecast of the delay distribution over a rotation $\{t_1, \ldots, t_n\}$. For this purpose we need the following definition of a delay distribution.

Definition 3 We refer to $d(k)$ as the probability distribution of the delay at time instant k. Each component of the distribution $d(k)$ is defined as $d_i(k) = \mathbf{P}[X_k = i]$ and represents the probability that at time instant k the delay state is i.

It is easy to see that for any time instant k the components $d_i(k)$ sum to 1, that is $\sum_{i=0}^{N} d_i(k) = 1$. The key property of the propagation of the delay distribution $d(k)$ is formulated in the next theorem, based on the Chapman-Kolmogorov equation (see for instance Cassandras and Lafortune (1999)), which can be easily proved.

Theorem 1 Given a delay distribution $d(k)$ at the time instant k of the Markov chain $\{X_k\}_{k=0,\dots,n}$, the delay distribution $d(k+1)$ at the next time instant can be calculated as

$$d(k+1) = P(k)^T \cdot d(k),\tag{4}$$

where $P(k)$ is the corresponding transition matrix.

By iterating equation (4) we immediately obtain the following property.

Corollary 2 Let be $d(0)$ the initial delay distribution. The distribution $d(n)$ at time instant n can be calculated as

$$d(n) = (\prod_{k=0}^{n-1} P(k))^T \cdot d(0).\tag{5}$$

Note that in our application the initial delay δ_0 is given. Therefore we initialize the delay distribution $d(0)$ as follows:

$$d_i(0) = \begin{cases} 1 & \text{if } \delta_0 \in C_i \\ 0 & \text{otherwise} \end{cases}$$

that is 1 for the delay state associated with δ_0 and 0 otherwise. The distribution $d(n)$ represents the delay forecast for the time instant n, that is when the sequence of travels $\{t_k\}_{k=1,\dots,n}$ has been completed.

4 USING THE MODEL ON THE HUPAC NETWORK

The delay propagation model of Section 3 is based on delay patterns observed in the past. In order to obtain the delay distributions and to generate the transition matrices needed in the Markov model, we analyzed the delay data of 20'000 compositions that circulated on the Hupac transportation network. The Hupac source data contained the arrival delays in minutes of each composition.

4.1 Delay Clusters

First of all we classified the delays of the compositions in five clusters, as this rough classification is the one usually adopted by the train dispatchers in our case study. This choice is arbitrary and it might not suit every other rail freight company. Data mining algorithms could also be used to find out the best separation into clusters, but we did not applied them in this work. In Table 1 we report for each cluster C_l its span in minutes and the percentage of the compositions with an arrival delay $\delta \in C_l$.

In the following we refer to $\mathcal{X}$ as the set of five delay states $\mathcal{X} = \{C_0, \dots, C_4\}$ or simply $\mathcal{X} = \{0, \dots, 4\}$.

4.2 Delay Distributions Over Relations

A first overview of the source data reveals that, on average, some relations are much more affected by delays than others. An explanation of this phenomenon might be that as the circulation plan is iterated,

Table 1: Delay clusters.

cluster	extension (min)	compositions (%)
C_0	$\delta = 0$	69.2
C_1	$0 < \delta \leq 60$	10.6
C_2	$60 < \delta \leq 180$	7.8
C_3	$180 < \delta \leq 360$	4.4
C_4	$\delta > 360$	8.0

also the delays occur periodically according to similar patterns, caused for instance by critical idle times or lack of efficiency at certain terminals or by systematic traffic congestions.

In order to visualize the different delay patterns, we aggregated the delay data per relations. We remind that a relation connects a departure and a destination terminal. The Hupac transportation network is composed by about 170 relations. It has to be observed that there is a great variance in the intensity of the traffic on the relations. Some are run daily or even more times a day, others only weekly. In our analysis we focus on relations that provide a high number of samples. In Table 2 we show the delay distributions of 30 relevant relations that are run with a very high frequency. In the observation period of six months each relation reported has been served between 150 and 500 times. In the first column we report the codenames of the departure and destination terminals of the relations. In the central column we show the distributions of the observed delays with respect to the five clusters $C_0, \ldots, C_4$. The last two columns contain the expected value (Avg) and the standard deviation (Std) of the corresponding delay distribution. The expected value of a given delay distribution d is calculated as $\bar{C} = \sum_{i=0}^{4} i \cdot d_i$. As an example, on the relation AARA-STAB 98% of the observed trains had no delay, while 2% were in delay cluster C_2. The expected value is 0.03, which means that, on average, trains in this relation run in the delay cluster C_0. The low value 0.23 of the standard deviation suggests that the relation AARA-STAB is not affected by significant delay fluctuations.

The relations in Table 2 are ordered by an increasing value of the averages. The table clearly shows that there is a strong difference in the delay patterns of the relations. Some relations never report delays, others occasional delays and finally there are relations with frequent heavy delays. The moderate size of the standard deviations suggests that the delay patterns are quite stable on the observation period. However, it is interesting to remark that in general the standard deviations grow together with the averages. This fact means that usually the relations with frequent heavy delays are also the more unpredictable.

Table 2: Delay distributions of 30 relations.

From	To	C_0	C_1	C_2	C_3	C_4	Avg	Std	From	To	C_0	C_1	C_2	C_3	C_4	Avg	Std
ANMU	BIRR	1.00	0.00	0.00	0.00	0.00	0.00	0.00	AARA	ANMU	0.70	0.11	0.08	0.03	0.08	0.69	1.24
MALM	TAUL	1.00	0.00	0.00	0.00	0.00	0.00	0.00	BUS2	MALU	0.68	0.12	0.10	0.03	0.07	0.70	1.20
TAUL	MALM	1.00	0.00	0.00	0.00	0.00	0.00	0.00	ANMU	BASE	0.60	0.22	0.07	0.03	0.08	0.79	1.22
HOTA	TAUL	1.00	0.00	0.00	0.00	0.00	0.00	0.00	SINN	MICE	0.66	0.05	0.12	0.16	0.01	0.80	1.21
TAUL	HOTA	1.00	0.00	0.00	0.00	0.00	0.00	0.00	BUS2	ANMU	0.71	0.05	0.05	0.06	0.13	0.86	1.46
AARA	STAB	0.98	0.00	0.02	0.00	0.00	0.03	0.23	BASE	ANMU	0.61	0.14	0.12	0.03	0.10	0.88	1.33
BUS2	POME	0.96	0.02	0.02	0.00	0.00	0.06	0.31	BUS2	KOEL	0.57	0.15	0.14	0.04	0.10	0.98	1.35
POME	BUS2	0.84	0.13	0.02	0.01	0.00	0.21	0.53	KOEL	BUS2	0.41	0.29	0.16	0.05	0.09	1.11	1.25
BUS2	TAUL	0.91	0.02	0.03	0.01	0.03	0.23	0.83	MALU	BUNA	0.24	0.42	0.18	0.09	0.07	1.34	1.15
STAB	AARA	0.81	0.15	0.04	0.00	0.00	0.24	0.53	TAUL	BUS2	0.34	0.23	0.17	0.04	0.22	1.54	1.52
MALU	SCHD	0.81	0.07	0.09	0.02	0.01	0.34	0.78	MALU	BUS2	0.38	0.17	0.15	0.11	0.19	1.55	1.54
BIRR	ANMU	0.81	0.06	0.06	0.02	0.05	0.46	1.06	ROTT	NOVA	0.53	0.05	0.03	0.06	0.33	1.63	1.85
SCHD	MALU	0.82	0.04	0.04	0.01	0.09	0.50	1.21	NOVA	ROTT	0.45	0.08	0.09	0.07	0.31	1.72	1.77
BUNA	MALU	0.82	0.04	0.04	0.01	0.09	0.50	1.21	ANMU	BUS2	0.27	0.26	0.17	0.07	0.23	1.76	1.51
MICE	SINN	0.58	0.33	0.03	0.00	0.06	0.65	1.03	ANMU	AARA	0.12	0.20	0.23	0.38	0.07	2.10	1.15

4.3 Propagation of Delays Through Consecutive Relations

After the results obtained through the statistical analysis of the delays on single relations, we focused our study on the delay propagation patterns on couples of consecutive relations. More precisely, according to the procedure defined above, starting from the historical delay data set of the Hupac network, we generated the transition matrices $P(r_1, r_2)$ for all the couples of relations $\{r_1, r_2\}$ that have been served consecutively.

In the observed period we identified 233 different couples of consecutive relations. For each couple $\{r_1, r_2\}$ we generated its histogram $H(r_1, r_2)$ and the corresponding transition matrix $P(r_1, r_2)$. As an example we report in Table 3 the transition matrix of the couple of relations {BUS2-KOEL,KOEL-BUS2} based on 166 samples.

Table 3: Transition matrix of the couple of relations {BUS2-KOEL,KOEL-BUS2}.

	C_0	C_1	C_2	C_3	C_4
C_0	0.39	0.43	0.10	0.04	0.04
C_1	0.34	0.16	0.34	0.05	0.11
C_2	0.36	0.35	0.18	0.04	0.07
C_3	0.56	0.22	0.11	0.00	0.11
C_4	0.32	0.09	0.27	0.14	0.18

According to Definition 2 we read from Table 3 for example that 39% of the trains that arrived on time on the relation BUS2-KOEL, came back on time to BUS2, while 43% of the trains on time the way there report a delay in cluster C_1 on the way back.

As already mentioned, equation 2, which defines the transition probabilities, requires that each row of the transition matrix sums to one. In case the histogram that originates the transition probability matrix (see equation 2) contains too few data, it might happen that this property is not verified; in such cases we inserted in the corresponding row of the transition matrix $P(r_1, r_2)$ the distribution obtained by the sum of all the rows of the histogram $H(r_1, r_2)$. In this way we add a weighted average of all the other rows, thus making an extrapolation from other observations on the same relation. However, it has to be pointed out that usually this situation does not arise when the number of available samples for a couple of relations is sufficiently high.

A detailed observation of the 233 histograms $H(r_1, r_2)$ and of the corresponding transition matrices $P(r_1, r_2)$ makes clear that the way delays are propagated from one relation to the next strongly depends on the sequence of relations considered. Since it is not possible to present here the whole set of histograms and transition matrices, in Table 4 we show the 20 most frequent couples of relations, characterized by four statistical indicators. The first column of Table 4 contains the departure and the destination terminals of the first relation r_1 and of the second relation r_2. In order to characterize a histogram $H(r_1, r_2)$ it is useful to compute a delay distribution $d(r_1)$ by summing up all the columns of $H(r_1, r_2)$ and dividing by the number of samples in the histogram. In an analogous way a distribution $d(r_2)$ is obtained by the sum of the rows of $H(r_1, r_2)$ divided by the number of samples. The distribution $d(r_1)$ can be interpreted as the delay distribution right after the arrival on the first relation r_1, while $d(r_2)$ is the analogous distribution after r_2. The indicators Avg 1 and Avg 2 are the expected values of the two distributions $d(r_1)$ and $d(r_2)$, respectively, while Std 1 and Std 2 are the corresponding standard deviations. The indicators Avg 1 and Avg 2 in Table 4 clearly show that on certain sequences delays tend to be amplified, while on others they tend to be reduced or even absorbed. For instance, the couples of relations {SINN-DUIT,DUIT-SINN} and {AARA-ANMU,ANMU-AARA} present a clearly increasing delay pattern. On the other hand sequences like {BUS2-SINN,SINN-BUS2} and {TAUL-BUS2,BUS2-TAUL} have a decreasing delay pattern. Due to this strong differentiation in the delay propagation patterns observed on couples of relations it is reasonable to calculate a delay forecast by means of the Markov chain model.

Table 4: Delay propagation on 20 couples of relations, averages and standard deviations.

Dep 1	Dest 1	Dep 2	Dest 2	Avg 1	Avg 2	Std 1	Std 2
BUS2	POME	POME	BUS2	0.00	0.21	0.00	0.52
SINN	DUIT	DUIT	SINN	0.00	0.94	0.00	1.25
AARA	VISP	VISP	AARA	0.00	1.10	0.00	1.79
ROTT	VRQU	VRQU	ROTT	0.00	2.30	0.00	1.69
POME	BUS2	BUS2	POME	0.17	0.00	0.38	0.00
AARA	KOEL	KOEL	AARA	0.52	0.68	0.67	1.24
SINN	MICE	MICE	SINN	0.69	0.79	1.15	1.15
AARA	ANMU	ANMU	AARA	0.80	2.03	1.29	1.14
MICE	SINN	SINN	MICE	0.81	0.75	1.18	1.19
BUS2	SINN	SINN	BUS2	0.84	0.58	1.51	0.85
KOEL	BUS2	BUS2	KOEL	0.92	1.24	1.09	1.38
DUIT	SINN	SINN	DUIT	0.96	0.00	1.25	0.00
BUS2	MALU	MALU	BUS2	1.04	1.99	1.26	1.59
BUS2	ANMU	ANMU	BUS2	1.05	1.99	1.47	1.54
BUS2	KOEL	KOEL	BUS2	1.26	1.19	1.39	1.23
ROTT	NOVA	NOVA	ROTT	1.45	1.71	1.82	1.74
ROTT	MALU	MALU	ROTT	1.59	1.63	1.36	1.36
MALU	ROTT	ROTT	MALU	1.69	1.56	1.34	1.36
TAUL	BUS2	BUS2	TAUL	1.73	0.00	1.63	0.00
NOVA	ROTT	ROTT	NOVA	1.75	1.44	1.74	1.83

4.4 Forecasting Delays

The transition matrices obtained as discussed in Section 4.3 enable us to forecast on the Hupac transportation network the risk of delays after a given rotation. In fact, given an initial delay distribution $d(0)$ and a sequence of transition matrices $\{P(k)\}_{k=0,\dots,n-1}$, the forecasted delay distribution $d(n)$ is calculated by means of equation (5).

As an example, let's consider a train with a delay of 80 minutes on the relation BUS2-KOEL, which is assigned to the rotation {KOEL-BUS2, BUS2-KOEL}. What kind of delay distribution might be expected after the rotation? In order to calculate the delay forecast $d(2)$ we set $d(0) := [0,0,1,0,0]^T$, as $\delta_0 = 80$ minutes belongs to the cluster C_2. Then there are two transition matrices involved: $P(0)$ is the transition matrix of the first couple of relations {BUS2-KOEL,KOEL-BUS2} shown in Table 3, and $P(1)$ is the matrix of the next couple {KOEL-BUS2, BUS2-KOEL}, reported in Table 5. The forecasted delay distribution $d(2)$ is calculated as $d(2) = (P(0) \cdot P(1))^T \cdot d(0) = [0.43, 0.21, 0.19, 0.05, 0.12]$. This result means that, according to the delay evolution observed in similar situations in the past, we can expect that the train will absorb the present delay with a probability of 43%, whereas a delay of cluster C_1 can be expected with a probability of 21%, and so on.

Table 5: Transition matrix of the couple of relations {KOEL-BUS2,BUS2-KOEL}.

	C_0	C_1	C_2	C_3	C_4
C_0	0.40	0.21	0.14	0.07	0.18
C_1	0.45	0.35	0.15	0.05	0.00
C_2	0.45	0.00	0.45	0.01	0.09
C_3	0.43	0.21	0.19	0.05	0.12
C_4	0.50	0.00	0.00	0.00	0.50

In order to compare the quality of the assignment of a train t_0 to different possible rotations it is necessary to reduce the information of the output distribution $d(n)$ to numerical indicators that can be easily interpreted.

Definition 4 Let be t_0 a train with a delay δ_0 assigned to a rotation $\{t_k\}_{k=1,\ldots,n}$. We define the *assignment index* $I(\{t_k\}, \delta_0)$ as

$$I(\{t_k\}, \delta_0) = \frac{1}{N} \sum_{i=0}^{N} i \cdot d_i(n), \tag{6}$$

where $d(n)$ is the output distribution after the rotation $\{t_k\}_{k=1,\ldots,n}$.

The assignment index $I(\{t_k\}, \delta_0)$ returns the expected value of the distribution $d(n)$ divided by N. Substantially, the assignment index assumes a value near to 0 if the expected delay is low, while it is near 1 if the expected delay belongs to the highest cluster C_N. In the previous example we obtain $I(\{t_k\}, 80) = 0.30$, which corresponds to the expected value 1.21 of the distribution $d(2)$.

Since the reliability of the forecast strongly depends on the number of samples considered in each histogram in the Markov chain, we define a reliability indicator.

Definition 5 Let be $H(k)$ the histogram related to the couple of consecutive relations $\{r(t_k), r(t_{k+1})\}$. We denote by $|H(k)|$ the total number of samples in $H(k)$, that is $|H(k)| = \sum_{i,j=0}^{N} h_{ij}(k)$. We define the *significance index* of the histogram $H(k)$ as

$$s(k) = \begin{cases} 0.1 & \text{if } |H(k)| < 5 \\ 0.02 \cdot |H(k)| & \text{if } 5 \le |H(k)| \le 50 \\ 1 & \text{if } |H(k)| > 50. \end{cases}$$

In other terms, to each histogram $H(k)$ we associate a scalar value between 0.1 and 1 that increases as the number of available samples increases. Now we are ready to define the reliability index $rel(\{t_k\})$.

Definition 6 Given a sequence of trains $\{t_k\}_{k=0,\ldots,n}$ we define the reliability index of the forecast on the sequence $\{t_k\}_{k=0,\ldots,n}$ as

$$rel(\{t_k\}) = \prod_{k=0}^{n-1} s(k). \tag{7}$$

It is easy to see that the value of the reliability index is near to 1 only if all the histograms involved in the sequence have a high significance value. Otherwise the reliability index assumes a value near to zero. In the case of the previous example we obtain a reliability index $rel(\{t_k\}) = 1$, because both transition matrices have a significance index equal to 1.

5   VALIDATION OF RESULTS

A question that arises naturally at this point is to what extent the delay distributions and the transition matrices based on the historical data are consistent with the distributions observed day by day on the transportation network. In order to validate the statistical results reported in Section 4, we compared the historical distributions with the ones observed on the Hupac transportation network on a control period of five months (from February to June 2011). The control data set contains the delays of about 17'000 trains. We aggregated them per relations and computed the delay distributions. Figure 1 shows a comparison of the averages of the 50 most frequently served relations. On the abscissa we report the averages of the historical data and on the ordinate the averages observed in the control set.

Figure 1 shows that there is a strong correlation between the delays in the control period and the historical delays. For instance, low delay averages in the historical data set are usually matched by low delay averages in the control set.

Moreover, during the control period we observed the delay propagation patterns on the 50 most significant couples of relations and compared them with the corresponding transition matrices derived from the historical data. In particular, we calculated for each couple of relations $\{r_1, r_2\}$ the difference between the delay average after relation r_2 (Avg 2) and the delay average after relation r_1 (Avg 1). We refer to this
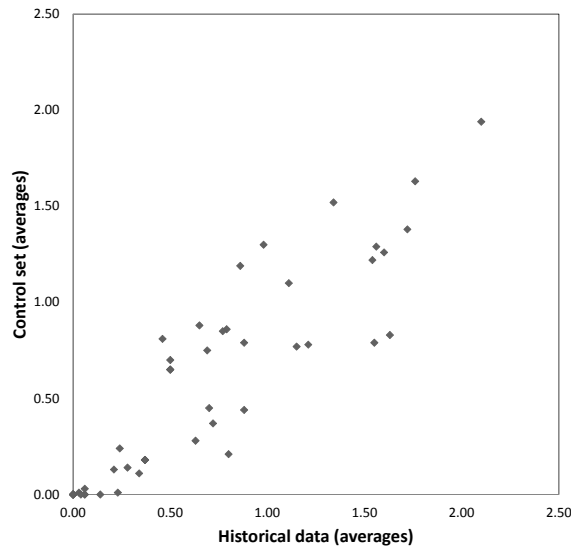
Figure 1: Delay averages, comparison on 50 relevant relations.

indicator as delay variation, because it expresses the variation of the delay average from the first relation r_1 to the second r_2. A positive variation indicates an increasing trend of the delays on the sequence $\{r_1, r_2\}$, on the other hand a negative variation means that the delays tend to decrease on the sequence. The advantage of the delay variation indicator is that it captures the propagation pattern of each couple of relations in a single number. In Figure 2 we report a comparison of the delay variations of 50 relevant couples of relations between the control period and the historical period. A clear linear correlation can be observed: to positive variations in the historical data usually correspond positive variations in the control period and similarly for negative variations.

This is mainly due to the periodicity of the circulation plan, which generates daily, or at least weekly, similar delay scenarios. Basically, this periodicity justifies the approach of forecasting delays on a railroad network by means of historical data. On the other hand, in some cases considerable deviations can be observed. An explanation might be that extraordinary events, such as railroad damages by bad weather conditions or severe accidents, might alter the usual delay patterns. Furthermore, a seasonal fluctuation of the delay distributions can be expected as well. In order to reduce the impact of these factors it would be necessary to consider a significantly larger delay data set. However, such an extensive data analysis lays beyond the scope of our present study.

6   CONCLUSIONS

In this paper we have presented an algorithm to forecast the propagation of delays in a rail freight network based on a Markov Chain model. We have then discussed its calibration and we have run a number of simulations to validate the results. The algorithm has been implemented in a Decision Support System, which provides support to intermodal transport operators for both planning the optimal circulation plan, and also in the operational dispatching of trains at the terminal level. In particular the delay propagation algorithm is used by the dispatcher to evaluate the future impact of train assignment choices on the future state of the train circulation. The same algorithm is also used in a matching algorithm which is used to compute efficient train assignment choices, considering also a number of performance indicators, such as the compliance of a composition to the desired structure of a composition serving a given relation. The
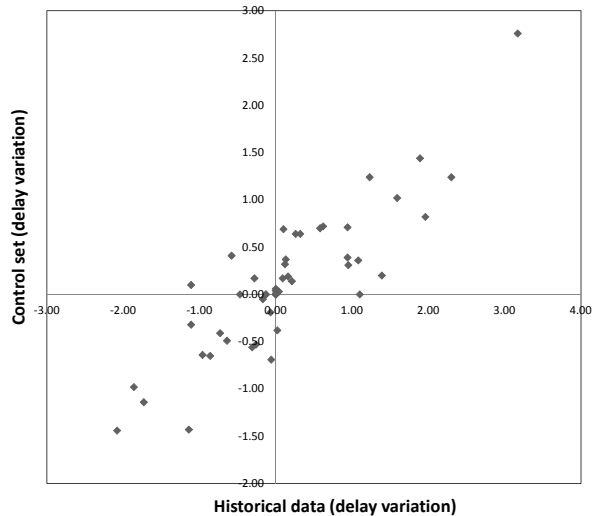
Figure 2: Delay variation, comparison of 50 relevant couples of relations.

DSS is currently used by the team of train dispatchers at Hupac who are routing more than 100 trains per day all over Europe.

7   ACKNOWLEDGEMENTS

REFERENCES

Acuna-Agost, R., P. Michelon, D. Feillet, and S. Gueye. 2011. "SAPI: Statistical Analysis of Propagation of Incidents. A new approach for rescheduling trains after disruptions". *European Journal of Operational Research* 215 (1): 227–243.

Assad, A. 1979. "Models for rail transportation". *Transportation Research Part A* 14:205–220.

Benna, T., and M. Gronalt. 2008, December. "Generic Simulation for Rail-Road Container Treminals". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Moench, O. Rose, T. Jefferson, and J. W. Fowler, 2656–2660. Piscataway, New Jersey: IEEE, Inc.

Caprara, A., M. Fischetti, and P. Toth. 2002. "Modeling and solving the train timetabling problem". *Operations Research* 50:851–861.

Cassandras, C. G., and S. Lafortune. 1999. *Introduction to Discrete Event Systems*. Norwell, Massachusetts: Kluwer Academic Publishers.

D'Ariano, A., D. Pacciarelli, and M. Pranzo. 2007. "A branch and bound algorithm for scheduling trains in a railway network". *European Journal of Operational Research* 183 (2): 643 – 657.

D'Ariano, A., and M. Pranzo. 2009. "An Advanced Real-Time Train Dispatching System for Minimizing the Propagation of Delays in a Dispatching Area Under Severe Disturbances". *Networks and Spatial Economics* 9 (1): 63–84.

European Commission Eurostat 2012. "Freight transport statistics".

Higgins, A., and E. Kozan. 1998. "Modeling Train Delays in Urban Networks". *Transportation Science* 32 (4): 346–357.

Huang, Y., M. D. Seck, and A. Verbraeck. 2010, December. "LIBROS-II: railway modelling with DEVS". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 2150–2160. Piscataway, New Jersey: IEEE, Inc.

Hupac 2012. "Profile of the Hupac Group".

Leilich, R. H. 1998, December. "Application of simulation models in capacity constrained rail corridors". In *Proceedings of the 1998 Winter Simulation Conference*, edited by D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 1125–1133. Piscataway, New Jersey: IEEE, Inc.

Lewellen, M., and K. Tumay. 1998, December. "Network simulation of a major railroad". In *Proceedings of the 1998 Winter Simulation Conference*, edited by D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 1135–1138. Piscataway, New Jersey: IEEE, Inc.

Lin, E., and C. Cheng. 2009, December. "YARDSIM: a rail yard simulation framework and its implementation in a major railroad in the U.S.". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 2532–2541. Piscataway, New Jersey: IEEE, Inc.

Minger, W. K., and J. N. Cetinich. 1969, December. "Association of American Railroads's Network Model". In *Proceedings of the 1969 Winter Simulation Conference*, edited by A. Ockene and P. Kiviat, 193–203. Piscataway, New Jersey: IEEE, Inc.

Özekici, S., and S. Sengör. 1994. "On a Rail Transportation Model with Scheduled Services". *Transportation Science* 28 (3): 246–255.

Rizzoli, A., N. Fornara, and L. Gambardella. 2002. "A simulation tool for combined rail/road transport in intermodal terminals". *Mathematics and Computers in Simulation* 59 (1–3): 57–71.

AUTHOR BIOGRAPHIES

JÁNOS BARTA studied mathematics at the Swiss Federal Institute of Technology of Zurich (ETHZ). From 1998 to 2003 he worked as a researcher and lecturer at the Zurich University of Applied Sciences (ZHAW). Since 2003 he is a lecturer at the University of Applied Sciences of Southern Switzerland (SUPSI) and since 2004 is an associate researcher with the Istituto dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA). His main research interests are combinatorial and robust optimization.

ANDREA EMILIO RIZZOLI received his Ph.D. in Control Engineering and Informatics from Politecnico di Milano (Italy) in 1993. He was a PostDoc researcher at CSIRO, Australia. Presently, he is a senior research scientist at IDSIA, Istituto Dalle Molle di studi sull'Intelligenza Artificiale, a research institute of the University of Applied Sciences of Southern Switzerland (SUPSI) and of the University of Lugano (USI). His current research interests focus on methodologies and techniques for model re-use and encapsulation in different domains, from environmental systems to transportation and logistics systems.

MATTEO SALANI received his PhD in Computer Science in 2006 from University of Milan and he is part of IDSIA since January 2010. He has been PostDoc researcher at EPFL (Transport and Mobility laboratory) from 2006 to 2009. He is active in the field of discrete optimization. In particular, his main scientific contributions are related to Column Generation methods applied to routing and scheduling problems. He published some relevant articles dedicated to the resource constrained elementary shortest paths problem.

LUCA MARIA GAMBARDELLA is research director at IDSIA, in Lugano. His major research interests and publications are in the area of optimization, simulation, swarm robotics, multi-agent learning, applied to academic and real-world problems. He has invented and developed several Ant Colony Optimization algorithms able to compute best-known solutions for many benchmark instances. He leads research and applied projects, funded among others by the Swiss National Science Foundation, the European Commission, the Swiss Technology and Innovation Commission, the Swiss Hasler Foundation and private companies.