Università degli Studi di Napoli Federico II

Facoltà di Ingegneria

Dipartimento di Ingegneria Aerospaziale

Dottorato di ricerca in Ingegneria Aerospaziale, Navale e della Qualità

Indirizzo Gestione della Qualità Totale

Quality, Availability and Lifecycle Cost of Transportation Systems

TUTORS
Prof. Pasquale Erto
Prof. Massimiliano Giorgio

DOTTORANDO
Michele Scuotto

COORDINATORE
Prof. Antonio Moccia

XXIV ciclo

To my parents, my family and

my sweet Marianna

Index

1 INTRODUCTION

In recent years the development of policies aiming at promoting the use of public transport, thus decreasing the demand for private transportation, assumed a continuously growing importance (Goldman, Gorham, 2006).

Promoting public transportation is in fact one of the main levers to ensure environmental sustainability: an increased use of public transport, and a consequent reduction of the number of vehicles on the road, may significantly lessen global vehicle emissions. So, great attention is given by Transport Authorities and researchers to define strategies aiming at encouraging the use of public transport. Among these, ensuring that passengers perceive high service quality is of paramount importance.

Under this perspective, service punctuality and customers' waiting time are of great concern (Van Hagen et al., 2007), (dell' Olio et al., 2011). Ensuring that a transport service has adequate on-time performances, besides the use of highly dependable technologies, requires the definition of organizational settings and maintenance management policies capable of minimizing service degradation in failure conditions.

So, three specific needs arise for Transport Authorities. First, service quality perceived by passengers should be measured and the effects of service improving strategies estimated via customer oriented, clear and easy to apply approaches, based on a solid stochastic and statistical framework. Second, Reliability, Availability, Maintainability (RAM) characteristics of transport systems should be controlled due to their strong relationship with service on-time performances, and hence with service quality. Third, the effects of organizational settings and maintenance management policies should be incorporated in the assessment of transport systems RAM performances.

The need to control RAM performances of transport systems and to embed, at the same time, the effects of organizational settings and maintenance management policies in RAM assessments, led the majority of Transport Authorities to focus on operational RAM indexes. These indexes, unlike inherent ones, directly account for the above effects and virtually all invitations to tender for transport systems contain specific operational RAM requirements for the whole system and its main constituents. Moreover, severe penalty payments are foreseen in the event that contractual RAM requirements are not met.

In addition, Transport Authorities are now assigning a substantial weight to RAM performances in tender evaluation processes, due to their strong relationship with service quality and system lifecycle cost.

Consequently, companies operating in the transport sector shall be capable of designing and controlling Quality and operational RAM performances of the systems they intend to deliver. In addition, they should be capable of quantifying the effects of reliability objectives on the system lifecycle cost since the tender phase.

A proper assessment of the above performances since the early stages of development, besides being a need arising from the market, is a twofold advantage. First, it allows to evaluate operational RAM performances associated to different design alternatives, thus enabling companies to choose the best technical proposal according to tender evaluation criteria. Second, it allows to timely define design guidelines to be taken into account to meet operational RAM requirements, thus avoiding costly and time consuming re-design or even retrofit activities during revenue operation.

In such a context, the definition of methods and tools for designing and controlling service quality, operational RAM performances and the effects of reliability objectives on the lifecycle cost is of crucial importance for both Transport Authorities and companies delivering transport systems. In fact, Transport Authorities aim at offering an high quality transport service, whereas companies delivering transport systems have to develop technical solutions able to perfectly fit tender evaluation criteria and hence to regard RAM performances as a source of competitive advantage.

1.1. Research Objectives

The main objective of this research consists in the development of probabilistic and statistical models aimed at aiding Transport Authorities and companies operating in the Transport sector in:

- Evaluating Quality of Transport services;
- Assessing the effects of organizational settings and maintenance management policies on operational availability of transportation systems;
- Assessing the effects of reliability objectives on the System Lifecycle cost.

In the field of service quality evaluation, lots of studies concerning the importance of transport service quality attributes relevant for the customers are available and different approaches have been proposed to measure service quality. These approaches, mainly rely on aggregate indexes based on the abovementioned attributes. Thus, if even these approaches allow Transport Authorities to know what really matters for customers and to identify areas characterized by a room for a service quality improvement, they cannot be easily used to estimate the effects of service changes neither as a self assessment tool. For such purposes, the need of quality indexes that can be calculated on the basis of commonly available data and information arises.

From this point of view, the advancement this research activity attempts to bring consists in proposing a stochastic approach capable of satisfying the above need.

As previously highlighted, another important concern of Transport Authorities and companies delivering transport systems is represented by operational RAM performances of transport systems. Here the main role is played by the fleet of vehicles and by decisions concerning fleet and maintenance service dimensioning. Fleet dimensioning is often performed by transit agencies adopting, as mandatory requirement, spare ratios recommended by funding agencies (e.g. Federal Transit Administration).Unfortunately, the above ratios are not available for all kinds of transport systems. In addition, using spare ratios doesn't allow to account for mutual interactions between fleet size and maintenance dimensioning. Thus, trade-off analyses cannot be performed. In order to overcome this issue, Markov models can be used. Unfortunately, not always Markov models work adequately: while the hypothesis of exponentially distributed failure times is usually met in practice, the hypothesis of exponentially distributed repair times is often not realistic and should be removed to obtain satisfactory results, since they depend on the repair time distribution. Considering non exponential repair times, leads one to handle non-Markovian processes, which are more difficult to treat, from both an analytical and numerical standpoint, than Markov processes. From this point of view, aim of this research activity consists in identifying an approach for the analysis of the effects of maintenance management decisions on the operational availability of a fleet of vehicles. More specifically, this research aims at defining an approach capable of taking into account all main factors influencing availability and of preserving, at the same time, analytical and numerical tractability of the resulting models.

As previously highlighted, great attention is given by Transport Authorities to the System lifecycle cost, which is greatly influenced by system reliability and maintainability characteristics. Thus, in order to be competitive, companies delivering transport systems should be capable of assessing the effects of reliability objectives on the system lifecycle cost. As far as I know, these assessments are not performed in the industry. In fact, in a typical industrial environment, lifecycle cost estimates are performed by analogy, on the basis of costs actually born to design, build, operate and maintain similar systems. Moreover, the translation of costs born for a given system in costs to be born for a new system is largely subjective and empirical, especially for costs related to the efforts to be sustained during the early stages of operation to improve the system reliability level. From this point of view, another aim of this research activity consists in developing a model capable of explaining and explicitly accounting for costs depending on reliability: failures during revenue operation, the acquisition of a given inherent reliability level and efforts required to reach the planned reliability target. More in detail, attention is focused on the formulation and parameter estimation of a failure intensity which properly models the failure rate behavior of complex repairable systems during the early stages of operation, including field testing, and the useful life.

1.2. Research Methodology

Being this research activity strongly stimulated by practical needs arising in the Transport sector, it has been devoted to define approaches and methodologies that are easy to apply and, at the same time, rigorous from a statistical and probabilistic point of view. So, for each of the considered engineering problems, the most important variables/factors are identified and a proper set of statistical and probabilistic tools to deal with them are chosen. Then, the main hypotheses arising from the previously chosen tools and methodologies to check them are identified; acceptability of approximations due to numerical procedures is checked as well. At last, in order to validate the proposed approaches, they are applied to real case studies and the acceptability of the underlying hypotheses is properly checked.

In order to ensure practical usability of the proposed methodologies, great attention has been given in managing the trade-off between their tractability and their capability to properly deal with the probabilistic features characterizing the considered engineering problems. In this sense, besides the precious suggestions of my supervisors, the collaboration with engineers and managers of the RAMS department of Ansaldo STS S.p.A. has been extremely helpful. Also the choice of commercial software, such as MATLAB® and Mircrosoft Excel, to implement routines required to apply the proposed methodologies has been made to facilitate their practical usability.

1.3. Products of the Research activity

The performed research activity led to the following publications/talks in International Conferences:

- Erto P, Giorgio M, Scuotto M (2010). *Statistical Quality Indexes for a public bus service*, in: Methods, Models and Information Technologies for Decision Support Systems, Pescara, 12-15 September , 199-202.
- Di Tommaso P, Giorgio M, Scuotto M, Testa A (2011). *Operational Availability evaluation of Complex Systems with non-exponential downtimes*, contributed talk in: Games and Decisions in Reliability and Risk, Belgirate 19-21 May http://www.mi.imati.cnr.it/conferences/gdrr11/talks.html.

The research activity also led to the following submission to an international journal:

- Erto P, Giorgio M, Scuotto M. *Statistical Tools for evaluating mass transport* *service quality*, submitted to Applied Stochastic Models in Business and Industry.

There is also the following paper in progress, to be submitted to an international journal:

- Di Tommaso P, Erto P, Giorgio M, Scuotto M, Testa A, *Operational Availability assessment of Transportation Systems with non-exponential downtimes.*

At last, a special mention to the Ansaldo STS Innovation Award 2010 has been gained with the following innovative idea (see attachments to chapter 1 for Ansaldo STS CEO acknowledgements):

- Lamberti I, Mormile T, Nardone R, Scuotto M, *Metro Model Sim: a tool for supporting metro systems Design with the aim of cost reduction, fulfilling performance requirements.*

1.4. Thesis Outline

This thesis consists of four distinct parts, an introductory framework and three chapters in which methodologies proposed to manage the three considered engineering problems are applied and discussed.

The framework aims at giving an overview of probabilistic tools used to formulate the models and methodologies proposed for the considered engineering problems. Attention is focused on stochastic processes used to develop the proposed approaches and methodologies.

In chapter 3, a service quality evaluation approach, based on a set of quality indexes related to the customer waiting time, is introduced and applied to evaluate the performances of a bus route operated by A.N.M. (Azienda Napoletana Mobilità). A proper checking of the main working hypotheses on the basis of real data shows that the above methodology can be very helpful for quality evaluation of a high frequency bus service.

In chapter 4 are presented stochastic models that can be used to evaluate the impact of different organizational settings and maintenance management policies on the operational availability of a fleet of vehicles. Different configurations concerning spare vehicles and maintenance crews are considered. Moreover, repair times are realistically assumed to be non-exponential random variables, thus leading one to manage non-Markovian stochastic processes. Fleet operational availability is computed via Device of Stages technique and, for comparison purposes, via Monte Carlo Simulation too. It is shown that the proposed non-Markovian models based on the Device of Stages technique can be more accurate than pure Markov models and formulas often used by practitioners. In addition, it is shown that they require less processing capability than that required by Monte Carlo Simulation.

In chapter 5, a lifecycle cost – reliability model is presented. More precisely, attention is focused on the formulation and parameter estimation of a failure intensity model capable of fitting early and useful life failures of a complex repairable system. The model allows to count via a Non Homogeneous Poisson Process the failures that must be financially supported during the early and useful life of a complex repairable system. Model validity is checked fitting the proposed model to the failure process experienced by trains running in the Copenhagen driverless metro system during the first two years of operation. In addition, an illustrative application of the proposed Lifecycle cost – Reliability model is performed, in order to highlight that the proposed model is capable of explicitly accounting for the influence of the reliability objectives on costs related to system acquisition, development, early and useful life operation.

2 THE ENGINEERING PROBLEMS AND THE STOCHASTIC FRAMEWORK

2.1. Introduction

In this chapter, an overview of probabilistic tools used to formulate the models and methodologies proposed for the considered engineering problems is given. First, basic definitions and general concepts concerning stochastic processes are recalled. Then, attention is focused on stochastic processes involved in each application developed to cope with the engineering problems described in the previous chapter. More in detail, an overview on point process is provided. These processes have been extensively used to develop the applications described in chapters 3 and 5. More precisely, the Renewal Process (RP) has been found to be adequate in modeling the bus headways (i.e. times between two consecutive bus arrivals at a given stop), whereas the Homogenous Poisson Process (HPP) is adequate to model passengers arrivals at a given stop associated to an high frequency transport service. A Non Homogenous Poisson Process (NHPP) based on a hyperbolic failure intensity has been found to be adequate to model the failure rate behavior during the first two years of operation of the fleet of light rail vehicles running in the Copenhagen Metro System, thus representing a solid basis to model the number of failures to be financially supported during the early stages of operation, including testing, and the useful life of a complex repairable system.

Once provided an overview on point processes, Markov and non-Markovian stochastic processes are introduced. An extensive use of these stochastic processes is made in chapter 4, where the importance of the inherent repair distribution in operational availability assessments is highlighted. Phase-type distributions, which allow to treat non-Markovian processes without losing numerical and mathematical tractability of Markov processes, are characterized as well.

2.2. Stochastic processes: basic concepts and definitions

A stochastic process is a family of random variables $\{X(t)|t \in T\}$, defined on a given probability space, indexed by the parameter *t*, where *t* varies over an index set *T*. Experimental observations $\{X(t_1),...,X(t_k)\}$, $t_1,...,t_k \in T$ constitute a realization of the process. The set of all possible values that random variables can take is called the state space.

If the state space of a stochastic process is discrete, it is called a discrete state process. If the state space is continuous, then the stochastic process is a continuous-state process. Stochastic processes can be discrete-parameter processes or continuous-parameter processes if the parameter set *T* is discrete or continuous, respectively.

In order to fully characterize a stochastic process, it is necessary to specify in which way the joint distribution of the random variables constituting the process can be derived.

Important characteristics of a stochastic process are the following functions:

- mean value function $m_X(t) = E\{X(t)\}$, which for each fixed *t* provides the mean of the random variable $X(t)$;
- auto-covariance function $C_{XX}(t_1,t_2) = Cov\{X(t_1), X(t_2)\}$, which provides information concerning stochastic dependencies between each couple of random variables constituting the process, $X(t_1)$ and $X(t_2)$.

It is to note that for $t_1 = t_2 = t$, $C_{XX}(t,t) = Var\{X(t)\}$, which provides, for each fixed *t*, the variance of the random variable $X(t)$.

Another useful function, is autocorrelation $K_{XX}(t_1,t_2)$. It easy to show that the following equation holds:

$$K_{XX}(t_1,t_2) = E\{X(t_1),X(t_2)\} = C_{XX}(t_1,t_2) + m_X(t_1)m_X(t_2)$$

A stochastic process $\{X(t)|t \in T\}$ is said to be (strictly) stationary if for any $t_1,...,t_n$ and *s* in *T*, the random vectors $\{X(t_1),...,X(t_n)\}$ and $\{X(t_1+s),...,X(t_n+s)\}$ have the same distribution. This means that the joint statistics of X of all orders are unaffected by a shift in time. A stochastic process is wide sense stationary if $m_X(t) = m_X(t+s)$ and $C_{XX}(t_1,t_2) = C_{XX}(t_1+s,t_2+s)$, $\forall t,t_1,t_2,s \in T$. Strict stationarity always implies wide sense stationarity but not vice-versa.

If statistical averages associated to a stochastic process are asymptotically equal to time averages, the process is ergodic. Ergodicity can be formally expressed via the following equation:

$$E\{\phi[X(t)]\} = \lim_{T \longrightarrow \infty} \frac{1}{2T}\int_{-T}^{T} \phi[X(t)]dt$$

where $\phi(\cdot)$ is a real function.

If a stochastic process is wide sense stationary and ergodic, it is possible to obtain various realizations of the process “cutting” a single realization in as many parts. This is very important to obtain accurate estimates of the mean value and auto-covariance functions. However, it is not easy to test whether a stochastic process is wide sense stationary and ergodic or not. In addition, it is to underline that wide sense stationarity does not imply ergodicity.

2.3. Point Processes

A Point Process is a continuous-parameter stochastic process in which random variables obtained for each fixed *t*, $N(t)$, express the number of events (e.g. failures, customer arrivals at a bus stop) occurred up to time *t*. Of course $N(t)$ may only assume non-negative integer values and for each $t_2 > t_1$ is $N(t_2) \geq N(t_1)$, being $N(0) = 0$. The family of random variables $\{N(t), t \geq 0\}$ defines a counting process. Other interesting variables associated to a point process are, for example, the time to k^{th} event measured from time 0, T_k, and the time between the $(k-1)^{th}$ and k^{th} arrival, X_k. The above variables are linked with one another via the following equations:

$$X_k = T_k - T_{k-1}$$
$$T_k = \sum_{i=1}^{k} X_i$$

where $T_0 = 0$ and $k = 0,1,2,...$

In addition, it is easy to show that $\{N(t) \geq k\}$ and $\{T_k \leq t\}$ are equivalent events. Thus, a point process is fully specified, when the counting process $\{N(t), t \geq 0\}$ or all joint distributions of T_k or X_k are specified.

A first important function associated to a point process is the expected number of events $M(t) = E\{N(t)\}$. This function provides , for each fixed *t*, the expected value of the random variable *N*(*t*). This function can be expressed in different manners, based on the relationship between *N*(*t*) and T_k .In fact, being $G_k(t) = \Pr\{T_k \leq t\}$ it is easy to understand that $\Pr\{N(t) \geq k\} = G_k(t)$. In addition, $\Pr\{N(t) = k\} = \Pr\{N(t) \geq k\} - \Pr\{N(t) \geq k+1\}$. Thus, it follows that:

$$M(t) = \sum_{k=0}^{\infty} k \Pr\{N(t) = k\} = \sum_{k=0}^{\infty} k[G_k(t) - G_{k+1}(t)] = \sum_{k=1}^{\infty} G_k(t)$$

Another important function is the intensity function, $z(t)$, defined as follows:

$$z(t) = \lim_{\Delta t \longrightarrow 0} \frac{\Pr\{N(t, t+\Delta t) \geq 1\}}{\Delta t}$$

If the above limit exists, the quantity $z(t)\Delta t$ provides, for small values of Δt, the probability of observing at least one event in $]t, t+\Delta t]$. It is possible to show that if the point process is orderly (i.e. the probability of observing simultaneous events is equal to zero), the following equation holds:

$$z(t) = \frac{dM(t)}{dt} = \sum_{k=1}^{\infty} g_k(t)$$

being $g_k(t)$ the probability density function (pdf) of the time to the k^{th} event measured from time 0.

An interesting random variable associated to a point process is the forward recurrence time, W_t, defined as the time to the next event measured from an arbitrary point in time *t*.

2.3.1. Homogeneous Poisson Process

A Homogeneous Poisson Process is a continuous-time orderly counting process defined as follows:

$$\Pr\{N(a_i, b_i) = k_i; i = 1,\ldots,k\} = \prod_{i=1}^{k} \frac{[C(b_i - a_i)]^{n_i}}{n_i!} \exp[C(b_i - a_i)]$$

where C is a non-negative constant.

From the above definition, the following properties arise:

- The number of events $N(a_i, b_i)$ over a finite interval $(a_i, b_i]$ follows a Poisson distribution with mean $C(b_i - a_i)$;

- The number of events $N(a_i, b_i)$ $(i = 1,...,k)$ in non-overlapping intervals are stochastically independent random variables. This means that a Homogenous Poisson Process is characterized by independent increments;
- The probability distribution of the number of events counted in any time interval only depends on the length of the interval. This means that a Homogeneous Poisson Process is characterized by stationary increments;
- The intensity is constant and equal to *C*.

It is possible to show that the above process is orderly. Thus, the intensity function coincides with the first derivative of the mean value function:

$$M(t) = \int_0^t C dx = Ct$$

From the properties above, it follows that the time between two consecutive events are stochastically independent random variables; in addition they are exponentially distributed with parameter *C*. Thus, it follows that T_k is an Eralng random variable with scale parameter *C* and shape parameter *k*. Moreover, due to the fact that the Homogeneous Poisson Process is characterized by stationary and independent increments, the forward recurrence time is an exponential random variable with parameter *C*. So, for a Homogeneous Poisson Process the random variables X_i and W_t have the same distribution, which is independent on *t* and on the number of events occurred up to *t*; in other words, for this process, the memory-less property holds.

2.3.2. Non Homogeneous Poisson Process

A Non Homogeneous Poisson Process is a continuous-time orderly counting process defined as follows:

$$\Pr\{N(a_i,b_i)=k_i; i=1,...,k\}=\prod_{i=1}^{k}\frac{\left[\int_{a_i}^{b_i} z(t)dt\right]^{n_i}}{n_i!}\exp\left[\int_{a_i}^{b_i} z(t)dt\right]$$

From the above definition, the following properties arise:

- The number of events $N(a_i,b_i)$ over a finite interval $(a_i,b_i]$ follows a Poisson distribution with mean $\int_{a_i}^{b_i} z(t)dt$;
- The number of events $N(a_i,b_i)$ $(i=1,...,k)$ in non-overlapping intervals are stochastically independent random variables. This means that a Non Homogenous Poisson Process is characterized by independent increments;
- A Non Homogenous Poisson Process is characterized by non-stationary increments. In other words, the probability distribution of the number of events counted in any time interval, besides depending on the length of the interval, depends on the position of the lower limit of the considered interval.

In addition, the random variables X_k are not in general stochastically independent, neither identically distributed.

The mean value function of a Non Homogenous Poisson Process is:

$$M(t)=\int_0^t z(t)dx$$

The cumulative distribution function (cdf) of the forward recurrence time $K_t(w)$ can be obtained as follows:

$$K_t(w)=1-\Pr\{N(t,t+w)=0\}=1-\exp\left[-\int_t^{t+w} z(t)dt\right]$$

The expected value of the forward recurrence time, $E\{K_t(w)\}$, can be obtained as follows:

$$E\{K_t(w)\} = \int_0^{+\infty} [1 - K_t(w)] dw = \int_0^{+\infty} \exp\left[\int_t^{t+w} z(t)dt \right] dw$$

It can be noted that $K_t(w)$ and $E\{K_t(w)\}$ only depend on *t* and not on process history up to time *t*.

2.3.3. Renewal Process

A renewal process is a continuous-time orderly counting process in which times between events are positive, independent and identically distributed (IID) random variables.

Just like other point processes, a renewal process can be specified in three standard ways:

- Specifying the joint distributions of the arrival epochs T_1, T_2,…;
- Specifying the joint distributions of the times between events X_1, X_2,…;
- By the joint distributions of the counting random variables, $N(t)$ for $t > 0$.

The simplest characterization is through the times between events X_i, since they are IID.

Such a stochastic process is called renewal process because it probabilistically it starts over at each arrival epoch, T_i. The cdf of the time to the k^{th} event, T_k, can be expressed as follows:

$$\Pr\{T_k \le t\} = \Pr\{N(t) \ge k\} = F_X^{(k)}(t)$$

Where $F_X^{(k)}(t)$ is the k-fold convolution of the time between arrivals cdf.

The mean value function, $M(t)$, can be expressed as follows:

$$M(t) = E\{N(t)\} = \sum_{k=0}^{\infty} k\left[\Pr\{N(t) \geq k\} - \Pr\{N(t) \geq k+1\}\right] = \sum_{k=0}^{\infty} k\left[F_X^{(k)}(t) - F_X^{(k+1)}(t)\right] = \sum_{k=0}^{\infty} F_X^{(k)}(t)$$

The mean value function, in the context of renewal processes is also called renewal function. The first derivative of the mean value function with respect to t , $m(t)$, is called renewal density and it can be expressed as follows:

$$m(t) = \frac{dM(t)}{dt} = \sum_{k=0}^{\infty} f_X^{(k)}(t)$$

Where $f_X^{(k)}(t)$ is the k-fold convolution of the time between arrivals pdf.

An important relationship between Renewal function and renewal density is given by the renewal equation:

$$M(t) = F_X(t) + \int_0^t F_X(t-x) \cdot m(x) \cdot dx$$

In general, it is not easy to determine analytical expressions for $M(t)$ and $m(t)$. However, it is possible to show that as t tends to infinity, a renewal process tends to behave like a Homogeneous Poisson Process. Thus, $m(t)$ and $M(t)$ tend to be constant and linear respectively.

For some applications, it is interesting to determine the expression of the forward recurrence time (time to the next event measured from a specific point in time *t*) cdf, $F_W(w|t)$. The forward recurrence time cdf can be expressed as follows:

$$F_W(w|t) = F_X(t+w) - \int_0^t \left[1 - F_X(t+w-x)\right] \cdot m(x) \cdot dx$$

Where $F_X(\cdot)$ is the cdf of the time between events.

The above expression can be easily justified: the first term includes probability of all renewal realizations with at least a renewal point in $[0,t]$, whereas the second term includes probability of all realizations with at least a renewal point in $[0,t]$ and no renewal points in $[t,t+w]$. It is possible to show that as *t* tends to infinity, the forward recurrence time cdf associated to a Renewal Process, $F_W^{\infty}(w)$, can be expressed as follows:

$$F_W^{\infty}(w) = \frac{\int_0^w [1 - F_X(w)] \cdot dw}{E\{X\}}$$

2.4. Markov Processes

Let Ω be a finite or countable set. The stochastic process $\{X(t) | t \in T\}$ is a Markov process if the following property holds:

$$\Pr\{X(t) = x | X(t_n) = x_n, ..., X(t_0) = x_0\} = \Pr\{X(t) = x | X(t_n) = x_n\}$$

$$\forall\, t_0 < ..... < t_n < t;\ t_k \in T$$

The above property is known as memoryless property, since the state probability distribution at a given point in time *t* only depends on the current state $X(t_n)$.

In order to completely define a Markov Process it is necessary to define an initial-state probability vector $\pi(0)$ and transition probability functions $p_{ij}(v,t)$ over the interval [*0,t*] for each couple of states. The initial-state probability vector $\pi(0)$ is such that $\pi_j(0) = \Pr\{X(0) = j\} \quad j \in \Omega$, whereas transition probability functions are defined as follows:

$$p_{ij}(v,t) = \Pr\{X(t) = j | X(v) = i\} \quad i, j \in \Omega \quad 0 \le v \le t$$

Transition probability functions are characterized by the following main properties:

$$p_{ij}(t,t) = \begin{cases} 1 & if \quad i = j \\ 0 & otherwise \end{cases}$$

$$\sum_{j \in \Omega} p_{ij}(v,t) = 1 \quad \forall i \in \Omega \quad 0 \le v \le t$$

A Markov process is homogeneous if the following property holds:

$$p_{ij}(t) = p_{ij}(0,t) = \Pr\{X(t+v) = j | X(v) = i\} \;\; \forall t \in T \quad \forall v \ge 0$$

A Markov process is regular if only a finite number of state transitions can be observed in a finite time interval. If the state space Ω is finite, the Markov process is certainly regular.

When dealing with Markov processes, it is important to classify the process states. The states can be classified via the random variable $\tau(i) := \inf\{t > 0 : X_t = i\}$, which represents the time in which the process visits the state *i* for the first time for *t* > 0.

If $\Pr\{\tau(i) = \infty | X_0 = i\} > 0$, the state *i* is transient, whereas if $\Pr\{\tau(i) = \infty | X_0 = i\} = 0$ the state *i* is recurrent. Recurrent states are positive recurrent if $E\{\tau(i) | X_0 = i\} < \infty$ and null recurrent if $E\{\tau(i) | X_0 = i\} = \infty$. A state from which it is impossible to leave is called absorbing state.

2.4.1. Chapman-Kolmogorov equations and transition rate

The Chapman-Kolmogorov equation allows to express transition probability from the state *i* entered at time *v* to the state *j* at time *t*, visiting a generic state *k* at time *u*. The equation takes the following form:

$$p_{ij}(v,t) = \sum_{k \in \Omega} p_{ik}(v,u) p_{kj}(u,t) \qquad \forall i,j \in \Omega \quad 0 \le v < u < t$$

A Markov process cannot be easily treated via the above equation. Thus, the transition rate is introduced, which is involved in the Kolmogorov Differential Equations. Transition rate from a state *j*, $q_j(t)$, measures how quickly, at time *t*, the state *j* is left by the process and is formally defined as follows:

$$q_j(t) = -\frac{\partial p_{jj}(v,t)}{\partial t}\Big|_{v=t} = \lim_{h\to 0}\frac{p_{jj}(t,t) - p_{jj}(t,t+h)}{h} = \lim_{h\to 0}\frac{1 - p_{jj}(t,t+h)}{h}$$

Thus, it is possible to obtain:

$$p_{jj}(t,t+h) = 1 - q_j(t)h + o(h) \quad being \quad \lim_{h\to 0}\frac{o(h)}{h} = 0$$

In the same fashion, it is possible do define a transition rate from state *i* to a state *j*:

$$q_{ij}(t) = \frac{\partial p_{ij}(v,t)}{\partial t}\Big|_{v=t} = \lim_{h\to 0}\frac{p_{ij}(t,t+h) - p_{ij}(t,t)}{h} = \lim_{h\to 0}\frac{p_{ij}(t,t+h)}{h}$$

and hence:

$$p_{ij}(t,t+h) = q_{ij}(t)h + o(h) \quad being \ \lim_{h\to 0}\frac{o(h)}{h} = 0$$

2.4.2. Kolmogorov Differential Equation

The Chapman-Kolmogorov equation can be arranged as follows:

$$p_{ij}(v,t+h) = \sum_{k\in\Omega} p_{ik}(v,u)p_{kj}(u,t+h) \qquad 0 \le v < u < t$$

Thus, it is possible to write:

$$p_{ij}(v,t+h) - p_{ij}(v,t) = \sum_{k\in\Omega} p_{ik}(v,u)[p_{kj}(u,t+h) - p_{kj}(u,t)] \qquad 0 \le v < u < t$$

And hence:

$$\frac{p_{ij}(v,t+h) - p_{ij}(v,t)}{h} = \frac{\sum_{k\neq j} p_{ik}(v,u)[p_{kj}(u,t+h) - p_{kj}(u,t)]}{h} + p_{ij}(v,u)\frac{p_{jj}(u,t+h) - p_{jj}(u,t)}{h}$$

$$0 \le v < u < t$$

Taking the limits $h \to 0$ and $u \to t$ it is possible to obtain:

$$\frac{\partial p_{ij}(v,t)}{\partial t} = \sum_{k \neq j} p_{ik}(v,t) q_{kj}(t) - p_{ij}(v,t) q_j(t)$$

which represents the Kolmogorov differential equation. Each transition probability function associated to a Markov Process can be obtained solving the above equation.

If a Markov Process is stationary, the Kolmogorov differential equation can be written as follows:

$$\frac{\partial p_{ij}(t)}{\partial t} = \sum_{k \neq j} p_{ik}(t) q_{kj}(t) - p_{ij}(t) q_j(t)$$

The above expression can be also arranged in the following matrix form:

$$\dot{P} = P(t)Q$$

Where:

- P is the transition probability matrix;
- Q is the infinitesimal generator of the Markov process. The generic element $Q(i,j)$ of Q is q_{ij}, whereas $Q(i,i)$ coincides with - q_i.

The infinitesimal generator is characterized by the following properties:

$$q_{ij} > 0 \quad if \quad i \neq j$$
$$q_{ij} < 0 \quad if \quad i = j$$
$$\sum_{j \in \Omega} q_{ij} = 0 \quad \forall i \in \Omega$$

The Kolmogorov differential equation for homogeneous and regular Markov processes admits the following unique solution:

$$P(t) = e^{Qt} = I + Qt + \ldots.. + \frac{(Qt)^k}{k!} + \ldots.$$

In general, it is not easy to obtain an analytical expression of $P(t)$. An analytical expression of $P(t)$ can be easily obtained when the cardinality of Ω is small, using eigenvectors and eigenvalues of Q. The simplest situation can be encountered when Q has distinct eigenvalues. In this case, Q is also characterized by distinct eigenvectors and it can be easily diagonalized:

$$Q = X^{-1} D X$$

Thus:

$$P(t) = e^{Qt} = X^{-1} e^{Dt} X$$

The analytical expression of $\pi(t)$ can be obtained as $\pi(t) = \pi(0)P(t)$.

Using the Kolmogorov differential equation, it is also possible to characterize the distribution of the sojourn time of the process in a given state. Such a task, can be easily performed considering a two state Markov process characterized by the following infinitesimal generator:

$$Q = \begin{bmatrix} -\lambda & \lambda \\ 0 & 0 \end{bmatrix}$$

for which the second state is absorbing.

Solving the Kolmogorov differential equation, the following transition probability matrix can be obtained:

$$P(t) = \begin{bmatrix} e^{-\lambda t} & 1 - e^{-\lambda t} \\ 0 & 1 \end{bmatrix}$$

From the above expression, it is easy to note that the sojourn time of the process in the state 1 is exponentially distributed. In general, the sojourn time of a Markov process in a given state is an exponentially distributed random variable. However, the sojourn time of the process in a set of states is not exponentially distributed.

2.4.3. Subordinated Process, Embedded Markov chain and main theorems

In order to define the subordinated process and the embedded Markov chain, it is useful to introduce the concept of Markov chain. Basically, a Markov chain represents a particular case of Markov process, for which the parameter *t* is discrete. Just like for a Markov Process, a Markov chain is homogeneous if the transition probability matrix *P* does not depend on *t*.

Starting from a Markov chain with transition probability matrix *R*, it is possible to obtain a Markov process generating transitions via a Homogeneous Poisson Process with parameter η:

$$p_{ij}(t) = \sum_{k=0}^{\infty} \frac{(\eta t)^k}{k!} e^{-\eta t} R^k(i,j) = e^{-\eta t} e^{\eta t R(i,j)} = e^{\eta t(-1+R(i,j))}$$

Such a Markov process is called subordinated process and the transition probability matrix is:

$$P(t) = e^{\eta t(-I+R)}$$

Thus, $Q(i,j) = \eta R(i,j),\ Q(i,i) = \eta(R(i,i)-1)$. The infinitesimal generator Q depends on η, which has to be such that $\eta \geq \max_i \{-q_{ii}\}$. In fact, *R* is a stochastic matrix and

$R = I + \frac{Q}{\eta}$.

Starting from a Markov process, the embedded Markov chain can be obtained sampling the process at each point in time in which a transition occurs. The transition matrix of the embedded Markov chain can be characterized as follows:

$$R(i,j) = \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} = -\frac{q_{ij}}{q_{ii}} \qquad i \neq j$$

The embedded Markov chain is very important since the states of a Markov process can be classified on the basis of the embedded Markov chain: a given state of a Markov process is transient/recurrent if it is transient/recurrent for the embedded Markov chain. In addition, a Markov process is irreducible if the embedded Markov chain is irreducible (basically irreducibility means that it is possible to get any state from any state).

In addition, the following theorems hold for the stationary probability vector:

- **Theorem 1**: If a Markov process is irreducible and recurrent, there exists $P^* := \lim_{t \to \infty} P(t)$, $P^*(i,j) = \overline{\pi}(j) \quad \forall i \in \Omega$ and $\overline{\pi}$ is the stationary probability vector. In addition, two only situations can occur: $\overline{\pi} = 0$ or $\sum_{j \in \Omega} \overline{\pi}(j) = 1$. In the former case all states are null states, in the latter case all states are recurrent states.
- **Theorem 2**: If a Markov process is recurrent with infinitesimal generator *Q*, the stationary probability vector is the unique solution of the equation $\overline{\pi} Q = 0$

2.5. Non Markovian Processes

For Markov processes, the probability that the process enters a given state at time *t* only depends on the current state. When other forms of dependency affect the process evolution, the stochastic process is non markovian. Non markovian stochastic processes can be further classified on the basis of the form of stochastic dependence affecting process evolution. Before classifying non Markovian processes, some definitions are needed.

2.5.1. Markov renewal sequence and Markov renewal process

The sequence $\{(Y_n, S_n), n \geq 0\}$ is a Markov renewal sequence with state space *I*, if the following property holds:

$$\Pr\left\{Y_{n+1} = j \cap S_{n+1} - S_n \le x \middle| Y_n = i, S_n, ..., Y_0, S_0\right\} = \Pr\left\{Y_{n+1} = j \cap S_{n+1} - S_n \le x \middle| Y_n = i\right\} =$$
$$= \left\{Y_1 = j \cap S_1 \le x \middle| Y_0 = i\right\} \quad \forall n \ge 0, \forall i, j \in I$$

The vector-valued stochastic process $\boldsymbol{N}(t) = (N_j(t),\ j \in I)$ is defined as a Markov renewal process, where:

$$N_j(t) = \sum_{n=1}^{N(t)} Z_j(n)$$
$$z_j(t) = \begin{cases} 1 & if \quad Y_n = j \\ 0 & otherwise \end{cases}$$
$$N(t) = \sup\left\{n \ge 0 : S_n \le t\right\}$$

$N_j(t)$ is the number of times state j is visited by time t, $N(t)$ is total number of state changes by time t. For Markov renewal processes, future evolution only depends on the current state of the process at Markov renewal points S_k, that is process evolution depends only on the current state solely at specific time epochs.

2.5.2. Markov regenerative process and main non-Markovian processes

A stochastic process $\left\{Z(t) \middle| t \ge 0\right\}$ is a Markov regenerative process if there exists a Markov renewal sequence $\left\{(Y_n, S_n), n \ge 0\right\}$ of random variables such that all conditional distributions of $\left\{Z(S_n + t) \middle| Z(u) \cap Y_n = i, 0 < u < S_n\right\}$ coincide with $\left\{Z(S_n + t) \middle| Y_n = i\right\}$ and $\left\{Z(t) \middle| Y_0 = i\right\}$.

Thus, at each S_n the process evolution does not depend on the history before it. For a Markov regenerative process, the stochastic process between two consecutive renewal points could be any continuous time stochastic process. In other words different local behaviours are allowed between two consecutive Markov regenerative points. Thus, this family of stochastic processes allow to define a wide class of non Markovian processes.

For example, a Semi-Markov process is a Markov regenerative process such that no state changes can occur between two consecutive Markov regenerative points. Such a process can be viewed as a Markov regenerative process for which the length of time between two Markov regenerative points depends on the current state and on the state to be entered next.

Differently from Semi-Markov processes, for other Markov regenerative processes state changes are allowed between two consecutive Markov regenerative points. As an example, this is the case of Semi-regenerative stochastic processes, for which the stochastic process between two Markov renewal points is a Markov process.

In order to deal with Markov regenerative processes, it is important to define the kernel *K*(*t*) and the local kernel *E*(*t*).

Each element $K_{ij}(x)$ of the kernel $K(x)$ is defined as follows:

$$K_{ij}(x) = \Pr\left\{Y_{n+1} = j \cap S_{n+1} - S_n \le x \middle| Y_n = i\right\} = \Pr\left\{Y_1 = j \cap S_1 \le t \middle| Y_0 = i\right\}$$

It is possible to note that $\{Y_n, n \ge 0\}$ is a Markov chain with transition probability matrix of $K_{ij}(\infty)$. Such a chain is called embedded Markov chain of the Markov regenerative process.

Each element $E_{ij}(x)$ of the kernel $E(x)$ is defined as follows:

$$E_{ij}(x) = \left\{M(x) = j \cap S_{n+1} - S_n > x \middle| Y_n = i\right\} = \Pr\left\{M(x) = j \cap S_1 \le x \middle| Y_0 = i\right\}$$

It is possible to show that the transition probability matrix of a Markov regenerative process *P*(*t*) satisfies the following generalized Markov renewal equation:

$$P(t) = E(t) + K * P(t)$$
$$K_{iu} * P_{uj}(t) = \int_0^t P_{uj}(t-x)\, dK_{iu}(x)$$

This is the general form of the equation that allows to obtain the transition probability matrix. It can be further particularized for each specific non markovian process and the corresponding numerical solution can be more or less complicated. In some specific cases, to be addressed in the following section, it is possible to obtain the state probability vector of a non markovian process without losing numerical and analytical tractability of Markov processes.

2.5.3. Modelling non exponential sojourn times in a Markovian setting: Phase-Type distributions

It has been pointed out that the sojourn time of a Markov process in a given state is an exponentially distributed random variable, whereas the sojourn time of the process in a set of states is not. Thus, when dealing with non markovian processes, each process state for which the sojourn time is not exponential can be modeled by a proper arrangement of multiple “stages” for which the sojourn time is exponential. By this way, a non markovian process can be treated without losing numerical and mathematical tractability of Markov processes. This is the basic idea behind the Device of Stages technique. In general, the application of this technique leads to approximate results. Nevertheless, any desired degree of accuracy can be obtained, based on the following theoretical result (Schassberger, 1973):

$$F(s) = \lim_{n \longrightarrow \infty} \sum_{k=1}^{n} p_{k,n} \left(\frac{\lambda_{k,n}}{\lambda_{k,n} + s} \right)^{r_{k,n}} \quad p_{k,n} > 0, \lambda_{k,n} \in (0,\infty), r_{k,n} \in \mathbb{N}$$

where $F(s)$ is the Laplace-Stieltjes transform of a non negative random variable. In other words, each non negative random variable is the weak limit of mixtures of Erlang distributions.

From a practical standpoint, the application of this technique is not exempt from drawbacks. As an example, each state with a non exponential sojourn time should be modeled by means of a reasonably low number of stages, in order to control the dimension of the state space associated to the underlying Markov process. Also the number of parameters associated to each arrangement of stages should be controlled, in order to avoid numerical problems when parameters have to be determined. In addition, each arrangement of stages should be characterized by a reasonably simple structure, so that the underlying Markov model can be easily generated.

Before dealing with the problem of approximating a non exponential random variable via an arrangement of exponential stages, it is worthy to formally characterize the sojourn time in a set of exponential stages. Such a random variable is Phase-Type distributed.

Given a Markov process with *m+1* states, such that states *1,…,m* are transient and the state *m+1* is an absorbing state, a Phase-Type distribution is the distribution of time from the above process's starting until absorption in the absorbing state. The infinitesimal generator of the considered Markov process, is characterized by the structure reported below:

$$Q = \begin{bmatrix} S & S^0 \\ \overline{0} & 0 \end{bmatrix}$$

where $S \in \Re^{m \times m}$, $S^0 = -S\overline{u}$, $\overline{u} = [1...1]^T$, $\overline{u} \in \Re^m$ and $\overline{0} \in \Re^{1xm}$.

The initial probability vector is $(\overline{\alpha}, \alpha_{m+1})$, being $\overline{\alpha} \in \Re^m$. The Phase-Type distribution cdf is:

$$F_T(t) = 1 - \overline{\alpha} \cdot \exp(S \cdot t) \cdot \overline{u}$$

where $\overline{\alpha} \cdot \exp(S \cdot t) \cdot \overline{u}$ is the probability that the process only visited the *m* transient states by time *t*.

The Phase-Type distribution pdf is:

$$f_T(t) = -\overline{\alpha} \cdot \exp(S \cdot t) \cdot S \cdot \overline{u}$$

A first example of Phase-Type distribution is the Erlang with parameters *r*, *λ* for which, given *r=3*, we obtain:

$$S = \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ 0 & 0 & -\lambda \end{bmatrix} \quad \overline{\alpha} = [1 \quad 0 \quad 0]$$

Also mixtures of Erlang random variables are Phase-Type distributions. For example, for a mixture of Erlang random variables with $r_1 = r_2 = 2$ it is possible to write:

$$S = \begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & 0 \\ 0 & -\lambda_1 & 0 & 0 \\ 0 & 0 & -\lambda_2 & \lambda_2 \\ 0 & 0 & 0 & -\lambda_2 \end{bmatrix} \quad \overline{\alpha} = [\alpha_1 \quad 0 \quad 1-\alpha_1 \quad 0]$$

As previously highlighted, it is possible to approximate an arbitrarily distributed sojourn time via a Phase-Type distribution. The choice of a proper Phase-Type distribution depends on the random variable to be approximated. For example, a Weibull distribution with an increasing hazard rate function can be approximated via an Erlang random variable, whereas a Weibull distribution with a decreasing hazard rate function can be approximated by means of an Hyperexponential distribution (i.e. a mixture of exponential random variables) (Bobbio, Cumani, 1983). Several techniques can be used to determine the parameters of the Phase-Type distribution. Among these, the simplest one is the moment matching technique. This technique consists in determining the desired parameters equating the moments of the Phase type distribution with the moments of the random variable to be approximated. Obviously, the number of equations is equal to the number of parameters to be determined. In addition, the random variable to be approximated shall have a number of finite moments equal to the number of parameters to be determined. When a Weibull with an increasing hazard rate function has to be approximated, the following system of equations shall be solved to determine the parameters of the Erlang random variable:

$$\begin{cases} \dfrac{n}{\lambda} = \alpha \cdot \Gamma\left(\dfrac{\beta+1}{\beta}\right) = m_1 \\ \dfrac{n \cdot (n+1)}{\lambda} = \alpha^2 \cdot \Gamma\left(\dfrac{\beta+2}{\beta}\right) = m_2 \end{cases}$$

The number of stages n can be obtained rounding to the nearest integer the number $\dfrac{m_1^2}{m_2 - m_1^2}$, whereas λ can be obtained solving the first equation. As an example, in the following figure is reported a Weibull pdf with parameters $\alpha = 1, \beta = 1.5$ and the approximating pdf of an Erlang random variable with parameters $n = 2, \lambda = 2.22$.

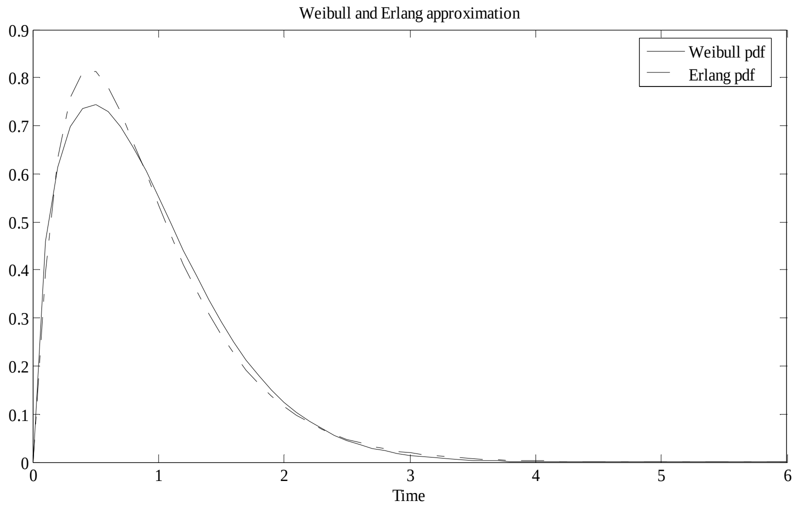


Figure 2-1: Weibull pdf $\alpha = 1, \beta = 1.5$ (solid line) and Erlang approximation pdf $n = 2, \lambda = 2.22$ (dash-dot line)

When a Log-normal distribution has to be approximated, mixtures of Erlang random variables with the same shape parameter *n* can be used (Johnson, Taaffe, 1989). Such an approach, if *r* is the number of Erlang random variables in the mixture, leads to solve a system of *2r* equations: *r-1* “branching” probabilities, *r* scale parameters and the common shape parameter *n* have to be determined. However, when the standard deviation of the Log-normal random variable is low when compared to the mean, the use of a single Erlang distribution might be satisfactory. As an example, in the following figure is depicted a Log-normal pdf with parameters $\mu = 1, \sigma = 0.3$ (solid line) and the corresponding Erlang approximation with parameters $n = 11, \lambda = 3.87$ (dash-dot line).

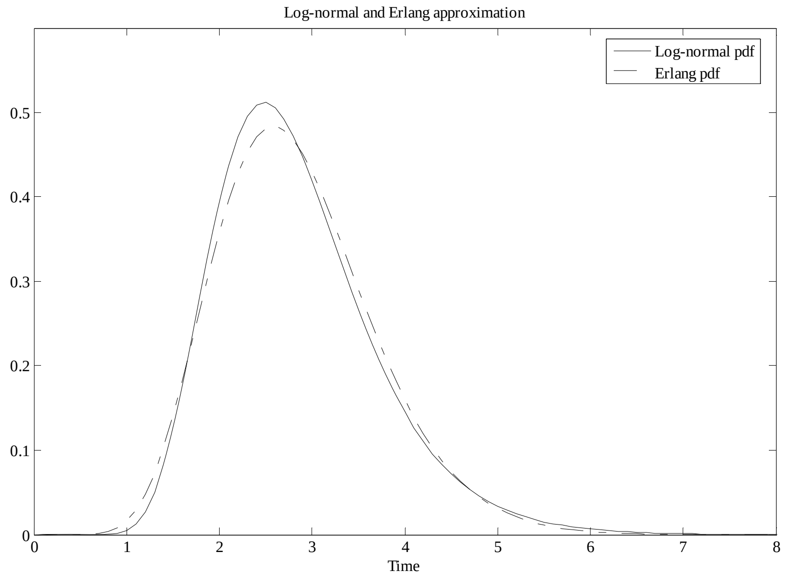


Figure 2-2: Log-normal pdf $\mu = 1, \sigma = 0.3$ (solid line) and Erlang approximation pdf $n = 11, \lambda = 3.87$ (dash-dot line)

It has to be noted that an Erlang distribution can also be used to approximate deterministic distributions, since the Erlang distribution is the Phase-Type distribution with the minimum variance (Aldous, Shepp, 1987).

3 QUALITY EVALUATION OF A MASS TRANSPORT SERVICE

Sometimes, a measure of the service quality is substantially suggested by the market. Often, a suggested and effective service quality measure is the price that customers are willing to pay. However, this approach cannot be adopted when prices are established on the basis of political considerations, as in the case of an urban mass transport service. In this instance, quality evaluations have to be performed focusing on the service characteristics which customers are more concerned about. Among these, the main characteristic is the time spent by customers at the stop points waiting for the transport mean (e.g. a bus).

With special reference to this important characteristic, in this chapter a general methodology is proposed which can be used to measure the quality of a mass transport service with short headways, when the time schedule is unknown to customers. The proposed approach has been tested using a real data set of bus arrival times from the route 181 of the Azienda Napoletana Mobilità S.p.A, the neapolitan mobility company.

3.1. Introduction

High quality service offered to passengers by a urban mass transport company is the main lever for encouraging people to use mass transit instead than private cars. Appropriate quality indexes have to be adopted to measure the offered quality level, which shall be continuously monitored and, when possible, improved by means of adequate policies and/or operational choices.

In many cases, a convincing service quality measure is the price customers are willing to pay for availing it. Unfortunately, this kind of mechanism doesn’t hold for urban mass transport services, since prices are established on the basis of political considerations, to promote the public transport usage. A quality characteristic of a mass transport service which customers are very concerned about is the service punctuality (Van Hagen et al., 2007). Unfortunately, an urban mass transport service is normally characterized by short headways that prevent customers to rely on a time table. For such a transport service, being the concept of punctuality not applicable, the focus shall move to the customers waiting time, which distribution strongly depends on scheduled frequency of passage (headway) and service regularity (i.e. Inspection Paradox) (Stein, Dattero, 1985). In this experimental context, specific quality indexes, based on the customer waiting time, are needed. The indexes proposed in this chapter exactly possess this characteristic. They have an unambiguous operative meaning and are easy to calculate. The only little obstacle to their use consists in the fact that the waiting time distribution, needed to perform the considered analysis, is not simple to obtain, unless samples of waiting time data collected among targeted customers are available**.** Indeed, this kind of surveys is very costly and time consuming and, as a matter of fact, is rarely performed in practice. On the contrary, information and/or data regarding the arrival processes of both transport means and customers, at the stop points are very often available.

Thus, in order promote the use of the proposed indexes a methodology is presented that can be adopted to obtain the customers waiting time distribution from the this latter kind of data (i.e. bus arrivals at the stop and customers' behaviour).

The methodology consists in two steps:

- modelling the arrival processes of both customers and transport means on the basis of the available data/information and some non restrictive hypotheses;
- obtaining the stationary distribution of the Forward Recurrence Time (Cox,

Isham, 1980) of the bus arrival process and using it to model the customers waiting time.

It is shown that, under the hypotheses considered in this chapter, the adopted approach gives a very good approximation for the exact customers waiting time distribution.

The proposed methodology is applied to the route 181 of the public bus transport service managed by A.N.M. (Azienda Napoletana Mobilità S.p.A, (www.anm.it)). The study is performed using a real data set of bus arrival times provided by A.N.M. All the main assumed hypotheses have been successfully tested on the basis of the available set of data.

3.2. Formulation Of The Waiting Time Distribution

As previously remarked, in this chapter the customers waiting time distribution is (indirectly) obtained combining models describing bus and customers arrival process at a specific bus stop of interest.

It is assumed that times between bus arrivals (i.e., the Headways, *H*) are identically distributed random variables, which cumulative density function is indicated as $F_H(\cdot)$. Moreover, it is assumed that the customers arrival process at the stop of interest is a Homogenous Poisson Process stochastically independent on the bus arrival process.

As additional hypothesis, for sake of simplicity, it is assumed that buses have infinite load capacity (i.e. customers waiting time ends when the first bus arrives).

Under the hypotheses given above, the customer waiting time cumulative distribution function (CdF) can be formulated as follows:

$$F_W(w) = \int_0^{\tau} F_W(w|t)\, f_T(t)dt \qquad (3.1)$$

Where:

- $[0,\tau]$ is the length of the time horizon on which the proposed method is applied;
- $F_W(w|t)$ is the forward recurrence time distribution from the time *t*, which coincides with the waiting time distribution of a customer that reaches the bus stop at time *t*;
- $f_T(t)$ is the probability density function (*pdf*) of the arrival time at the bus stop of the generic customer (i.e. a customer randomly chosen among those that reach the stop in the interval $[0,\tau]$);
- $F_w(w)$ is the (unconditional) forward recurrence time CDF, which coincides with the (unconditional) waiting time distribution of the generic customer.

In particular, assumed that at least one customer reaches the stop in the time interval $[0,\tau]$, as the arrival Process is Homogeneous Poisson, $f_T(t)$ is a uniform $[0,\tau]$ random variable (i.e. $f_T(t)=1/\tau$) (Ross, 1996).

Unfortunately, as previously mentioned, Equation (3.1), is not easy to use, unless customers waiting times are directly observed.

As an instance, if even one assumes that sequences of successive headways constitute a Renewal Process (i.e. if the additional hypothesis that headways are *s*-independent is made), in order to obtain the forward recurrence time distribution from the time *t,* $F_W(w|t)$, it would be necessary to solve the following equation (Cox, 1967):

$$F_W(w|t)=F_H(t+w)-\int_0^t \left[1-F_H(t+w-u)\right]m(u)\,du \qquad (3.2)$$

which is not an easy task, since the Renewal Density, $m(u)$, is often unavailable in closed form (Arnold, Groeveld, 1981).

Anyway, a nice approximation for $F_w(w)$, for large τ, is represented by the following stationary forward recurrence time distribution, $F_W^\infty(w)$ (Cox, Isham, 1980), (Daley, Vere-Jones, 2002):

$$\begin{aligned} F_W^\infty(w) &= \lim_{\tau\to+\infty}\int_0^\tau F_W(w|t)\, f_X(t)dt = \lim_{\tau\to+\infty}\frac{1}{\tau}\int_0^\tau F_W(w|t)dt = \\ &= \frac{1}{E\{H\}}\cdot\int_0^w (1-F_H(h))\,dh \end{aligned} \tag{3.3}$$

where $E\{H\}$ is the mean headway.

On the basis of the *CDF* (3.3) the following value is obtained for the mean customer waiting time $E\{W\}$:

$$E\{W\} = \frac{E\{H\}}{2} + \frac{Var\{H\}}{2E\{H\}} \tag{3.4}$$

a result that is well known to experts involved in transit service planning (Ceder, 2007), (Osuna, Newell, 1972). From (3.4) follows that the mean customer waiting time increases with the headway variance. As to say: the generic customer that reaches the bus stop at time *t*, randomly chosen in $[0,\tau]$, picks an interval between bus arrivals that is larger, in mean, than $E(H)$. In fact, long intervals are more likely to be selected than short ones.

3.3. Service Quality Indexes

In this section the proposed service quality indexes are introduced. The first one, the well known System Service Dependability (SSD) (Heimann, 1972), (Fielding, 1979), (Silcock, 1981), (Erto et al., 1995) is here proposed in a customized version which allows its use in the considered experimental context. The other two indexes, the Density of Hope of End Waiting (DHEW) and Mean Residual Waiting Time (MRWT), are reformulated in terms of waiting time. The intrinsic meaning of all the considered indexes is discussed in details. Moreover the diagnostic capabilities of the indexes are pointed out via a simple sensitivity analysis.

3.3.1. SYSTEM SERVICE DEPENDABILITY

In the field of mass transport services, widespread measures of service quality are based on the ratio of the number of successes over a total number of trials, called Dependability (Erto et al., 1995), where the term success is specifically defined for each specific application.

In several studies regarding Bus and Railway Systems, the System Service Dependability (SSD) has been defined as the ratio of the number of customers incurring in a delay not higher than a tolerable value, say *d*, to the total number of served customers (Heimann, 1972), (Silcock, 1981). For Airline Services, a System Dependability measure is given by the Dispatch Reliability, defined as the ratio of the number of flights that depart without a cancelation or a delay higher than a tolerable value to the number of scheduled flights (Fielding, 1979).

In this study the success is defined as the occurrence of a waiting time not greater than a tolerable value, say w_T. So, the Dependability is defined as follows:

$$SSD = F_W\left(w_T\right) = \Pr\left\{W \le w_T\right\} \tag{3.5}$$

3.3.2. DENSITY OF HOPE OF END WAITING

The Density of Hope of End Waiting (DHEW) is defined as:

$$DHEW = \lim_{\Delta w \to 0} \frac{HEW(w, \Delta w)}{\Delta w} \tag{3.6}$$

where the *HEW* (namely, the Hope of end Waiting) is given by:

$$HEW(w, \Delta w) = \frac{F_W(w + \Delta w) - F_W(w)}{1 - F_W(w)}$$

The DHEW fits the concept of Density of Hope of End Delay (Erto et al., 1995) to the case of absence of time table considered in this chapter.

The HEW depends on two variables, w and Δw. The HEW gives, for each *w*, the probability that the Waiting time is less than or equal to $w + \Delta w$ given that the time already spent at the stop is *w.*

Since the Density of Hope of End Waiting is obtained via equation (3.6) it is a function of w only. It is the analogous, in terms of waiting time, of the Hazard Rate function, used in Reliability. The DHEW provides useful diagnostic information about the transport service under study, in fact:

- An increasing DHEW indicates that the transport service is robust with respect to the causes of disorder, since the probability that the waiting ends increases as the already accumulated waiting time increases;
- A constant DHEW indicates that the transport service is indifferent with respect to the causes of disorder, since the probability that the waiting ends doesn't depend on the waiting time already accumulated at the bus stop;
- A decreasing DHEW indicates that the transport service is not able to react to the causes of disorder (i.e. it is weak), since the probability that the waiting ends decreases as the waiting time increases.

3.3.3. MEAN RESIDUAL WAITING TIME

The Mean Residual Waiting Time (MRWT) is defined as follows :

$$MRWT(w) = E\left\{W - w \middle| W > w\right\} = \frac{\int_{w}^{\infty}\left(1 - F_W\left(x\right)\right)dx}{1 - F_W\left(w\right)}$$

It is inspired to the concept of Mean Residual Life, well known in Reliability. The Mean Residual Waiting Time represents the mean of the residual waiting time, $W - w$, calculated under the hypotesis that the waiting time already accumulated at the bus stop is larger than w. The MRWT is proportional to the inverse of the DHEW, in fact:

- An increasing MRWT indicates that the expected residual waiting time increases as the waiting time increases;
- A constant MRWT indicates the expected residual waiting time doesn't depend on the waiting time already accumulated at the bus stop;
- A decreasing MRWT indicates the expected residual waiting time decreases as the waiting time increases.

3.3.4. SENSITIVITY OF THE INDEXES TO THE HEADWAY VARIANCE

Figure 3-1 shows results obtained for the proposed indexes for different values of the Headway Variance, which is used to determine different service regularity. In the performed analysis the Headway is modelled as a Gamma variable, which use is motivated in section 3.4.1. A mean Headway of 12 minutes is assumed. The values $\alpha = 1, 2, 5$ are respectively adopted for the shape parameter, in order to specify different Headway variances (the higher α the smaller the headway variance). Moreover the limiting case of a perfectly regular transport service is considered for which the headway is exactly equal to 12 minutes.

Figure 3-1-a depicts the waiting time CdF $F_W(\cdot)$, strictly linked to SSD via equation (3.5). As service regularity increases (i.e. variance of *H* decreases), the SSD increases. For $\alpha = 1$ the headway is an exponential random variable, thus, because of the memoryless property, the identical exponential distribution is also obtained for the waiting time.

Figures 3-1-b and 3-1-c show the DHEW and the MRWT computed on the basis of the Waiting Time Distribution (3.3). Of course, for $\alpha = 1$, being the waiting time Exponentially distributed, both the DHEW and the MRWT are constant. Obviously the DHEW increases and the MRWT decreases with service regularity.

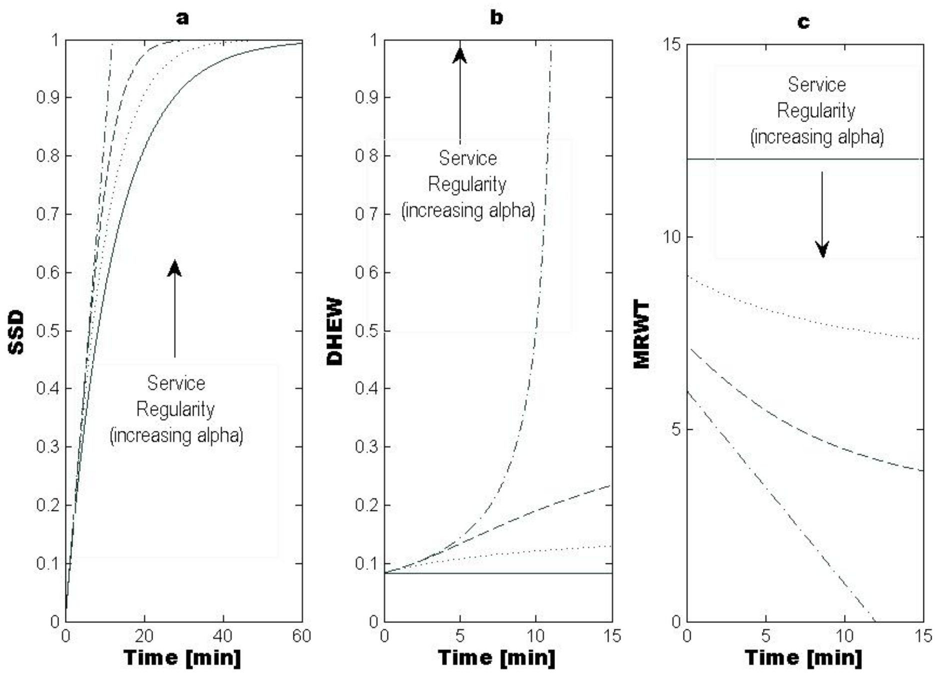


Figure 3-1: Service Quality Indexes in different regularity conditions

3.4. A first Case Study: Quality evaluation of the route 181 operated by A.N.M.

The service quality indexes introduced in section 3.3 have been used to analyse the route 181 of Azienda Napoletana Mobilità S.p.A. The route is characterized by 55 bus stops distributed along a route of 13.9 km, that connects the "Campi Flegrei" railway station to "Piazza Medaglie d'oro" and vice-versa. Arrival Times of buses at main stops, maxi-nodes, are collected in a Data Recording System that allows monitoring in real time the service performances. For this exemplificative application, attention has been focused on the Via Caravaggio stop. A real data set (see Attachments to chapter 3) of bus arrival times is used to check that a Gamma RP fits the bus arrival process. The same data set was used to estimate the Gamma distribution parameters.

No data about the customers arrival process are available. Thus, the hypothesis that the customers arrival process is an Homogenous Poisson Process stochastically independent on the bus arrival process, was not checked. Anyway, it is to note that when a mass transport service is characterized by quite short headways, as in the considered application, this hypothesis seems to be very convincing (Kho et al., 2005).

The section is structured as follows. In section 3.4.1 appropriate statistical tests of hypotheses are adopted to check the bus arrival process assumptions. In section 3.4.2 the accuracy of the approximation adopted to model the waiting time distribution is discussed in some details. At last, in section 3.4.3 results of the performed analysis are presented.

3.4.1. Statistical Hypothesis checking

The Gamma model for time between bus arrivals (Larson, Odoni, 1981), (Billi et al., 2003) has been considered :

$$f_H(h) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} h^{\alpha-1} e^{-\frac{h}{\beta}} \tag{3.9}$$

For parameter estimation purposes the mid-point imputation approach is adopted, which for our one minute precision data (see Attachments to chapter 3) and an estimated median headway of about 11 minutes, doesn't affect the validity of the performed analyses (Law, Brookmeyer, 1992).

The hypothesis of Renewal Process has been checked via the Generalised Anderson Darling (GAD) Test, which presents good power against both monotonic and non-monotonic alternatives (Kvaloy et al., 2001).

The available sequences of headways have been treated as a unique realization of a Renewal Process in order to check both the presence of day by day (i.e. inter day) and headway by headway (i.e. intraday) heterogeneities.

The GAD score 0.97 is obtained which, at 5% level of significance, doesn't give evidence against the null hypothesis (the 5% GAD test rejection limit is 2.49).

Goodness of fit of the Gamma model has been checked via a modification of the Locke's test (Locke, 1976) based on the consistent BKR test of independence (Lukacs, 1995), (Blum et al., 1961), (Wilding, Govind, 2008).

The randomized 98 pairs (one observation has been randomly discarded) used to perform the test are reported in Appendix B. The 5% rejection limit for the BKR test has been calculated adopting the Gaussian "portable" approximation of the BKR statistic given in (Mudholkar, Wilding, 2005). For the considered number of pairs the rejection limit 4.62 has been obtained. The BKR statistic 3.10 has been computed on the basis of the available data. Thus, the considered sets of pairs gives no evidence against the null hypothesis of Gamma distributed headways.

Table 3-1 reports the Maximum Likelihood Estimates of the shape, α, and scale, β, parameters of the Gamma distribution. The Maximum Likelihood Estimates of the Mean and Variance are also reported:

Headway Model Parameters		Headway Mean and Standard Deviation [min]	
α	4.14	Mean	12.96
β	3.13	Standard Deviation	6.37

Table 3-1: Headway model parameters and elementary statistics

3.4.2. Checking Approximation Accuracy

In this chapter the limiting distribution (3.3) is used, for a finite τ , to approximate the exact customer waiting time distribution (3.1). In order to check the approximation accuracy a simple approach based on the Phase-Type Renewal Processes theory (Neuts, 1978) is adopted, which enables to avoid searching solutions for equations (3.1) and (3.2), that are quite cumbersome to solve.

Note that the Gamma random variable isn't a PH one unless the shape parameter is an integer. Thus, a *Gamma*(5,β) distribution is considered for the headway. Moreover, the value $\beta = 2.59\ \text{min}^{-1}$ has been used for the scale parameter, in order to set the mean of Headways to the estimated value 12.96 (see Table 3-1).

For any given mean headway value and finite time τ , the approximation gets worse as the variance decreases. So the performed check, made considering a Gamma(5,2.59) RP, is conservative. The CTMC $\{X(t), t \geq 0\}$ associated to the Gamma(5,β) RP is characterized by the following generator Q^*:

$$Q^* = \begin{bmatrix} -\beta^{-1} & \beta^{-1} & 0 & 0 & 0 \\ 0 & -\beta^{-1} & \beta^{-1} & 0 & 0 \\ 0 & 0 & -\beta^{-1} & \beta^{-1} & 0 \\ 0 & 0 & 0 & -\beta^{-1} & \beta^{-1} \\ \beta^{-1} & 0 & 0 & 0 & -\beta^{-1} \end{bmatrix} \qquad (3.11)$$

where $X(t)$ may be viewed as the (transient) state of the Gamma (5,β) distribution visited by the renewal process at time *t*.

Following the main results given in (Kao, Smith, 1992) it is possible to show that that the forward recurrence time *CdF* at time *t*, $F(W|t)$, for this renewal process can be also formulated as:

$$F(W|t)=\sum_{i=1}^{5} v_i(t)\frac{\gamma(5+1-i, w/\beta)}{\Gamma(5+1-i)} \tag{3.12}$$

where:

- $v_i(t)$ $i=1,...,5$ are the state probabilities at time *t* associated to the CTMC $\{X(t), t\geq 0\}$. They can be obtained solving $v(t)=v(0)\cdot\exp(Q^* t)$ where $v(0)=[1 \quad 0 \quad 0 \quad 0 \quad 0]$;
- $\gamma(s,\cdot x)$ is the lower incomplete gamma function: $\int_0^x z^{s-1}\cdot e^{-z}dz, \quad s>0$ (Abramowitz, Stegun, 1972);
- $\Gamma(s)$ is the gamma function: $\int_0^\infty z^{s-1}\cdot e^{-z}dz, \quad s>0$.

Equation (3.12) can also be interpreted using the total probability theorem. In fact it is easy to recognize that if $X(t)=i$ the forward recurrence time is $\text{Gamma}(5-i+1,\beta)$ distributed, thus $F(W|t)$ can be obtained multiplying the forward recurrence time distributions conditioned on $X(t)$ by the correspondent state probabilities.

As stated in section 3.2, the arrival time of the generic customer at the stop of interest is Uniform $[0,\tau]$ distributed and is stochastically independent on the bus arrival process. So the customer waiting time *CdF* , $F_W(w)$, can be expressed as follows:

$$F\left(W\right)=\frac{1}{\tau}\int_{0}^{\tau}\sum_{i=1}^{5}v_i\left(t\right)\frac{\gamma\left(5+1-i,\,{}^{w}\!/_{\beta}\right)}{\Gamma\left(5+1-i\right)}dt \tag{3.13}$$

Figure 3-2 reports the stationary forward recurrence time *pdf* (dashed line) and the exact waiting time density obtained via equation (3.13) in the case of a τ = 3 hour long time interval (i.e. the length of the 7-10 AM daily observation interval).

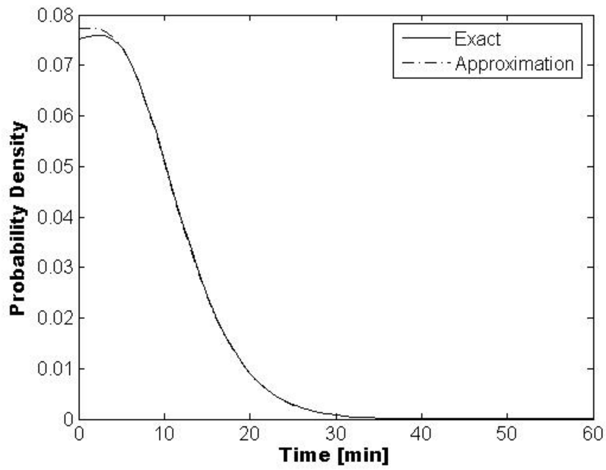


Figure 3-2: Comparison between the exact customer waiting time and forward recurrence time densities

Even for the selected, very small, τ value the stationary forward recurrence time distribution constitutes a very good approximation for the exact customer waiting time distribution. Obviously the larger is the time interval of interest, the better the approximation.

3.4.3. SERVICE QUALITY EVALUATION

In this paragraph, the limiting waiting time distribution (equation (3.3)) obtained using the estimated headway distribution, that is a Gamma(4.14, 3.13), is used to evaluate the service quality indexes introduced in Section 3.3. The resulting limiting waiting time distribution, which cannot be expressed in analytic form, is represented in Figure 3-3.

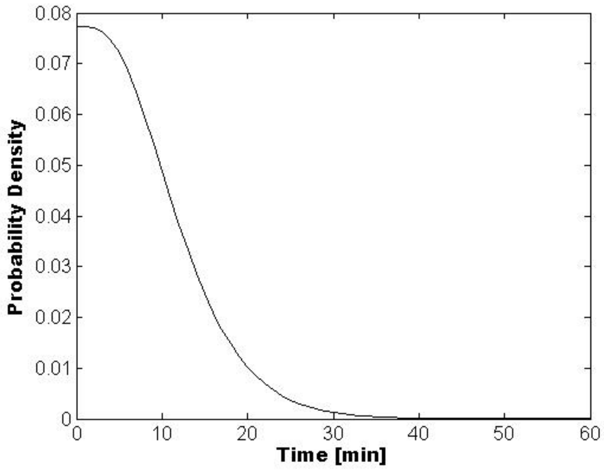


Figure 3-3: Estimated Waiting Time pdf

Results obtained for the indexes are represented in Figure 3-4. In the same figure, for comparison purposes, results obtained in the case of a perfectly regular transport service are represented, which are the best possible to attain with the assumed (i.e. estimated) mean headway value. A visual analysis of Figure 3-4 highlights that the DHEW is a concave increasing function (see figure 3-4-b); this means that the service is robust, but its robustness cannot be easily perceived by customers. In fact, for example, the MRWT dynamic (figure 3-4-c) shows that after 10 minutes already spent at the stop the expected residual waiting time is still about 5.5 minutes, whereas the initial mean waiting time for the generic customer is 8.04 minutes. Thus, after 10 minutes spent at the bus stop the mean residual waiting time decreases of (about) 2.5 minutes only, a result that may discourage customers to wait (consider that in the case of the perfectly regular transport service in figure 3-4-c after 10 minutes spent at the bus stop the mean the residual waiting time is 1.50 minutes).

Moreover, as an example, the SSD plot evidences that the generic customer waiting time exceeds 7.45 minutes (1.15 times the half headway) with probability 0.45 (i.e. the 45% of the customers wait more than 7.45 minutes) whereas in the case of a perfectly regular transport service the generic customer waiting time exceeds 7.45 minutes with probability 0.42. Thus not big differences are evidenced by this index if a small value is considered for the reference waiting time. On the contrary if the SSD is calculated with reference to a waiting time of 14 minutes, larger than the mean headway, a value of about 0.85 is obtained for the service under study, whereas the value 1 is obtained for the perfectly regular service.

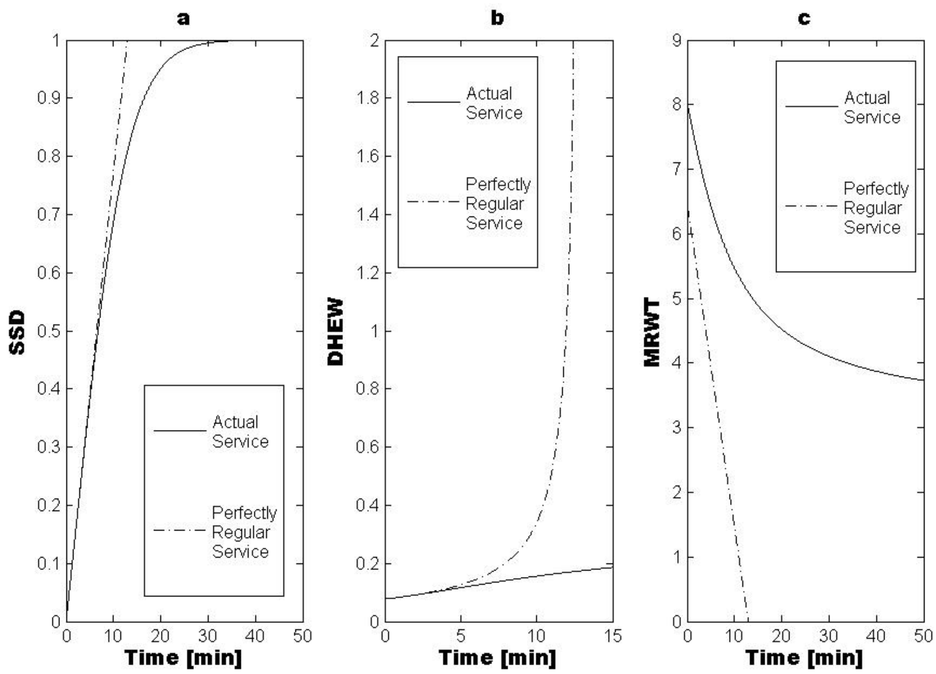


Figure 3-4: Service quality indexes for the route 181- Comparison between actual service (continuous line) and perfectly regular service (dash-dotted line)

Equation (3.4), as long as results in figures 3-1 and 3-4, gives evidence that reducing the headway variance allows to obtain noteworthy quality improvement. This led

Transport Authorities to develop, implement and test various service-regularizing strategies (Hounsell et al., 2008), (Panglinan et al., 2008). Besides improving service quality, reducing headway variance leads to reduce operating costs (i.e. vehicle operating hours per day, labor, energy and maintenance costs). For example, as shown via the comparison performed in figure 3-5, given the generic customer mean waiting time, if the headway standard deviation reduces from 6.37 to 5 minutes the system quality improves (in terms of both DHEW and MRWT) and (see equation 3.4) requested mean headway increases from 12.96 to 14.33 minutes. This apparently "small" change allows to save about 340 runs per year, only considering the working days and the time interval 7.00-10.00 am.

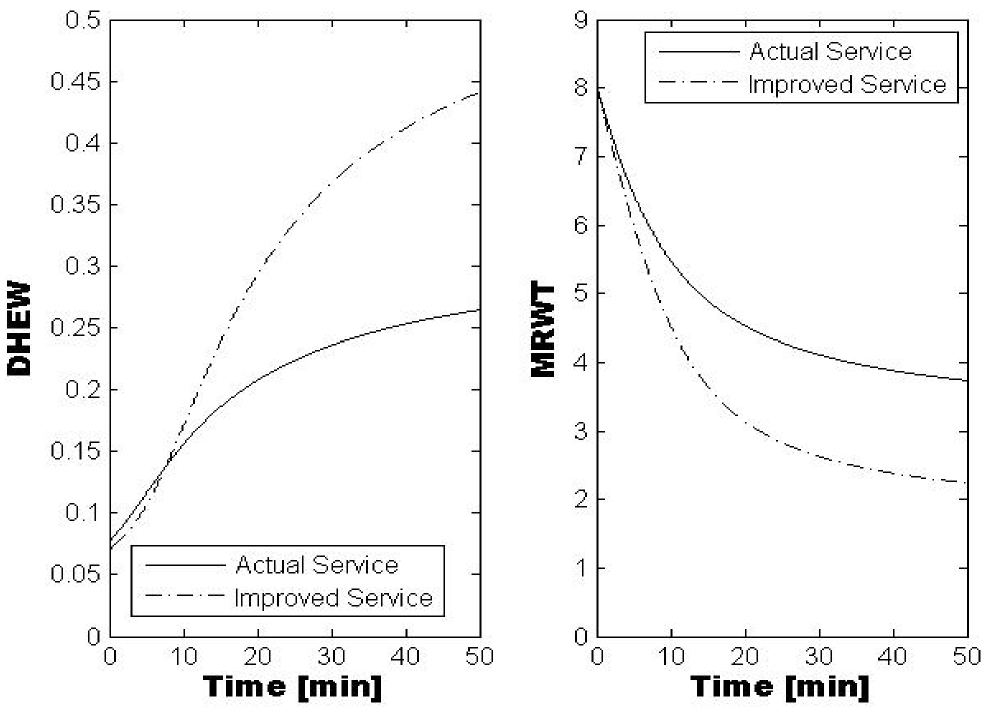


Figure 3-5: Comparison between actual (solid line) and forecasted service (dash-dotted line)

3.5. Conclusions

A set of service quality indexes related to the customer waiting time has been introduced and applied to evaluate the performances of a real bus route. The stationary distribution of the Forward Recurrence Time has been adopted to model the generic customer waiting time.

Conditions in which the proposed approach can be used are specified and the main working hypotheses have been checked. It is shown that the proposed quality indexes are characterized by a precise operative meaning and don't require expensive and extensive data collection. On the basis of data that are usually available in practice, the proposed indexes enable quantifying the effect of service regularity on customers waiting time. On the basis of this inference, a service improvement strategy can be formulated and the expected results estimated.

4 AVAILABILITY ASSESSMENT OF TRANSPORTATION SYSTEMS WITH NON EXPONENTIAL DOWNTIMES

4.1. Introduction

In the last decades Reliability, Availability and Maintainability (RAM) characteristics of transport systems have been assuming a continuously growing importance**.** Indeed, quantitative RAM requirements for the whole system and its main constituents (e.g. Fleet, Traction system, Overhead Catenary System) are nowadays set in practically all the invitations to tender for transport systems. In addition, such requirements are almost always defined via operational indexes which strongly depend on organisational strategies. So, in order to well operate under this new perspective, it is absolutely necessary to dispose of complex stochastic models, which allow to account for the effects of management strategies/decisions on transport systems RAM. In many cases, effective models can be developed via Markov theory, which relies on the assumption that failure and repair times are exponentially distributed. Unfortunately, not always markovian models work adequately. In fact, while the hypothesis of exponentially distributed failure times is usually met in practice, the hypothesis of exponentially distributed repair times is often not realistic and should be removed to obtain satisfactory results. In this chapter, some non-Markovian stochastic models are presented that can be used to evaluate the effect of managerial decisions concerning maintenance service (i.e. number of spare trams, numbers of servers) on operational availability of a fleet of trams in presence of non exponential repair times. Modelling/computation is practically performed adopting the Device of Stages (DOS) technique (Cox, Miller, 1965), (Singh et al., 1973), (Neuts, 1981). In addition, Monte Carlo Simulation (MCS) (Asgarpoor, Mathine, 1997), (Billinton, Li, 1994) is also carried out and results obtained are compared with those obtained via DOS technique. The influence of spare vehicles and number of servers on fleet operational availability is studied and discussed. The effect of inherent repair time distribution (Knessl et al., 1987), (Cooper et al., 1998), (Gupta, Srinivasa Rao, 1998), (Colini et al., 2009), on short and long term operational availability is also investigated and highlighted. Main advantages and drawbacks of the two considered approaches (DOS&MCS) are discussed.

4.2. A second Case Study: Operational Availability assessment of the AnsaldoBreda fleet

Operational Availability of a fleet of identical trams, produced by AnsaldoBreda for ANM (Azienda Napoletana Mobilità), to operate a tramway line in the city of Naples, is analysed. The study is performed only taking into account failures of the traction/braking subsystems (two traction modules and one braking module for each tram), which mainly affect fleet availability. Values of inherent Mean Time To Failures (MTTF), 3000 hours, and Mean Time To Repair (MTTR), 3 hours, of the single traction/braking subsystem were directly provided by AnsaldoBreda.

In order to model the stochastic dependence among the states of different trams, generated by the presence of the queue, the whole fleet is modelled as a multi-state system. Operational availability of the fleet is computed in the following spare trams/servers configurations:

1)	0 spare trams and 1 server	2)	1 spare tram and 1 server
3)	0 spare trams and 2 servers	4)	1 spare tram and 2 servers

Table 4-1: spare trams/servers configurations considered for the case study

Fleet Availability under configurations 1 and 2 is evaluated via model B_1, whereas under configurations 3 and 4 model B_2 is adopted. These models are described in sections 4.3 and 4.4 respectively. The application is developed on the basis of the following hypotheses:

i. The transport service operates "24 hours a day";
ii. Planned service requires at least twenty trams available;
iii. The generic tram is withdrawn from service when a failure occurs to at least one module of its traction/braking subsystem;
iv. Time that a failed tram spends to reach the workshop and time that the spare tram (when present) spends to replace a failed tram is not considered;
v. Only first level maintenance is performed at the workshop (i.e. spare modules are used to replace the failed ones, which are repaired off-line with the aim of maximizing fleet availability);
vi. Limitless spare modules are available at the workshop;
vii. Time spent to perform preventive maintenance is not considered;
viii. Failure times and inherent repair times of trams are mutually independent random variables, r.v.;
ix. In both 1 and 2 server case, trams to be repaired are arranged in a single (common) queue, served according to a First In First Out (FIFO) discipline;
x. Servers work independently. Maintenance crews are assumed to have identical skills and experience (consequently, inherent repair times are identically and independently distributed r.v.);

xi. At any point in time, probability that more than 3 trams need to be repaired is set to zero. This simplifying assumption doesn't significantly affect the results for the considered MTTR and MTTF values.

In this chapter the inherent repair time is assumed to be an Erlang-3 r.v. This assumption allows to easily apply the DOS approach. However, alternative choices are discussed in some details.

4.3. Model B_1(1 Server)

In presence of 1 server and an Erlang-3 inherent repair time, the stochastic process describing the system state is Semi-Regenerative (Çinlar, 1975). In fact, in presence of a single server, operating according to a FIFO discipline, after each repair completion the process enjoys absence of memory, that is process evolution from then onward depends on the current process state (number of trams to be repaired) and not on previous history**.** In this case (see hypotheses in section 4.2), four states can be defined for the fleet, identified by the label $\{d=0,1,2,3\}$, which specifies the number of trams at the workshop. The state space and transitions between states are reported in Figure 4-1, being s=0 (i.e. number of spare trams) under configuration 1 and s=1 under configuration 2.

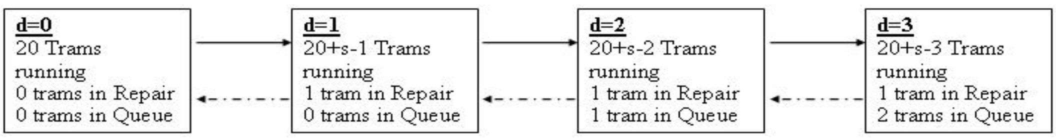


Figure 4-1: Model B_1 - State Space

Fleet is available if $20+s-d \geq 20$. Transition rates due to failures (continuous arrows) are time invariant, (these rates only depend on the number of running trams), transition rates due to repairs (dash-dot arrows) depend only on the time elapsed since the last repair completion.

4.3.1. DEVICE OF STAGES TECHNIQUE

The basic idea behind the DOS technique is that of modeling a single state with non exponential sojourn time by a proper arrangement of multiple stages in which sojourn time is exponential. Theoretically, DOS allows to model/approximate any non negative r.v. (Schassberger, 1973). An eventual drawback of this technique is that, in some cases, depending on the shape of the random variable to model, it may lead to a very large number of stages for some states. The state space in Figure 4-1 leads to the model in Figure 4- 2.

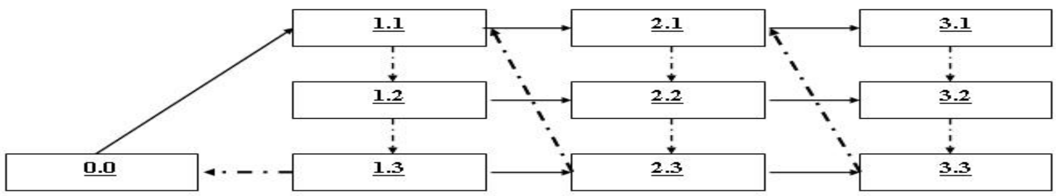


Figure 4- 2: Model B_1 DOS- State Space and transitions

As it is shown in the figure above, in order to identify states with $d \geq 1$ it is now necessary to specify not only the number of trams to be repaired, *d*, but also the stage, *r*, of the ongoing repair. In the case *d*=0 the label *r* is conventionally set to 0 to signify that the server is not busy. Transitions due to failures (solid arrows) make *d* increase by 1 and leave *r* unchanged. Transitions due to repairs (dash-dot arrows), occurring at a rate $\mu = 3 / \mathrm{MTTR}$, make *d* decrease by 1 only when the stage $r = 3$ is left.

4.3.2. MONTE CARLO SIMULATION

MCS method is applied on the basis of the model depicted in Figure 4-1. In order to compute fleet Operational Availability via MCS, 10^5 realizations of the *chronological system State Transition Process* (Billinton, Li, 1994), over a discrete ($\Delta t = 0.1h$) time span of 300 hours, were generated. During each Monte Carlo trial, a realization of the process is obtained simulating the random walk which guides the system from one state to another, at different times. To do this, transition by transition, given the state the system occupies and the time at which this state was reached, it is determined the time at which the next transition occurs and the state reached as a consequence of this transition. In particular, $d = 0$ is characterized by an exponential sojourn time, since only transitions due to failures can occur and hence only state 1 can be entered. When $d \geq 1$ both failure and repair transitions can occur. So, for example, if the system state changes from 0 to 1 at time *t*, sojourn time in state 1, say $t_{\min 1}$, is defined as the smaller between a failure time , t_{F1}, and a repair time, t_{R1}, generated from two independent r.v, Exponential and Erlang-3 distributed, respectively. Thus, the System enters a new state at time $t + t_{\min 1}$. The new state is *d*=0 if $t_{F1} \geq t_{R1}$, it is *d*=2 otherwise. In the latter case, a new failure time, t_{F2}, generated from the appropriate exponential r.v., is compared with the residual repair time, $t_{R1} - t_{F1}$. Sojourn time in state *d*=2, $t_{\min 2}$, is defined as the smaller between t_{F2} and $t_{R1} - t_{F1}$. The new state will be *d*=1 if $t_{F2} \geq t_{R1} - t_{F1}$, state *d*=3 otherwise. When *d*=3 only transition to state *d*=2 is allowed. The sojourn time in state *d*=3 is equal to $t_{R1} - (t_{F1} + t_{F2})$. Such a procedure is repeated until total time reached 300 hours. Operational Availability was computed averaging over the results of 10^5 trials. Stateflow and Simulink Charts developed to perform MCS under configuration 1 are reported in attachments for chapter 4. Failure and Repair Times were obtained via MATLAB Exprnd and Gamrnd functions.

4.4. Model B_2 (2 Servers)

When 2 servers and Erlang-3 inherent repair times are considered, the process describing the system state is Generalized Semi-Markov (Ciardo et al., 1994). In fact, in presence of 2 servers, process enjoys absence of memory only when the workshop is empty (as a consequence of the previous transitions). The state space and the transitions between states are now those reported in Figure 4- 3:

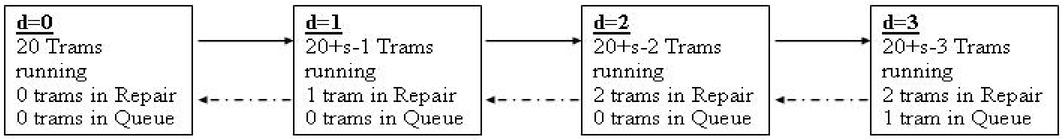


Figure 4- 3: Model B_2 - State Space and transitions

4.4.1. DEVICE OF STAGES

For $d \leq 1$, model is the same as that described in section 3.1. To identify states with $d \geq 2$ it is now necessary to specify the number of trams in the workshop, *d,* and the stages, *r1* and *r2,* of the two ongoing repair activities. Then, repair transitions for $d \geq 2$ are organized as shown in Figure 4- 4.

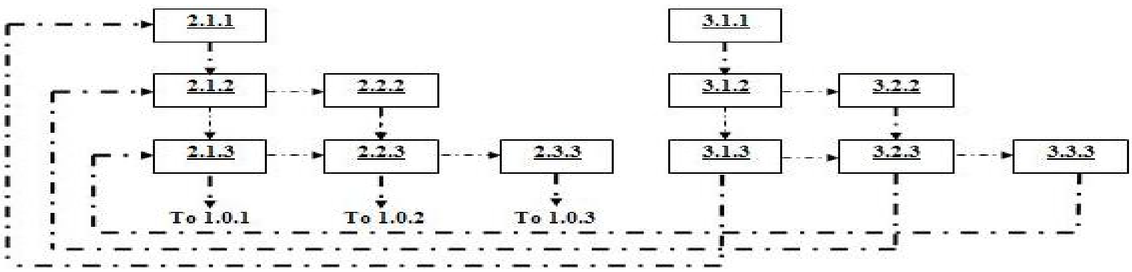


Figure 4- 4: Model B_2 DOS- State Space and repair transitions

Repair transitions depicted by thick arrows occur at a rate equal to 6/MTTR. Those depicted by fine arrows occur at a rate equal to 3/MTTR. Transitions due to failures make *d* increase by 1 and leave unchanged the stage(s) of ongoing repair(s) (see also section 4.3.1). Only states such that $r_1 \leq r_2$ are defined in the model. The use of this expedient allows to split states $d = 2$ and $d = 3$ in only 6 stages instead of 9, without loss of useful information. In the case *d=0*, *r1=r2=0* indicates that both servers are not busy. Similarly, for *d=1*, *r1=0* and *r2≠0* indicates that only one server is busy.

4.4.2. MONTE CARLO SIMULATION

Implementing MCS for the two servers case is a little bit more difficult than the previous case, because for $d \geq 2$ it is necessary to keep memory of the instants in which both ongoing repairs started. In order to simplify this task MCS was performed starting from the model described in section 4.4.1, adopting a procedure similar to that outlined in section 4.3.2.

4.5. Fleet Availability

Results obtained via both DOS and MCS are reported in the following Figure 4-5. All configurations (Table 4-1) are considered. It is assumed that at time 0 all trams are available.

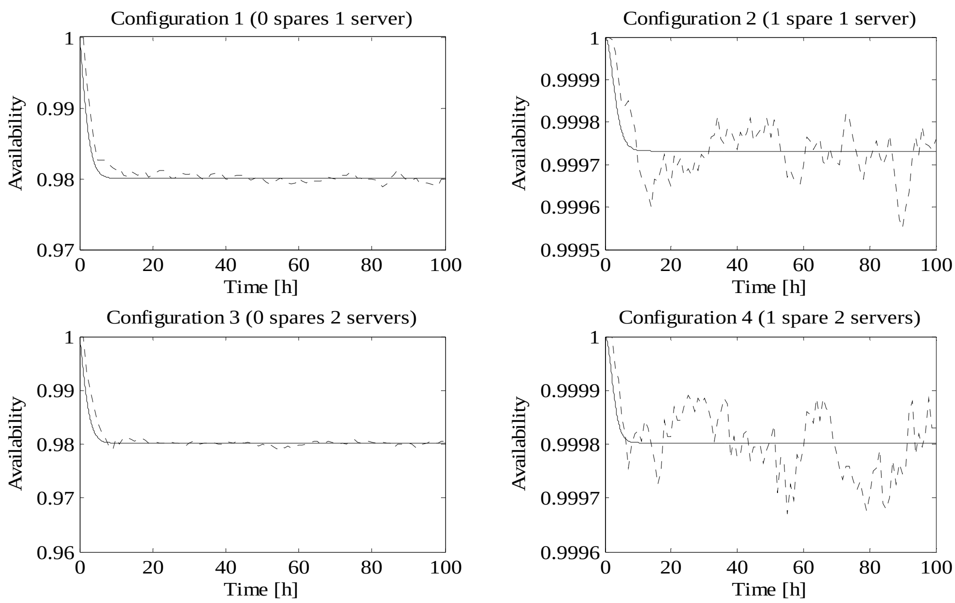


Figure 4-5: Fleet Availability – DOS (solid line) vs MCS (dashed line)

MCS has been performed as described in 4.3.2 and 4.4.2. Fleet availability via DOS technique has been numerically obtained via a discrete-time approximation (step size equal to 0.1 hours) of the differential equation describing the evolution of the state probability vector. Of course, for the Erlang-3 inherent repair time, DOS technique leads to exact state probabilities of the involved non-Markovian processes (see sections 4.4.3 and 4.4.4). Results obtained via DOS (solid line) and MCS (dashed line) are in quite good agreement one each other. The highest fleet availability value is obtained (as expected) under configuration 4, both in the transient and steady phase. It is possible to note that the presence of a spare tram determines an important improvement of fleet operational availability, while the impact of the number of servers is less relevant. About MCS, it is to note that 10^5 Monte Carlo trials don't allow to perform accurate evaluations of the considered availability level (see fluctuations in figure 4-5). Consequently, considered that even more severe requirements are usually defined in the invitations to tender, the DOS approach is usually to be preferred to MCS.

4.6. Comparisons With Other Approaches

In this section the results provided by models B_1 and B_2 are compared to those obtained via the following:

1) Model *C*: the model often used by practitioners, which neglects the effect/presence of the queue and assumes that the spare tram is (always) available with probability 1;

2) Model D_i $\{i=1,2\}$,: a Markov model with inherent exponential repair time.

Comparisons between results obtained via the above models $L=C,D_1,D_2$ and $M=B_1, B_2$ are made adopting the following index:

$$\Delta_{LM}\%(t) = [(A_L(t) - A_M(t))/(1 - A_M(t))] \times 100 \quad L = C, D_1, D_2 \quad M = B_1, B_2$$

which represents the percent difference in unavailability at time *t* computed via the compared models. All computation are performed assuming that at time 0 all trams are available.

4.6.1. MODEL C VERSUS MODELS B_1 AND B_2

Model *C* calculates availability, $A_C(t)$, via the following equation:

$$A_C(t) = \sum_{i=20-i}^{20} \binom{20}{i} p(t)^i [1 - p(t)]^{20-i} \qquad s = 0,1 \tag{4.1}$$

where *s* is the number of spare trams (available with probability 1) and:

$$p(t) = \frac{\mu}{\mu+\lambda} + \frac{\lambda}{\lambda+\mu}\exp\left[-(\lambda+\mu)t\right]; \quad \lambda = \frac{1}{MTTF}; \quad \mu = \frac{1}{MTTR}; \quad s = 0,1$$

Model *C* assumes that spare trams availability is 1 for every $t \geq 0$.

For *s*=0 equation (4.1) reduces to the equation proposed in (Trivedi, 2002) for a series of *n* identical systems in presence of *n* identical servers operating independently of one another.

Results obtained applying this model may significantly differ from those obtained via models B_1 and B_2. This mainly depends on the number of servers, the number of spare trams, the shape of the repair time distribution, MTTF and MTTR of the single unit (mainly on MTTF/MTTR). Plots of $\Delta_{CB1}\%(t)$ and $\Delta_{CB2}\%(t)$ are reported in Figure 4-6.

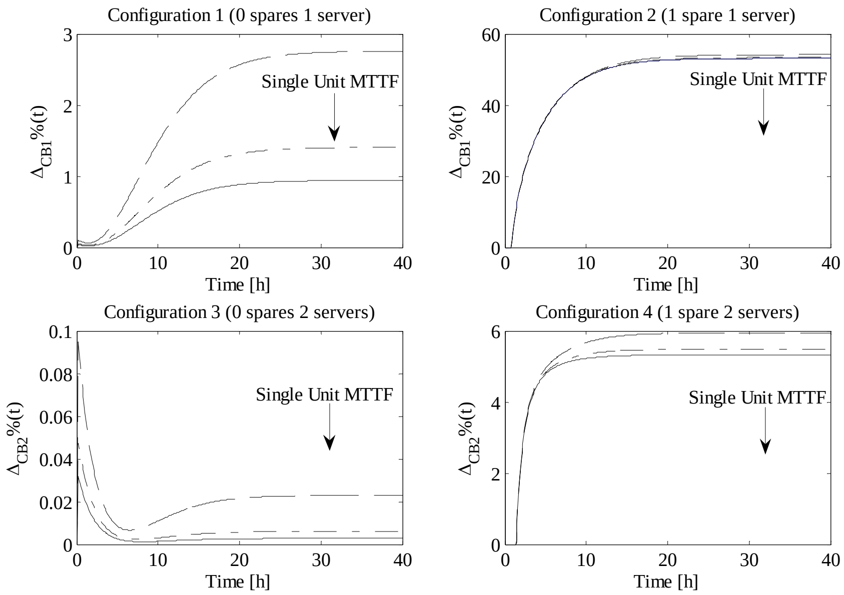


Figure 4-6: plots of indexes $\Delta_{CB1}\%(t)$ and $\Delta_{CB2}\%(t)$

For each configuration, three different values are considered for the MTTF: 1000, 2000, and 3000 hours, respectively. Related results are reported in Figure 4-6. The MTTF increases as indicated by the arrow.

In all cases model *C* leads to overestimate fleet availability. For given a MTTF, $\Delta\%$ decreases as the number of servers increases and increases as *s* passes from 0 to 1. The worst results are obtained under configuration 2, where $\Delta\%$ approaches 50%.

4.6.2. MODELS D_I VERSUS MODELS B_I

Mathematical details of the well known models D_i are skipped for sake of brevity. Plots of $\Delta_{D1B1}\%(t)$ and $\Delta_{D2B2}\%(t)$ obtained under each configuration are depicted in Figure 4-7.

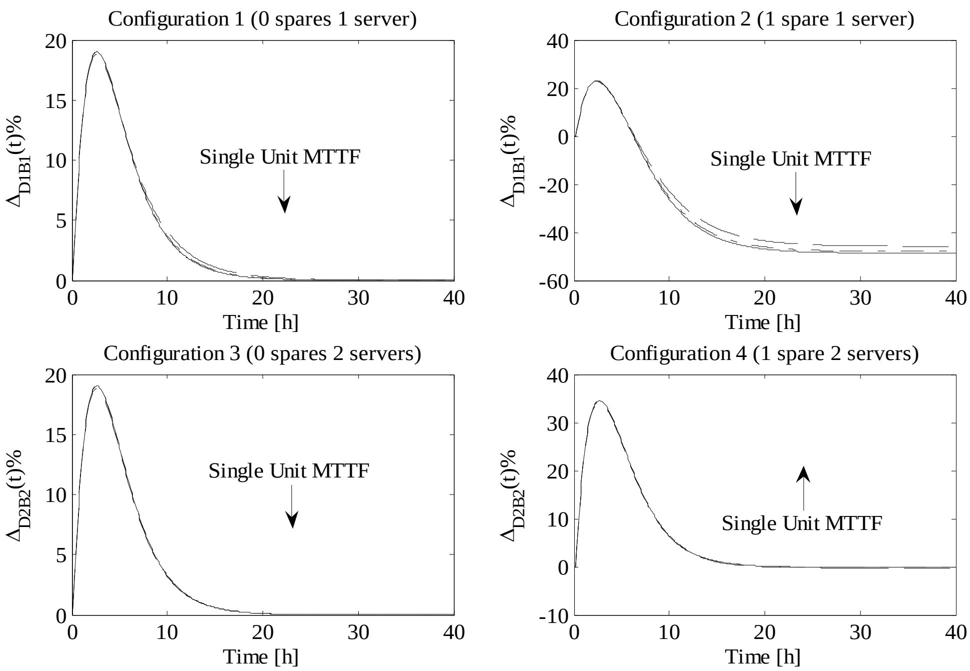


Figure 4-7: plots of indexes $\Delta_{D1B1}\%(t)$ and $\Delta_{D2B2}\%(t)$

As in section 4.6.1, three different values are considered for the MTTF (i.e. 1000, 2000, and 3000 hours). Figure 4-7 shows that the effect of MTTF in this case is not relevant. On the contrary, the effect of different repair time distributions on fleet operational Availability is clear. Configuration 2 is the one for which the effect of different repair time distributions is more evident. The related plot shows that during the first (say 5 hours) the index $\Delta\%$ increases and assumes positive values; subsequently it decreases settling on a constant negative value. Plots related to configurations 1,3 and 4 give a clear evidence of the effect of the repair time distribution on transient phase availability. On the contrary the effect on steady phase availability, although certainly present, cannot be appreciated on the basis of a visual analysis. More in detail, availability in configurations 1 and 3 (i.e. 0 spare trams) is higher under the Exponential than under the Erlang-3 time to repair, both in the transient and steady phase. For configurations 2 and 4 (i.e. 1 spare tram) transient phase availability is higher under the Exponential repair time, while steady phase availability is higher under the Erlang-3 repair time. The observed effects (given the MTTR) seem to be mainly due to differences in skewness and variance of the two considered repair time distributions. Skewness affects availability since probability of performing quicker than average (i.e. MTTR) repairs increases as the (positive) skewness increases. Variance affects Availability via the waiting time distribution, whose mean increases as the variance of the repair time distribution increases, as a consequence of the so called inspection paradox (Stein, Dattero, 1985).

Obtained results, at least in the steady phase, are in accordance with results obtained in literature for finite source queues (Knessl et al., 1987), (Gupta, Srinivasa Rao, 1998). Obviously, the performed analysis doesn't allow defining a mathematical relationship between the fleet operational availability and the repair time distribution, neither it enables one to a-priori establish how significantly the repair time distribution may affect availability. Despite this, it clearly highlights that fleet operational availability is influenced by the repair time distribution, both in the transient and in the steady phase, even though the discussed effect may appear not always relevant. Nonetheless, a proper modelling of the repair time distribution can be always considered worthy, even if this leads to more complicated mathematics. In fact, in the considered applicative field, penalties due to poor availability performance are usually so severe to make intolerable even a very little loss of accuracy.

4.7. Conclusions

The effect of managerial decisions concerning maintenance service, namely the number of spare trams and the number of servers, on the operational availability of a fleet of trams was analysed. The realistic hypothesis that *Inherent* repair times are non-exponential random variables was assumed. The impact on fleet availability of this latter assumption was evaluated via appropriate non-Markovian processes. Device of Stages technique has been used to compute fleet operational availability under four different configurations (i.e. operational scenarios). Availability computation was also performed via Monte Carlo Simulation, highlighting drawbacks this approach may lead too. It is shown that the proposed non-Markovian models allow to adequately evaluate the impact on fleet operational availability of the considered factors, which affect both transient and steady phase availability. Results provided by the proposed models were also compared to those obtained via two simpler models: a pure markovian model and a model which is usually adopted by practitioners. Obtained results evidence that the latter simpler models cause a loss of accuracy that is not possible to tolerate. In fact, in the considered applicative field, the severity of penalties foreseen in the case of poor availability performance is very high. Indeed, this kind of penalties, which are customarily set as a percent of Operation & Maintenance payments, can go up to tens of millions of Euros (e.g. in the case of a fleet of 20 trains and a System lifecycle of 20 years this kind of penalty approaches 10 millions of Euros).

5 QUALITY VS COSTS: THE LIFE CYCLE COST

5.1 Towards a whole Lifecycle Cost – Reliability Model

Since promoting public transportation is one of the main levers to ensure environmental sustainability, great interest has been focused by Transport Authorities on policies aiming at promoting the use of public transport. In this respect, ensuring that passengers perceive a high service quality is of paramount importance. In addition, an increased use of public transport can be achieved making it relatively affordable with low fares. Thus, Transport Authorities have to manage an important trade-off between Quality and costs.

As previously highlighted, service quality is mainly related to service on-time performances, which strongly depend on systems reliability and on the operators' capability of minimizing service degradation in failure conditions. Thus, the above trade-off can be in large part explained managing the trade-off between reliability and costs from a lifecycle perspective (i.e. taking into account costs to be born during the whole system lifecycle). Of course, managing this trade-off on an integrated basis, since the early stages of the system lifecycle is strategically important also for companies delivering transportation systems: costs associated to different lifecycle phases are strongly interrelated and, even if a large portion of costs is the direct result of activities pertaining to systems operation and support, the commitment of these costs depends on decisions made during conceptual/preliminary design (Fabrycky, Blanchard, 1991), (Wheatcroft, 1985).

Lifecycle cost estimates are usually performed in the industry by analogy, on the basis of costs actually born to design, build, operate and maintain comparable systems. However, the translation of costs born for a given system in costs to be born for a new system is largely subjective and empirical. Thus, it is not possible to properly address the problem of analyzing the effects of reliability targets on the system lifecycle cost. In order to make this possible, it is necessary to develop a model that explains and explicitly accounts for costs related to systems reliability: failures during revenue operation, the acquisition of a given inherent reliability level and efforts required to reach the planned reliability target. From this point of view, modeling the failure rate behavior of a complex system during the whole lifecycle, from field testing to revenue operation, is very important since the development and support costs are strictly related to the reliability improvement process and to the steady-state reliability level. Thus, in order to formulate a Lifecycle cost – Reliability model, it is at first necessary to focus on the formulation and parameters estimation of an intensity model capable of fitting early and useful life failures. Once a model capable of explaining the failure rate behavior is found, relationships between reliability, acquisition costs, support costs and costs to be born to reach the desired reliability level can be identified. Following the above approach, a Lifecycle cost – Reliability model is formulated. Its main characteristic is the concept of inertia pertaining to the reliability improvement process, which largely explains costs to be born, during the testing phase and the early stages of operation of a given system, to reach the planned reliability target.

5.2 Fitting Early And Useful Life Failures Via The Hyperbolic Model

In the context of modelling/forecasting the Life Cycle Cost of a complex repairable system, the need for an intensity function capable of explaining the system failure rate behavior during both early and useful life arises. The idea of modelling the failure rate behavior of complex repairable systems via a Non Homogeneous Poisson Process (NHPP) is not new. As an example, starting from Duane's investigations, which led to the so called learning curve approach (Duane, 1964), Crow proposed to model the failure process of a complex repairable system during the testing phase (i.e. reliability growth) via a NHPP with a power law intensity (Crow, 1974). On the basis this idea, the Hyperbolic failure intensity model, which enables counting via a NHPP the failures that must be financially supported during the early stages of operation, including field testing, and the useful life of a complex repairable System, is presented. This three parameters model is characterized by a decreasing failure intensity approaching a non-zero lower bound, which makes it suitable to model the failure rate behavior during the above lifecycle phases. A two parameters version of this failure intensity model has been introduced in (Erto, 1988). Moreover, the functional form proposed in (Erto, 1988) has been used in (Erto, Palumbo, 2005) to characterize the hazard rate function of a reliability model for non-repairable systems. In addition, the proposed three parameters model is quite close to the functional form assumed by the Army Maturity Projection Model, used for reliability projections, when the number of B-mode failures (i.e. failure modes that receive corrective action during development and testing, see) tends to infinity (Ellner, Wald, 1995). Since reliability improvements of a complex repairable system are due to corrective actions implemented during testing, the counting process of failures experienced during system testing is not characterized by independent increments. So, a justification for the use of a NHPP to model the failure rate behavior of a complex repairable system during the early stages of operation, including testing, and the useful life is necessary. To this aim, conditions in which reliability performances of a system can be, on average, explained by the considered NHPP are briefly outlined. Then, procedures to obtain the Maximum Likelihood and Minimum Chi - square estimates of model parameters, for grouped failure data, are outlined and discussed. Model suitability to fit the counting process of failures of complex repairable systems during both the early stages of operation and the useful life is checked fitting the model to the failure process experienced by the fleet of trains operated within the M1 and M2 lines of the Copenhagen Metro System during the first two years of operation.

5.2.1 The Hyperbolic Model

The proposed failure intensity, which generalizes the one proposed in (Erto, 1988), is:

$$\lambda(t) = r + \frac{a}{bt+1} \qquad a > 0,\ b > 0, r > 0$$

which is strictly decreasing from the early maximum value a+r to the asymptotic minimum r.

The parameter a is called limiting decrease of the failure intensity (Erto, 1988). It may be viewed as the overall failure rate, at the beginning of system testing and operation, due to B-mode failures. The parameter r is the steady state failure intensity. It may be viewed as the failure rate due to failure modes that not receive corrective action (i.e. A-mode failures). The parameter b is a scale parameter, depending on the unit of measurement of the system age (e.g. calendar time, miles and so on), t. The application of such a model to count, via a NHPP, failures a complex repairable system undergoes during testing and operation is probabilistically justified in (Ellner et al., 1998). In fact, under the following assumptions:

- time to first occurrence of the generic B-mode failure is assumed to be an Exponential random variable (r.v.) with parameter λ_i;
- the mode to mode variation in the initial B-mode failure rates is a Gamma r.v. with scale parameter *b* and shape parameter *ab*/*k*, being k the number of different B-mode failures;

the stochastic process that counts the number of first occurrences of B-mode failures by time *t* converges, as *k* tends to infinity, to a NHPP with the following mean value function $M_I(t)$:

$$M_1(t) = \frac{a}{b}\ln(bt+1)$$

so taking into account A mode failure rate, the following mean value function is obtained:

$$M(t) = rt + \frac{a}{b}\ln(bt+1)$$

whose first derivative with respect to *t* coincides with the proposed failure intensity model.

5.2.2 Maximum Likelihood Estimation

Given the number of failures a system faced in some (non-overlapping) operating windows, the Likelihood function *L* is:

$$L(a,b,r|n_i,\Delta t_i) \propto \prod_{i=0}^{m-1} \frac{\left[r(t_{i+1}-t_i)+\frac{a}{b}\ln\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right]^{n_i}}{n_i!} \exp\left\{-\left[r(t_{i+1}-t_i)+\frac{a}{b}\ln\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right]\right\} \quad (5.1)$$

Where:

- *m* is the number of operating windows;

- n_i is the number of observed failures within the i^{th} operating window
- $\Delta t_i = t_{i+1} - t_i$ is the amplitude of the i^{th} operating window.

Taking the logarithm of equation (5.1), it is possible to obtain:

$$\log\left(L\left(a,b,r \middle| n_i,\Delta t_i\right)\right) = \sum_{i=0}^{m-1} n_i \log\left[r\Delta t_i + \frac{a}{b}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right] - \sum_{i=0}^{m-1}\log\left(k_i!\right) - rt_m - \frac{a}{b}\log\left(bt_m+1\right) \quad (5.2)$$

It is possible to show that the ML estimate of the total number of failures at the end of the observation period coincides with the total number of observed failures. In fact, letting $c = \frac{r}{a}$, the mean value function can be expressed as $M(t) = c \cdot a \cdot t + \frac{a}{b}\ln(bt+1)$. If τ is the total observation time, we obtain:

$$M(\tau) = \gamma = c \cdot a \cdot \tau + \frac{a}{b}\ln(b\tau+1) \Rightarrow a = \frac{\gamma}{c\tau + \frac{\ln(b\tau+1)}{b}}$$

Thus:

$$z(t) = \left[c + \frac{1}{(bt+1)}\right]\frac{\gamma}{c\tau + \left[\ln(b\tau+1)/b\right]}$$

$$M(t) = \frac{ct + \left(\ln(bt+1)\right)/b}{c\tau + \left(\ln(b\tau+1)\right)/b} \cdot \gamma$$

If the first n times to failure of a given repairable system are known, the likelihood function is:

$$\begin{aligned}
L &= \exp\left[-M(\tau)\right] \cdot \prod_{i=1}^{n}\left[z(t_i)\right] = \\
&= \exp(-\gamma)\prod_{i=1}^{n}\left\{\left[c + \frac{1}{(bt_i+1)}\right]\frac{\gamma}{c\tau + \left[\left(\ln(b\tau+1)\right)/b\right]}\right\} = \\
&= \exp(-\gamma) \cdot \gamma^n \cdot \prod_{i=1}^{n}\left[\frac{c + (bt_i+1)^{-1}}{c\tau + \left[\left(\ln(b\tau+1)\right)/b\right]}\right]
\end{aligned}$$

The log-likelihood, *l*, can be expressed as follows:

$$l = \ln L = -\gamma + n\ln(\gamma) - n\ln\left[c\tau + \left(\ln(b\tau+1)\right)/b\right] + \sum_{i=1}^{n}\ln\left[c + (bt_i+1)^{-1}\right]$$

Taking the first derivative of the log-likelihood with respect to $M(\tau)=\gamma$ and equating it to zero it is possible to obtain:

$$\frac{dl}{d\gamma} = -1 + \frac{n}{\gamma} = 0 \Rightarrow n = \gamma$$

The same result stands for the case of grouped data. Based on the above result, equation (5.2) can be expressed in terms of only 2 parameters, say *b* and *r*:

$$\log\left(L(b,r|n_i,\Delta t_i)\right) = \sum_{i=0}^{m-1} n_i \log\left\{r\Delta t_i + \frac{[N(t_m)-rt_m]}{\log(bt_m+1)}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right\} - \sum_{i=0}^{m-1}\log(n_i!) - N(t_m) \quad (5.3)$$

The function (5.3) can be maximized via Quasi-Newton methods, preferably providing the analytical expression of the gradient of (5.3) as input for the maximization routine:

$$\frac{\partial\log(L)}{\partial b} = \sum_{i=0}^{m-1}\frac{n_i\left\{\left[t_{i+1} - \frac{t_i(bt_{i+1}+1)}{bt_i+1}\right]\frac{1}{bt_{i+1}+1} - \frac{t_m\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)}{2(bt_m+1)}\right\}}{\left\{r\Delta t_i + \frac{[N(t_m)-rt_m]}{\log(bt_m+1)}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right\}}\frac{[N(t_m)-rt_m]}{\log(bt_m+1)}$$

$$\frac{\partial\log(L)}{\partial r} = \sum_{i=0}^{m-1}\frac{n_i\Delta t_i - \frac{n_i t_m}{\log(bt_m+1)}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)}{r\Delta t_i + \frac{(N-rt_m)}{\log(bt_m+1)}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)}$$

In order to ensure convergence of the optimization routine, good starting values should be chosen. A very simple and practical approach in finding starting values capable of ensuring a high chance of convergence consists in a-priori specifying the parameter b and then estimating parameters a and r via Ordinary Least Squares. Numerical experiments have shown that, in order to ensure convergence, it is only important to match the correct order of magnitude of the parameter b. So, the above procedure can be repeated many times, choosing as starting values the values b and r for which the highest r-square is obtained. The proposed numerical procedure has been tested on 100 replications of a NHPP with Hyperbolic intensity with parameters specified in Table 5-1 :

a [failures/km]	3,00E-03
b [km^{-1}]	1,00E-04
r [failures/km]	1,00E-03

Table 5-1: model parameters for numerical experiments

Different values for Δt_i and T_m have been considered. More precisely:

- Case 1: $\Delta t_i = 230000$ km, $T_m = 115000$ km;
- Case 2: $\Delta t_i = 5750$ km, $T_m = 115000$ km;
- Case 3: $\Delta t_i = 230000$ km, $T_m = 460000$ km;
- Case 4: $\Delta t_i = 5750$ km, $T_m = 460000$ km.

The proposed procedure in all cases converged to the solution. The performed numerical experiments revealed that, given the values of a and b, the m.l.e. of the parameter r can be strongly optimistic (or even negative). when the product bT_m is low. More in detail, for cases 1 and 2 optimistic results have been obtained 12 and 14 times respectively, whereas for cases 3 and 4 only 0 and 1 optimistic estimates for *r* have been obtained. Thus, on the basis of the above results one may argue that when a system undergoes a reliability program aiming at a 75% failure rate reduction, the m.l.e. of *r* in almost 10% may be strongly optimistic (or even negative) when $bT_m = 10$, whereas such a situation is quite unlikely when $bT_m = 40$. Of course the above probabilities may decrease as the product bT_m increases and/or the failure rate reduction decreases.

5.2.3 Minimum Chi-Square Estimation

Minimum Chi-square estimation is a point estimation technique consisting in minimizing the Chi-square statistic. Even if the Chi-square statistic is widely used for goodness of fit testing, its use for point estimation purposes is quite unpopular. In fact, differently from Maximum Likelihood, Minimum Chi-square estimators are often unavailable in closed form. In addition, since Maximum Likelihood estimation leads to estimators that are asymptotically equivalent to those obtained minimizing the Pearson's Chi-square statistic, at least for large samples one might concentrate on whichever procedure is easiest to undertake and this is usually Maximum Likelihood estimation.

In this section, a procedure to obtain Minimum Chi-square estimators for the Hyperbolic model parameters is outlined. Also in this case, the grouped failure data setting is considered.

In the considered setting, the Pearson's chi square function is:

$$\chi^2(a,b,r|n_i,\Delta t_i) = \sum_{i=1}^{m} \frac{\left[n_i - r(t_{i+1} - t_i) - \frac{a}{b}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right]^2}{r(t_{i+1} - t_i) + \frac{a}{b}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)} \qquad (5.4)$$

Where:

- m is the number of operating windows;
- n_i is the number of observed failures within the i^{th} operating window;
- $\Delta t_i = t_{i+1} - t_i$ is the amplitude of the i^{th} operating window.

The function in (5.4) can be minimized via Quasi-Newton methods, preferably providing the analytical expression of the gradient of (5.4) as input for the maximization routine. Differently from Maximum Likelihood estimation, in this case it is not express one of the three parameters as a function of the other two. In the following, the partial derivatives of the Chi-square function with respect to *a*, *b*, *r* are depicted:

$$\frac{\partial \chi^2}{\partial a} = \sum_{i=1}^{m} \frac{2\left[n_i - N\left(t_{i+1}\right)\right]\left[-\frac{1}{b}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)\right]N\left(t_{i+1}\right) - \left[n_i - N\left(t_{i+1}\right)\right]^2 \frac{1}{b}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)}{N\left(t_{i+1}\right)^2}$$

$$\frac{\partial \chi^2}{\partial b} = \sum_{i=1}^{m} \frac{2\left[n_i - N\left(t_{i+1}\right)\right]\left[-\frac{a}{b^2}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right) + \frac{a}{b}\left(\frac{t_{i+1}}{bt_{i+1}+1} - \frac{t_i}{bt_i+1}\right)\right]N\left(t_{i+1}\right)}{N\left(t_{i+1}\right)^2}$$

$$- \sum_{i=1}^{m} \frac{\left[n_i - N\left(t_{i+1}\right)\right]^2\left[-\frac{a}{b^2}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right) + \frac{a}{b}\left(\frac{t_{i+1}}{bt_{i+1}+1} - \frac{t_i}{bt_i+1}\right)\right]}{N\left(t_{i+1}\right)^2}$$

$$\frac{\partial \chi^2}{\partial r} = \sum_{i=1}^{m} \frac{2\left[n_i - N\left(t_{i+1}\right)\right]\left(t_{i+1} - t_i\right)N\left(t_{i+1}\right) - \left[n_i - N\left(t_{i+1}\right)\right]^2\left(t_{i+1} - t_i\right)}{N\left(t_{i+1}\right)^2}$$

being $N\left(t_{i+1}\right) = r\left(t_{i+1} - t_i\right) + \frac{a}{b}\log\left(\frac{bt_{i+1}+1}{bt_i+1}\right)$

In order to find good starting values for the optimization routine, it is possible to proceed as described in section 5.2.2.

5.2.4 Maximum Likelihood Vs Minimum Chi-Square

In this section Maximum Likelihood and Minimum Chi-square estimators are compared. To this aim, 1000 replications of a NHPP with Hyperbolic failure intensity with the following parameters :

a [failures/km]	1.68E-04
b [km^{-1}]	3.78E-06
r [failures/km]	8.40E-07

Table 5-2: First parameters set for Maximum Likelihood and Minimum Chi-square estimators comparison

have been generated.

Each process replication consists in the number of observed failures in 15 operating windows organized as follows:

Operating window	Lower Bound [km]	Upper Bound [km]
1	0	400000
2	400000	800000
3	800000	1200000
4	1200000	1600000
5	1600000	2300000
6	2300000	3000000
7	3000000	3700000
8	3700000	4500000
9	4500000	5300000
10	5300000	6100000
11	6100000	6900000
12	6900000	7700000
13	7700000	8570000
14	8570000	9460000
15	9460000	10350000

Table 5-3: Operating windows for comparisons

In the following table, are summarized main characteristics pertaining to both Maximum Likelihood and Minimum Chi-square estimators:

Model Parameters	***True Value***	***Maximum Likelihood***			***Minimum Chi Square***		
		Mean	*Variance*	*MSE*	*Mean*	*Variance*	*MSE*
a [Failures/km]	1.68E-04	1.82E-04	3.32E-09	3.52E-09	1.79E-04	3.12E-09	3.24E-09
b [1/km]	3.78E-06	4.49E-06	6.89E-12	7.39E-12	4.31E-06	6.75E-12	7.03E-12
r [Failures/km]	8.40E-07	7.09E-07	3.98E-12	4.00E-12	1.03E-06	4.33E-12	4.36E-12

Table 5-4: Mean Squared Errors for Maximum Likelihood an Minimum Chi-square estimators – first comparison

It can be noted that Minimum Chi – square estimators for parameters *a* and *b* are characterized by a lower mean squared error if compared to Maximum Likelihood estimates. On the contrary, Maximum Likelihood estimates of *r* are characterized by a lower mean squared error if compared to the Minimum Chi – square estimate of *r*.

In the considered setting, the product $bT_m \approx 8.7$. Thus, as observed in section 5.2.2, some optimistic estimates of *r* could be expected. Such a situation occurred in 30% of cases for both Maximum Likelihood and Minimum Chi – square estimation.

Other comparisons have been performed between Maximum Likelihood and Minimum Chi – square estimates on the basis of 1000 replications of a NHPP with the following parameters for the Hyperbolic failure intensity:

a [failures/km]	1.00E-03
b [km^{-1}]	1.00E-05
r [failures/km]	1.00E-05

Table 5-5: Second parameters set for Maximum Likelihood and Minimum Chi-square estimators comparison

Also in this case, each process replication consists in the number of observed failures in 15 operating windows organized as in the previous case (see Table 5-3).

In the following table, are summarized main characteristics pertaining to both Maximum Likelihood and Minimum Chi-square estimators:

Model Parameters	True Value	Maximum Likelihood			Minimum Chi Square		
		Mean	Variance	MSE	Mean	Variance	MSE
a [Failures/km]	1.00E-03	9.32E-04	2.39E-08	2.85E-08	9.24E-04	2.35E-08	2.93E-08
b [1/km]	1.00E-05	9.47E-06	7.16E-12	7.44E-12	9.33E-06	7.06E-12	7.51E-12
r [Failures/km]	1.00E-05	9.73E-06	8.12E-12	8.19E-12	1.01E-05	8.51E-12	8.54E-12

Table 5-6: Mean Squared Errors for Maximum Likelihood an Minimum Chi-square estimators – second comparison comparison

In this case, Maximum Likelihood estimates for all parameters are characterized by the lowest Mean Squared Error. Moreover, being bt_m higher than in the previous case no optimistic estimates for r have been obtained.

Based on the analyses performed above, one may conclude that Maximum Likelihood and Minimum Chi – square estimators are substantially equivalent in terms of mean squared error. In addition, their behaviour with respect to bt_m is substantially the same. The only advantage of Maximum Likelihood is due to the fact that it is easier to implement.

5.2.5 A third Case Study: the failure process of the Copenhagen Metro fleet

In order to show model suitability to fit the counting process of failures of complex repairable systems, it is fitted to the failure process experienced by the fleet of trains operated within the M1 and M2 lines of the Copenhagen Metro System. Available data consist in 20 non overlapping operating windows, covering the period of time between February 2003 and January 2005. For each operating window the fleet revenue kilometers and the number of failures are specified (see attachments to chapter 5). More in detail, a failure is meant as a technical problem to the vehicle and on board installations leading to train withdrawal from service, not necessarily by means of rescue procedures.

Following the procedure outlined in the previous section, ML estimates of model parameters have been obtained:

a [failures/km]	7,30E-04
b [km^{-1}]	2,57E-06
r [failures/km]	4,67E-05

Table 5-7: Maximum Likelihood estimates of model parameters

In figure 5-1-a the observed realization of the cumulative number of train removals from service (asterisks) and the estimated mean value function of the considered counting process (solid line) are depicted, whereas in figure 5-1-b the estimated (smooth line) and empirical failure intensity functions are depicted:

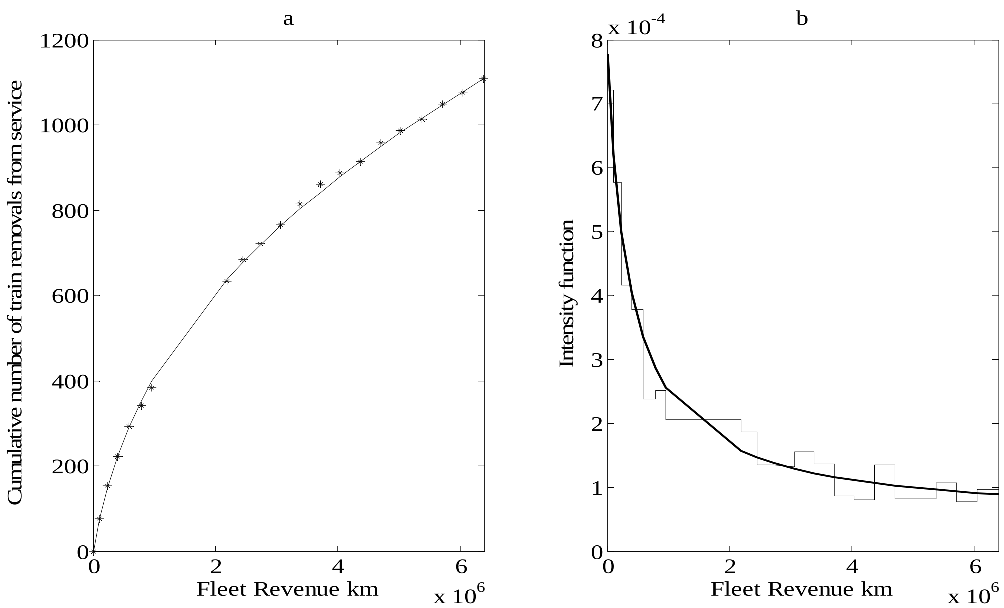


Figure 5-1: Observed and fitted cumulative number of failures and failure intensity

It can be noted that the proposed model seems very suitable to fit the counting process of failures experienced by the fleet of vehicles during the considered period. This has been confirmed by a chi-square Goodness of fit test, whose results are summarized in the following table:

χ^2 statistic	21,22
p-value (χ^2, 16 degrees of freedom)	0,17

Table 5-8: Chi-square goodness of fit test results

Since log-likelihood generates nearly elliptical contours centered at the maximum likelihood estimates (see Figure 5-2), the "normal approximation" (Meeker, Escobar, 1998) to obtain confidence intervals for the parameters seems to be suitable in this case. As an example, the 90% two sided confidence interval for *r* is [2.62E-05, 6.73E-05] failures/km and the 90% two sided confidence interval for *b* is [1.43E-06, 3.72E-06] km^{-1}.

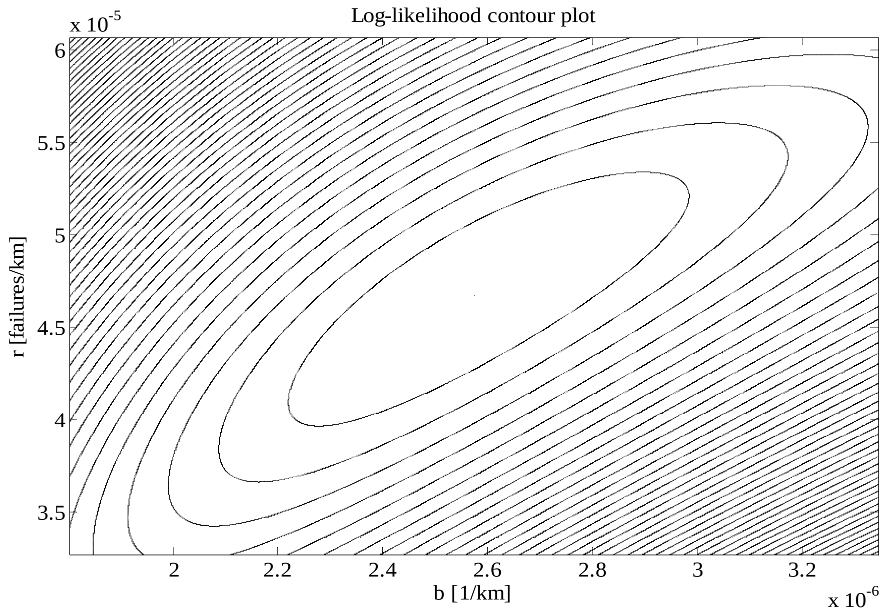


Figure 5- 2: Log-likelihood contour plot

5.3 The Cost Model

It is well known in literature that the reliability level of a given system explains about 66% of its Lifecycle Cost (Seger, 1983). More in detail, reliability affects costs to be born throughout all the main lifecycle phases, namely:

- Research, Development and Testing;
- Procurement;
- Operation and support.

In this section, a model capable of explaining the part of the Lifecycle cost depending on reliability investments is presented. The model is fully based on the failure intensity model introduced in section 5.2.1. Thus, it is capable of relating research, development, testing and maintenance costs to the reliability growth process. Moreover, the model allows to incorporate costs to be born to attain a given inherent reliability level. In addition, the number of model parameters is quite low and they are all characterized by a clear and precise operative meaning. The general formulation of the model is:

$$LCC\left(t\middle|a,b,r,A,B,C\right)=A\left[rt+\left(\frac{a}{b}\right)\ln\left(bt+1\right)\right]+B\cdot f\left(r\right)+\frac{Ca^2b^2}{2}\left[1-\frac{1}{\left(bt+1\right)^4}\right] \quad (5.5)$$

Where :

- $LCC\left(t\middle|a,b,r,A,B,C\right)$ is the system life cycle cost up to *t*;
- *t* is a relevant measure of the system age (calendar time, revenue kilometers and so on);
- *A* is an economic constant, accounting for in-service failures;
- *B* is an economic constant, accounting for the acquisition of a given inherent reliability level;

- C is an economic constant, accounting for engineering efforts to be sustained during development, testing and early stages of operation to reach the planned reliability target, r.

It is possible to note that the model structure encompasses costs to be financially supported due to failures during revenue service (first term), investment costs (second term), engineering and design costs during testing and early stages of operation to reach the planned reliability target (third term). In the following section, each of the three terms is discussed in some detail.

5.3.1 Costs due to failures during revenue service

These costs mainly pertain to corrective maintenance and losses due to service interruptions. For example, if mass transit vehicles are considered, corrective maintenance costs include labour due to first and second line repair activities, spare parts and consumables, testing equipment and maintenance tools. Thus, in order to estimate the corrective maintenance cost per failure, information included in corrective maintenance analyses (MTTR, tools, materials) are of crucial importance. Other critical factors are the service level of the repair shop and, last but not least, the system mission profile. Losses due to service interruptions can be quantified on the basis of penalties for delays foreseen in service contracts, which are often set considering ridership demands and timetables.

The above costs can be accounted for in the term A and the total cost of failures during revenue service up to t is given by:

$$A\left[rt + \left(\frac{a}{b}\right)\ln(bt+1)\right]$$

5.3.2 Investment Costs

These costs pertain to the acquisition of a specified inherent reliability level. They encompass engineering and design costs, research and development costs, factory acceptance tests and inspections/controls during production. The knowledge of a relationship between the above costs and reliability is of crucial importance to solve critical design problems, such as the reliability apportionment among assemblies constituting a complex system. In (Govil, Aggarwal, 1982) a complete overview of cost-reliability relationships is provided.

From a practical standpoint, especially when dealing with complex systems such as mass transit vehicles, a typical and important question one should be able to respond is "how much does the acquisition cost increase if the planned reliability target is 10% higher than the target set for a previously developed and fielded system?".

In order to answer such a question, (Long et al., 2007) noted that when the ratio of reliability investment to average production unit cost (APUC) is plotted against the percentage improvement in reliability on a log-log scale, the result is a straight line. More in detail, the proposed cost estimating relationship is:

$$\log\left(\frac{New\,MTBx}{Old\,MTBx}\right) = 0.343\log\left(\frac{Investment}{APUC}\right) - 0.81$$

From which it follows:

$$Investment = \left[\exp(0.81)\frac{New\ MTBx}{Old\ MTBx}\right]^{\frac{1}{0.343}} \cdot APUC \qquad (5.6)$$

The above relationship has been calibrated on the basis of several programs concerning military systems. Since used data were from a disparate sampling of systems, the above relationship is likely to be system and technology independent (Long et al., 2007).

5.3.3 Costs during field testing and early stages of operation

These costs relate to the development and implementation of measures aimed at mitigating failure modes not adequately addressed during system design and development (e.g. effects of environmental conditions on specific components), with the aim of attaining the planned reliability levels. Among the three identified cost components, this is for sure the most difficult to evaluate, due to the complexity and variety of activities to be carried out to attain the planned reliability goals. However, in macroscopic terms, following the idea proposed in (Erto, 1998), one may argue that the above costs are somehow proportional to the inertia which leads the system failure process to remain on the current intensity level. Thus, the above costs, at a given point in time, are proportional to the "inertia force" *FI*(*t*) pertaining to the failure process:

$$FI(t) = C\frac{d^2\lambda(t)}{dt^2} = C\frac{2ab^2}{(bt+1)^3}$$

where C is an economic coefficient.

Thus, in order to reduce the failure intensity from the starting value *a*+*r* to $\lambda(t)$, the total cost to be sustained, *K*, is:

$$K = -\int_{a+r}^{\lambda(t)} FI(t)\,d\lambda(t) = \int_0^t \frac{2a^2b^3C}{(bt+1)^4}dt = \frac{a^2b^2C}{2}\left[1-\frac{1}{(bt+1)^4}\right]$$

The above cost can be seen as the overall work from the start of field testing to time *t*, required to move the failure intensity from *a*+*r* to $\lambda(t)$.

5.4 Model Application example

In this section it is shown how the proposed Lifecycle cost – Reliability model can be used to define, for a given system, a reliability target such that the reliability-related Lifecycle cost is minimized. More in detail, the model is applied to identify, for a fleet of 28 light rail vehicles, the optimum reliability target (mean distance between train removals).

The above fleet is expected to run 380.000 km per month, for 30 years. Based on the above mission profile, the cost per single service failure (A) is expected to be 3.000 Euros.

In order to evaluate the expected investment costs as a function of the steady state reliability objective via equation (5.6), it is known that the Average Production Unit Cost (APUC) of a light rail vehicle with a steady state mean distance between removals of 16.000 km is about 500.000 Euros. In addition, from estimates provided in Table 5-7, it is assumed that the initial distance from the steady state reliability target is about 15.6 times the planned target and the parameter b is assumed to be equal to 2,57E-06 km^{-1}. At last, in order to estimate the coefficient C, it is known that, for a similar project, the aggregate consolidation costs sustained to reach the planned reliability target were of about 1 million of Euros.

Based on the above information, in the following figure is plotted the Reliability-related Lifecycle cost for different steady state reliability levels.

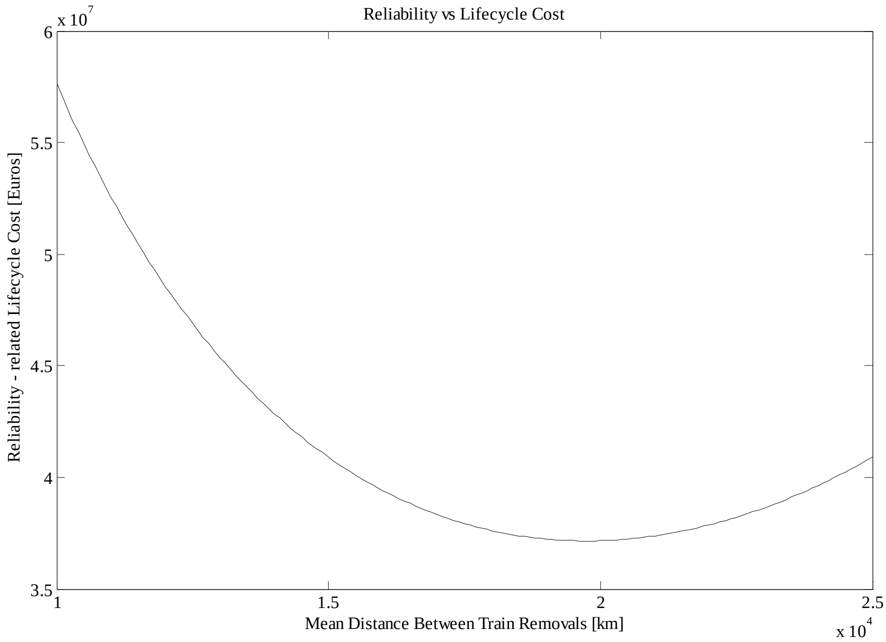


Figure 5- 3: Lifecycle Cost vs Reliability

Looking at the figure above, it is possible to conclude that for the considered fleet the minimum reliability-related Lifecycle cost can be obtained for a mean distance between train removals of about 20.000 km

5.5 Conclusions

A Lifecycle Cost – Reliability model, based on a three-parameters Hyperbolic failure intensity, has been defined. The model encompasses the relationships between reliability and the main cost components depending on reliability: costs due to service failures, investment costs required to reach a given inherent reliability level and efforts to be born to attain the planned reliability goals. For these costs, the innovative concept of inertia pertaining to the reliability improvement process has been used. Through an example, it has been shown that the model allows to set reliability targets such that reliability-related Lifecycle cost is minimized. However, the proposed model constitutes a first attempt to support, on the basis of quantitative and objective data, the definition of reliability targets for which the reliability-related Lifecycle cost is minimized since the early design phases. Nevertheless, further refinements are needed, especially to model efforts to be born during field testing and early stages of operation to attain the planned reliability levels, since in a typical industrial environment such a kind of cost information is not available with the level of detail required for modeling purposes.

6 CONCLUSIONS AND FUTURE RESEARCH

In this Thesis, a stochastic approach to manage Quality, Availability and Lifecycle Cost of Transportation systems has been outlined.

More specifically, a methodology to evaluate transport service quality has been defined. The proposed methodology relies on a set of innovative quality indexes based on the generic customer's waiting time. It allows to "on-line" monitor Quality of high frequency bus services by means of commonly available data, avoiding expensive data collection activities. In fact, a specific strategy has been formulated to obtain the waiting time distribution for a generic customer on the basis of commonly available data/information.

The adequacy of the above strategy and of the main working hypotheses has been checked by means of real data concerning the route 181 operated by A.N.M. The considered case study also allowed to highlight that the proposed approach enables to quantify the effects of service frequency and service regularity on the customers' waiting time. Moreover it allows to assess the current service quality level and to forecast how changes of the service frequency and/or regularity can affect quality perceived by customers. Since the proposed approach relies on the use of commonly available data, it can be successfully used by Transport Authorities as a self-assessment tool.

Moreover, a simple approach to assess the effects of the inherent repair time distribution, number of spare vehicles and number of maintenance crews on the operational availability of a fleet of vehicles has been defined. The considered approach allowed to easily handle complex Non - Markovian stochastic processes, and, in accordance with results found in Queueing theory, revealed that fleet operational availability may be significantly influenced by the inherent repair time distribution in both transient and steady phase. In addition, it has been shown that the proposed approach leads to operational availability evaluations that are more accurate than those usually performed in the industrial practice and that it may overcome some issues arising when Monte Carlo Simulation is used.

At last, a Lifecycle Cost – Reliability model, based on a three-parameters Hyperbolic failure intensity, has been formulated. Numerical procedures to obtain Maximum Likelihood and Minimum Chi-square estimates of the Hyperbolic model parameters have been developed as well. By means of reliability data concerning the first two years of operation of the Copenhagen Metro vehicles, it has been shown that the Hyperbolic model may be adequate to explain the failure rate behavior of complex repairable systems. Once identified a proper failure intensity model, a relationship between reliability and the main cost components depending on reliability has been developed. From this perspective, the innovative concept of inertia pertaining to the reliability improvement process has been used to model costs to be born, during the testing phase and the early stages of operation, to reach the planned reliability target. Through an example, it has been shown that the model allows to set reliability targets such that reliability-related Lifecycle cost is minimized. However, it has to be noted that the proposed model only represents a first attempt to explain, since the conceptual design stage, reliability-related Lifecycle cost. Of course, further refinements are needed and, under this perspective, availability of detailed operation, support and engineering cost data is of crucial importance.

REFERENCES

Abramowitz M, Stegun IA.Handbook of Mathematical Functions, With Formula, Graphs and Mathematical Tables, Eds., Dover publications, New York, 1972.

Aldous D, Shepp L (1987). The least variable phase-type distribution is Erlang, Stochastic models, 3.

Arnold BC, Groeveld RA. On excess life in certain renewal processes. J. Appl. Probab. 1981; 18 : 378–389.

Asgarpoor S., Mathine, M. (1997), Reliability Evaluation of Distribution Systems with Non-Exponential Downtimes, IEEE Transactions on Power Systems, 12 (2):579-584.

Billi C, Gentile G, Nguyen S, Pallottino S. (2003). Ripensando all'attesa alle fermate, Technical Report TR-03-17, Università di Pisa, Dipartimento di Informatica.

Billinton R., Li W., Reliability Assessment of Electric Power Systems using Monte Carlo methods, Plenum Press, New York, 1994.

Blum JR, Kiefer J, Rosenblatt M. Distribution free tests of independence based on the sample distribution function. Ann. Math. Statist. 1961; 32 : 485–498.

Bobbio A, Cumani A (1983). A Markov approach to wear-out modelling, Microelectronics and Reliability, 23.

Ceder A. Public Transit Planning and Operation: Theory, Modeling, and Practice, 1st ed. Butterworth –Heinemann : Oxford, 2007.

Ciardo G., German R., Lindemann C. (1994), A characterization of the stochastic process underlying a stochastic Petri net, IEEE Transactions on software engineering, 20 (7):506-515.

Çinlar E, Introduction to Stochastic Processes, Prentice-Hall, 1975.

Colini A., Erto P., Giorgio M., Testa A. (2009), A practical Markovian model of the availability and reliability of a mass transport service with non-exponential repair times, in : European Safety and Reliability Conference (ESREL), Prague, 1525-1532.

Cooper R.B., Niu S.C., Srinivasan M.M. (1998), Some refletions on the renewal-theory paradox in queuing theory, J. Appl. Math. Stochastic Anal. 11: 355-368.

Cox DR, Isham V. Point Processes, 1st ed. Chapman and Hall : London, 1980.

Cox DR. Renewal Theory, Methuen : London, 1967.

Cox D.R., Miller, H.D., The Theory of Stochastic Processes, Metheun & Co. Ltd, 1965.

Crow LH (1974). Reliability Analysis for Complex Repairable Systems, Reliability and Biometry, pp. 379 – 410, SIAM.

Daley DJ, Vere-Jones D. An Introduction to the Theory of Point Processes. Volume 1 : Elementary Theory and Methods, 2nd ed. Springer : New York, 2002.

dell'Olio L, Ibeas A, Cecin P (2011). The quality of service desired by public transport users, Transport Policy, 18:217-227.

Di Tommaso P, Giorgio M, Scuotto M, Testa A (2011). Operational Availability evaluation of Complex Systems with non-exponential downtimes, contributed talk in: Games and Decisions in Reliability and Risk, Belgirate 19-21 May http://www.mi.imati.cnr.it/conferences/gdrr11/talks.html.

Duane JT (1964), Learning curve approach to reliability monitoring, IEEE Transactions on Aerospace 2(2): 563-566.

Ellner PM, Wald L, AMSAA Maturity Projection Model, in Proceedings of Annual Reliability and Maintainability Symposium, 1995.

Ellner PM, Wald L, Woodworth J. A parametric Empirical Bayes Approach to reliability projection, in proceedings of Workshop on Reliability growth modelling, The Center for Reliability Engineering, 1998.

Erto P, Giorgio M, Lanzotti A. A statistical index of the quality of a mass transport service, in Proceedings of the Int'l. Conf. on Statistical Methods and Statistical Computing for Quality and Productivity Improvement, Aug. 1995, Seoul, 69-78.

Erto P, Giorgio M, Scuotto M (2010). Statistical Quality Indexes for a public bus service, in: Methods, Models and Information Technologies for Decision Support Systems, Pescara, 12-15 September , 199-202.

Erto P, Giorgio M, Scuotto M. Statistical Tools for evaluating mass transport service quality, submitted to Applied Stochastic Models in Business and Industry.

Erto P, Statistical model 'Life Cycle Cost-Reliability'of an Industrial Plant (in Italian), vol. 3, pp. 130 – 132 in “Atti della XXXIV Riunione Scientifica della Società Italiana di Statistica, Siena, maggio 1988, CEDAM, Padova 1988.

Erto P, Palumbo B (2005), Origins, Properties, and Parameters Estimation of the Hyperbolic Reliability Model, IEEE Trans. on Reliability, 54 (2): 276-281.

Fabrycky WJ, Blanchard BS. Life-Cycle Cost and Economic Analysis, Prentice-Hall, Englewood Cliffs, NJ,1991.

Fielding, J.P. (1979). A Transport Aircraft Reliability Formula, 2° National Rel. Conf., 2D/2/1-8.

Goldman T, Gorham R (2006). Sustainable Urban Transport: Four innovative directions, Technology in Society, 28:261-273.

Govil KK, and Aggarwal KK (1982). Cost versus reliability curve in reliability optimization problems, The QR Journal, pp. 71-72.

Gupta G.C., Srinivasa Rao T.S.S. (1998), On the analysis of single server finite queue with state dependent arrival and service processes M(n)/G(n)/1/K, OR Spektrum, 20:83-39.

Heimann DJ. The determination of transit system dependability, Proc. ARMS, 1979, Washington, 314-322.

Hounsell NB, Shretsha BP, Palmer S, Bowen T, D'Souza C. New Strategy Options for Bus Priority at Traffic Signals in London, presented at the European Transport Conference, October 2008, Noordwijkerhout (NL).

Johnson MA, Taaffe MR (1989). Matching moments to phase-type distributions: mixtures of Erlang distributions of common order, Stochastic models, 5.

Kao EPC, Smith MS. On excess, current and total life distributions of phase-type renewal processes. Naval Research Logistics 1992; 39 : 789-799.

Kho SY, Park JS, Kim YH, Kim EH. A development of Punctuality Index for Bus Operation. Journal of Eastern Asia Society for Transportation Studies 2005; 6 : 492-504.

Knessl C., Matkowsky, Schuss Z., Tier C (1987). The two repairmen problem: A finite source M/G/2 queue, SIAM J. Appl. Math. 47(2): 367-397.

Kvaloy JT,Lindqvist B, Malmedal H. A statistical test for monotonic and non-monotonic trend in repairable systems, in Proceedings of ESREL 2001 conference , 16–20 September 2001, Torino, 3 : 1563-2570.

Larson RC, Odoni AR. Urban Operations Research. Prentice Hall, Englewood Cliffs: NJ , 1981.

Law G, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. Statistics in Medicine, 11 : 1569-1578.

Locke C. A test for the composite hypothesis that a population has a gamma distribution Comm. Statist. Theory Methods 1976; 5 : 351-364.

Long AE, Forbes J, Hees J, Stouffer V (2007) Empirical relationships between reliability investments and lifecycle support costs, technical report.

Lukacs E. A characterization of the gamma distribution. Ann. Math. Statist. 1955; 26 :319-324.

Meeker WQ, Escobar LA, Statistical Methods for Reliability Data, John Wiley & Sons, 1998.

Mudholkar GS, Wilding GE. Two Wilson–Hilferty type approximations for the null distribution of the Blum, Kiefer and Rosenblatt test of bivariate independence. Journal of Statistical Planning and Inference 2005; 128 (1) : 31-41.

Neuts M.F., Matrix-Geometric Solutions in Stochastic Models: an Algorithmic Approach, Dover Publications Inc., 1981.

Neuts MF. Renewal processes of phase type. Naval Research Logistics Quarterly 1978 ; 25: 445-454.

Osuna EE, Newell GF. Control Strategies for an Idealized Public Transportation System, Transportation Science 1972; 6 (1) : 52-72.

Panglinan C, Wilson NHM, Dluger A. Bus Supervision Deployment Strategies and the Use of Real-Time AVL for Improved Bus Service Reliability, Transportation Research Record 2008 ; 2063 : 28-33.

Ross SM. Stochastic Processes, 2nd ed. Wiley : New York,1996.

Schassberger R.S., Warteschlangen, Springer-Verlag, Berlin, 1973.

Seger JK (1983). Reliability investment and Life-Cycle cost, IEEE Trans. on Reliability, 32 (3): 259-263.

Silcock DT. Measures of operational performance for urban bus services, Traf.Eng.Contr., 1981; 12 : 645-648.

Singh C., Billinton R., Lee S.Y. (1973), Reliability Modeling Using the Device of Stages, in: PICA, 22-30.

Stein WE, Dattero R. Sampling Bias and the Inspection Paradox. Mathematics Magazine 1985 ; 58 (2): 96-99.

Trivedi K., Probability and Statistics with Reliability, Queuing and Computer Science Applications, 2nd ed. John Wiley and Sons, New York, 2002.

Van Hagen M, Galetzka M, Pruyn A. Perception and evaluation of waiting time at stations of Netherlands Railways (NS), in Proceedings of European Transport Conference, Oct 2007, Noordwijkerhout (NL).

Wheatcroft P.A.C, (1985). DOCTOR — A Whole-Life Reliability Cost Model, International Journal of Quality & Reliability Management, 2 (1): 4 – 17.

Wilding EW, Govind SM. A gamma goodness of fit test based on characteristic independence of the mean and coefficient of variation. Journal of Statistical Planning and Inference 2008; 138 : 3813-3821.

ATTACHMENTS TO CHAPTER 1

Acknowledgements by the Ansaldo STS CEO

AnsaldoSTS

Una Società Finmeccanica

Via Paolo Mantovani, 3-5
16151 Genova
Tel. +39 010 6552111
Fax +39 010 655 2939

Attn : Michele Scuotto

HRMGPR/743

Genoa, 09th December 2010

Re : Ansaldo STS Innovation Award 2010

Dear Michele,

I have the pleasure to inform you that your proposal " MetroModelSim (MMS) was awarded the **Special Mention 2010** .

Congratulations!

Such special mention has been established this year as a reflection of the importance to build a portfolio of technologies and best practices to be exploited at the best by all our companies.

For your contribution, your company will award you with a cash prize of Euro 1250,00 to be paid to you in local currency with the next paycheck in accordance with the application legislation and company policy and procedures) .

Please share my congratulations with everyone supported you and your team in this task

Kind Regards

Sergio De Luca

Sergio De Luca

Ansaldo STS S.p.A.
Sede legale:
Via Paolo Mantovani, 3-5
16151 Genova
Tel. +39 010 6552111
Fax +39 010 6552939

Sede secondaria:
Via Argine, 425
80147 Napoli
Tel. +39 081 2431111
Fax +39 081 2432699

Direzione e coordinamento
Finmeccanica S.p.A.

Capitale sociale € 60.000.000,00
R.E.A. n. 421689
Iscrizione Registro delle Imprese di Genova
C.F. e P.I. n. 01371160662

www.ansaldo-sts.com

ATTACHMENTS TO CHAPTER 3

Bus Headways of the route 181 measured at Via Caravaggio stop (minutes)

1st week					2nd week					3rd week				
22	18	1	7	9	12	14	17	7	10	10	36	8	11	17
5	11	9	11	5	9	10	18	17	20	13	5	13	6	9
12	39	24	4	12	15	6	9	12	21	11	11	11	11	12
6	28	9	16	14	1	15	17	12	9	12	21	14	12	18
27	12	11	13	7	12	8	7	19	11	11	34	7	18	6
7	17	8	15	12	11	12	6	9	9	7	8	13	10	5
9	18	16	12	9	13	33	5	25	11	8	27	7	5	15
8	12	18	24	14	22	6	23	3	10	18	1	10	11	18
12	15	10	9	8	14	15	6	13	10	20	13	8	11	15
11		3	7	8	11	6	35	10	10	8	20	11	6	2
7		27	10	9	13	5	16	10	10	5		8	12	8
14		10	16	10	18		11	9	20	11		10	10	33
10		15	18	14	4			29	10			13	10	8
12				6	12							9	7	
12				8								11	11	
				23								9	15	
													9	

Pairs used for composite Gamma Hypothesis Testing

Pair No.	U	V	Pair No.	U	V	Pair No.	U	V
1	28	13.333	34	25	12.727	67	64	12.857
2	28	15.455	35	25	12.727	68	15	20.000
3	27	12.500	36	29	16.364	69	36	20.000
4	25	10.833	37	19	37.500	70	46	25.385
5	43	43.750	38	24	11.818	71	32	14.615
6	22	12.000	39	33	20.000	72	16	16.667
7	36	30.000	40	19	17.143	73	18	10.000
8	27	20.000	41	26	13.636	74	18	10.000
9	25	21.250	42	27	17.000	75	28	83.333
10	35	10.588	43	28	25.000	76	15	11.429
11	18	12.500	44	19	28.000	77	26	13.636
12	25	10.833	45	11	26.667	78	17	24.000
13	22	10.000	46	27	12.500	79	25	15.000
14	17	24.000	47	8	30.000	80	19	13.750
15	27	14.545	48	45	30.909	81	39	22.500
16	26	16.000	49	44	30.000	82	13	11.667
17	28	46.000	50	19	17.143	83	19	21.667
18	36	10.000	51	17	14.286	84	25	12.727
19	31	21.000	52	21	20.000	85	30	20.000
20	17	14.286	53	18	20.000	86	12	14.000
21	12	110.000	54	24	16.667	87	24	10.000
22	23	13.000	55	20	15.000	88	10	90.000
23	24	30.000	56	20	18.571	89	24	50.000
24	26	13.636	57	37	11.765	90	22	17.500
25	15	20.000	58	16	16.667	91	19	11.111
26	16	12.857	59	17	11.250	92	23	15.556
27	27	14.545	60	14	13.333	93	37	36.250
28	36	30.000	61	33	12.000	94	22	12.000
29	26	22.500	62	32	19.091	95	28	18.000
30	14	18.000	63	18	12.500	96	15	15.000
31	19	17.143	64	21	11.000	97	45	65.000
32	32	22.000	65	35	21.818	98	32	25.556
33	22	14.444	66	19	180.000			

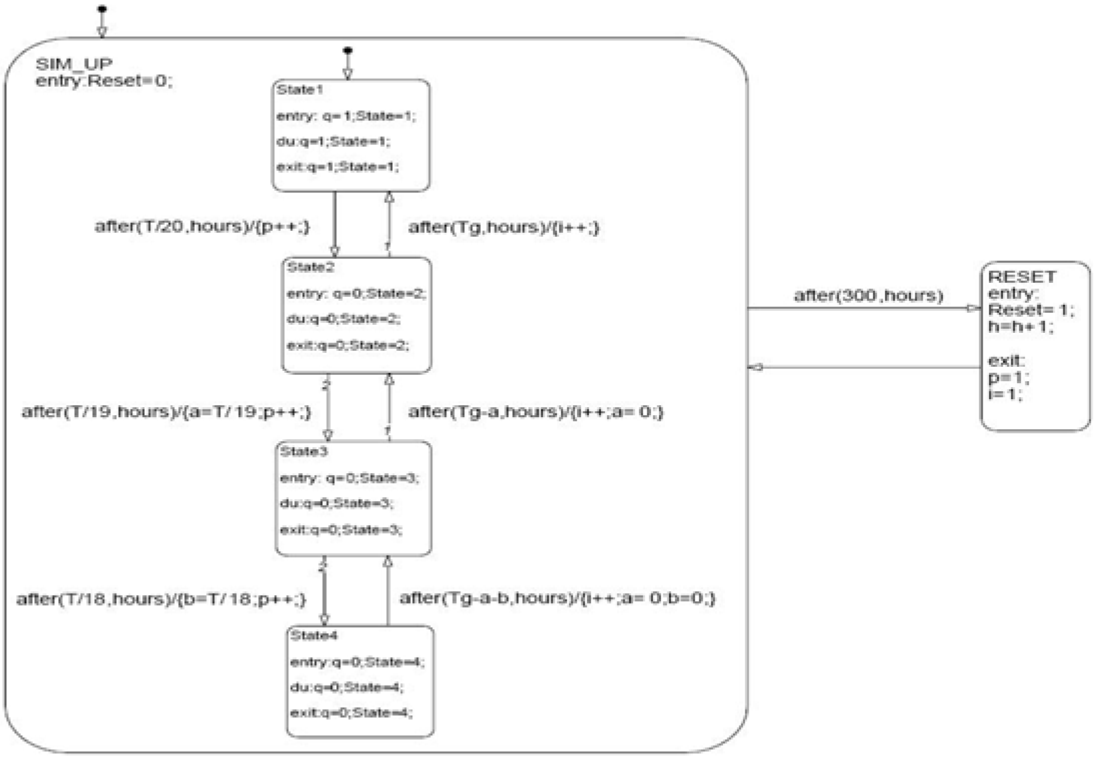
MATLAB® code for composite Gamma hypothesis testing

```
  %%Locke's Test based on BKR test of independece- BKR
statistic computation%%
  Data0=[
  ];
Data=Data0+0.5*ones(size(Data0));
y=randsample(197,1) %eliminate 1 datum since the sample is
odd
Data4Test=zeros(length(Data)-1,1);
  for i=1:y-1
      Data4Test(i,1)=Data(i,1);
  end
  for i=y:size(Data4Test)
      Data4Test(i,1)=Data(i+1,1);
  end
  Nums=[1:1:196];
  K=randsample(Nums,length(Data4Test))';
  Pairs=zeros(196/2,2); %make pairs
  for d=1:196/2
      Pairs(d,1)=Data4Test(K(2*d-1,1),1);
  end
  for d=1:196/2
      Pairs(d,2)=Data4Test(K(2*d,1),1);
  end
  UV=zeros(size(Pairs));
  for i=1:196/2
      UV(i,1)=Pairs(i,1)+Pairs(i,2);
  end
  for i=1:196/2
      w=Pairs(i,1)/Pairs(i,2);
      v=Pairs(i,2)/Pairs(i,1);
      if v>w;
          UV(i,2)=v;
      else
          UV(i,2)=w;
      end
  end
  scatter(UV(:,1),UV(:,2))
  UV(:,1)
  UV(:,2)
  N1=zeros(length(UV),1);
  for  k=1:length(UV)       %For each pair the number of
points in the 3rd orthant is quantified (axes pass for the
considered pair)
```

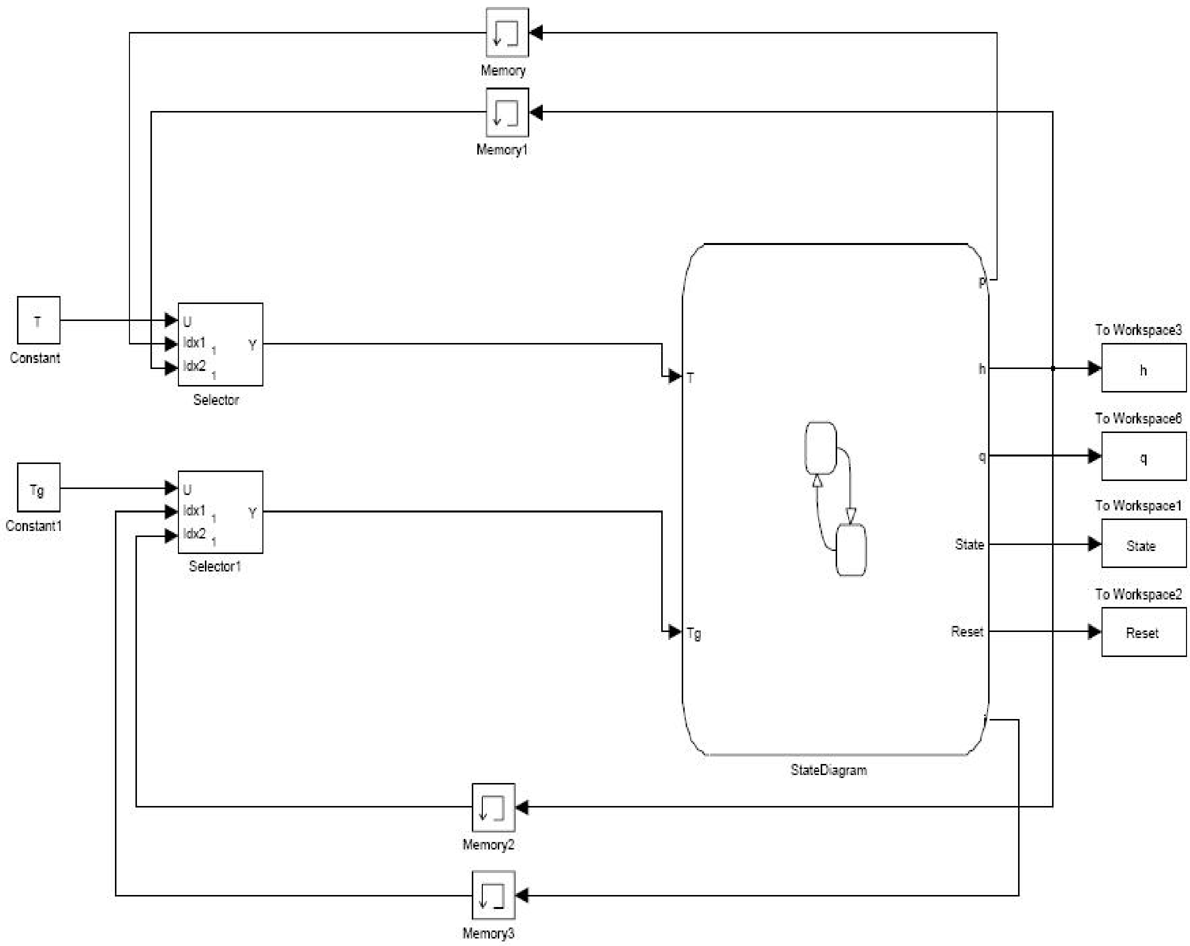
```
      for i=1:length(UV)
      if UV(i,1)<=UV(k,1) && UV(i,2)<=UV(k,2)
          N1(k)=N1(k)+1;
      end
      end
  end
  N1;
  N2=zeros(length(UV),1);    %for each pair the number of points in the 4rd orthant is quantified (axes pass for the considered pair)
  for k=1:length(UV)
      for i=1:length(UV)
      if UV(i,1)>UV(k,1) && UV(i,2)<=UV(k,2)
          N2(k)=N2(k)+1;
      end
      end
  end
  N2;
  N3=zeros(length(UV),1);  %for each pair the number of points in the 2nd orthant is quantified (axes pass for the considered pair)
  for k=1:length(UV)
      for i=1:length(UV)
      if UV(i,1)<=UV(k,1) && UV(i,2)>UV(k,2)
          N3(k)=N3(k)+1;
      end
      end
  end
  N3;
  N4=zeros(length(UV),1);   %for each pair the number of points in the 1st orthant is quantified (axes pass for the considered pair)
  for k=1:length(UV)
      for i=1:length(UV)
      if UV(i,1)>UV(k,1) && UV(i,2)>UV(k,2)
          N4(k)=N4(k)+1;
      end
      end
  end
  N4;
  S=zeros(length(UV),1);
  for i=1:length(UV)
      S(i,1)=(N1(i)*N4(i)-N2(i)*N3(i))^2;
  end
  nBn=(length(UV)^-4)*sum(S(:,1))
```

ATTACHMENTS TO CHAPTER 4

Stateflow chart for MCS



Simulink chart for MCS



MATLAB® code for Device of stages application : 1 maintenance crew

```
 %Computes fleet availability with erlang inherent repair
times
 stages=3;
 spares=1;
 dt=0.1;
 lambda=1/3000;
 MTTR=3;
 mu=stages/MTTR;
 Time=100;
 N=Time/dt;
 T=[0:dt:Time];
 nmax=3; %maximum number of failed trams
 Q=zeros(stages*nmax+1,stages*nmax+1);
 Q(1,2)=20*lambda*dt;
 for i=2:stages*spares+1
     Q(i,i+stages)=20*lambda*dt;
 end
 k=[1:1:(nmax-spares)];
 K=zeros(nmax-spares,stages);
 for i=1:stages
     K(:,i)=k';
 end
 c=(nmax-spares)*stages;
 R=zeros(1,c);
 R(1:stages)=K(1,:);
 for i=2:max(k)
 R((i-1)*stages+1:i*stages)=K(i,:);
 end
 for s=stages*spares+2:stages:stages*nmax+1-(2*stages-1)
     for j=1:stages
         Q(s+j-1,s+j-1+stages)=(20-R(s)+spares)*lambda*dt;
     end
 end
 for i=2:stages:stages*nmax
   for j=1:stages-1
         Q(i+j-1,i+j)=mu*dt;
     end
 end
 Q(stages+1,1)=mu*dt;
 for i=2*stages+1:stages:stages*nmax+1
         Q(i,i-2*stages+1)=mu*dt;
 end
```

```
for i=1:stages*nmax+1
    Q(i,i)=1-sum(Q(i,:));
end
p0=[1 zeros(1,stages*nmax)];
p=[p0;zeros(N-1,stages*nmax+1)];
for i=1:N
    p(i+1,:)=p(i,:)*Q;
end
Availability=zeros(N,1)
for i=1:N
    for j=1:stages*spares+1
    Availability(i)=Availability(i)+p(i,j);
    end
end
plot(T(2:length(T)),Availability)
Unavailability=ones(length(Availability),1)-Availability
```

MATLAB® code for Device of stages application : 2 maintenance crews

```
  %Computes the fleet availability of a fleet of 20 trams
with two maintenance crews
  %and a maximum of two spares

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%Variables%%%%%%%%%%%%%%
%%%
  stages=3;                                    %Erlang shape
parameter
  spares=0;                                  %Number of spare
Trams
  dt=0.1;
  lambda=1/3000;                  %Single Tram Failure Rate
1/h
  lambda1=(20+spares-1)*lambda; %Fleet failure rate with 1
tram under repair 1/h
  lambda2=(20+spares-2)*lambda;%Fleet failure rate with 2
Trams under repair
  MTTR=3;                  %Inherent Time To Repair of a tram
1/h
  mu=stages/MTTR;                              %Erlang Scale
Parameter
  mu1=mu;
  Time=40;                                       %Simulation
Time h
  N=Time/dt;
  T=[0:dt:Time];                                      %Clock
Variable
  nmax=3;                         %Maximum number of failed
Trams
  Q=zeros(stages+1+0.5*stages*(stages+1)*2);
%%%%%Transition Matrix%%%%%

  %%%%%%%%%%%Completed Repairs%%%%%%%%%%%%%%%%%%%%%%
  Q(stages+1,1)=mu*dt; %From last stage of repair (1 Tram
under Maintenance)
  K=zeros(stages,1);  %This Matrix Contains "From--->To"
states with 3 failed trams after repair completion
  K(1,2)=stages+1+(stages-1)*stages/2+1;
  for i=1:stages
      K(i,1)=1+stages+2*stages*(stages+1)/2-(i-1);
  end
  for i=2:stages
      K(i,2)=K(i-1,2)-(stages-(i-1));
```

```
  end
  K;
  Q(K(1,1),K(1,2))=2*mu*dt;
  for i=2:stages
      Q(K(i,1),K(i,2))=mu*dt;
  end
  for i=1:stages
      Q(stages+1+(stages)*(stages+1)*0.5-(i-1),stages+1-(i-
1))=mu*dt;
  end
  Q(stages*(stages+1)*0.5+stages+1,stages+1)=2*mu*dt;

  %%%%%Transitions Between repair Stages%%%%%%%%%%%%%%%%%%
Doubles=[stages+2;zeros(stages-2,1)];     %This     Matrix
Calculates states with 2 trams under repair and repairs at
same "stage"
  for i=2:stages-1
      Doubles(i)=Doubles(i-1)+i;
  end
  Doubles;
  for  i=2:stages-1  %Vertical  Transitions  2  trams  under
Repair
      for j=stages+1+1:(stages-1)*stages*0.5+stages+1
          if j==Doubles(i)
              Q(j,j+i)=2*mu*dt;
          else
          Q(j,j+i)=mu*dt;
          end
      end
  end
  for r=2:stages-1    %Vertical Transitions  3 Trams under
Repair
      for
s=stages+1+stages*(stages+1)*0.5+2:(stages)*(stages+1)*0.5*
2+stages+stages+1-2*stages
          if s==Doubles(i)+stages*(stages+1)*0.5
              Q(s,s+r)=2*mu*dt;
          else
          Q(s,s+r)=mu*dt;
          end
      end
  end
  for          i=stages*(stages+1)*0.5+stages+1+1:-1:stages
%Horizontal Transitions 2 Trams under repair
      Q(i-1,i)=mu*dt;
  end
  Cancel=zeros(stages,1); %The Previous Cycle contains not
allowed Tranitions. The matrix %contains states from which
transitions have to be erased
```

```
  Cancel(1,1)=stages+2;
  for i=2:stages
      Cancel(i,1)=Cancel(i-1)+i;
  end
  for i=1:stages
  Q(Cancel(i),Cancel(i)+1)=0;         %Erase  Not  Allowed
Transitions
  end
  Q(stages+2,stages+3)=2*mu*dt;
%Redefine the erased but allowed Transition (From stages
1,1 to 2,1)

  For
  i=stages*(stages+1)+stages+1:1:stages+1+stages*(stages+1)
*0.5+2
      Q(i-1,i)=mu*dt;
  end
  Q(stages+2,stages+4)=0;
  Cancel1=zeros(stages,1); %Same meaning of Cancel but for
the state with 3 failed Trams
  Cancel1(1,1)=stages+1+stages*(stages+1)*0.5+1;
  for i=2:stages
      Cancel1(i,1)=Cancel1(i-1)+i;
  end
  for i=2:stages-1
  Q(Cancel1(i),Cancel1(i)+1)=0;
  end
  Q(stages+1+stages*(stages+1)*0.5+1,stages+1+stages*(stage
s+1)*0.5+2)=2*mu*dt;
  Q(2,3)=mu*dt;                    %From state with one failed
tram
  for i=3:stages
      Q(i,i+1)=mu*dt;
  end

  %%%%%%%%%%%%%%%%%%%%%%%%%%%%Failures%%%%%%%%%%%%%%%%%%
%%%
  Q(1,2)=20*lambda*dt;
  L=zeros(stages,2);
  for i=1:stages
      L(i,1)=i+1;
  end
  L(1,2)=stages+2;
  for i=2:stages
      L(i,2)=L(i-1,2)+(i-1);
  end
  for i=1:stages
      Q(L(i,1),L(i,2))=lambda1*dt;
  end
```

```
for i=1+stages+1:stages*(stages+1)*0.5+stages+1
    Q(i,i+stages*(stages+1)*0.5)=lambda2*dt;
end
for i=1:stages+1+2*stages*(stages+1)*0.5
    Q(i,i)=1-sum(Q(i,:));
End
%%% State probability computation%%%%%%%%%%%%%%
p0=[1 zeros(1,stages+1+2*stages*(stages+1)*0.5-1)];
p=[p0;zeros(N-1,stages+1+2*stages*(stages+1)*0.5)];
for i=2:N
    p(i,:)=p(i-1,:)*Q;
end
Availability=zeros(N,1);
if spares==0
    Availability=p(:,1);
elseif spares==1
    for i=1:N
        Availability(i)=p(i,1);
    for j=2:stages+1
        Availability(i)=Availability(i)+p(i,j);
    end
    end
else
for i=1:N
        Availability(i)=p(i,1);
    for j=2:stages*(stages+1)*0.5+stages+1
        Availability(i)=Availability(i)+p(i,j);
    end
    end
end
plot(T(1:length(Availability)),Availability)
```

ATTACHMENTS TO CHAPTER 5

Grouped Failure data of the Copenhagen Metro Fleet

Km	Vehicle removals from service
106764	77
238534	76
404084	69
592224	71
793345	48
964446	43
2181895	250
2449481	50
2725183	37
3065337	45
3381437	49
3712025	45
4027063	27
4365609	27
4697897	45
5026259	27
5367903	28
5697052	35

Km	Vehicle removals from service
6043436	27
6385430	33

MATLAB® code for grouped data likelihood calculation

```
%computes -Loglikelihood and relative gradient of grouped
Hyperbolic data. T=interval bounds, N=failures in interval
% intensity=r+a/(bt+1) x(1)=asymptotic rate, r x(2)=b
  function [y,g]=Hypminusloglikebr(x,T,N)
  T1=[0;T];
  K=zeros(length(N),1);
  H=zeros(length(N),1);
  L=zeros(length(N),1);
  Ntm=sum(N(:,1));
  tm=T1(length(N)+1);
  logfin=log(x(2)*tm+1);
  for i=2:length(N)+1
      L(i-1)=real(log((x(2)*T1(i)+1)/(x(2)*T1(i-1)+1)));
  end
  for i=1:length(N)
      H(i)=real(x(1)*(T1(i+1)-T1(i))+((Ntm-
x(1)*tm)/logfin)*L(i));
  end
  for i=1:length(N)
      K(i)=real(N(i)*log(H(i)));
  end
  y=-sum(K(:,1));
  D1=zeros(length(N),1);
  for i=1:length(N)
      D1(i)=real(-(N(i)/H(i))*((T1(i+1)-T1(i))-
(tm/logfin)*L(i)));
  end
  g(1)=sum(D1(:,1));
  D2=zeros(length(N),1);
  for i=1:length(N)
      D2(i)=real(-(N(i)/(H(i)*logfin))*(Ntm-
x(1)*tm)*(T1(i+1)/(x(2)*T1(i+1)+1)-T1(i)/(x(2)*T1(i)+1)-
tm/(logfin*exp(logfin))*L(i)));
  end
```

```
  g(2)=sum(D2(:,1));
```

MATLAB® code for grouped data Chi-square calculation

```
%computes  chisquare  and  relative  gradient  of  grouped
Hyperbolic data. T=interval bounds,N=failures in interval
% intensity=r+a/(bt+1) x(1)=asymptotic rate, r x(2)=initial
rate, a x(3)=b
  function [y,g]= Hypchisquare(x,T,N)
  T1=[0;T];
  K=zeros(length(N),1);
  Nhat=zeros(length(N),1);
  for i=2:length(N)+1
      Nhat(i-1)=(x(2)/x(3))*log((x(3)*T1(i)+1)/(x(3)*T1(i-
1)+1))+x(1)*(T1(i)-T1(i-1));
  end
  Nhat;
  for i=1:length(K)
      K(i)=((N(i)-Nhat(i))^2)/Nhat(i);
  end
  y=sum(K(:,1));
  D1=zeros(length(N),1);
  for i=1:length(N)
      D1(i)=(N(i)^2)*(T1(i)-T1(i+1))/(Nhat(i))^2+(T1(i+1)-
T1(i));
  end
  g(1)=sum(D1(:,1));
  D2=zeros(length(N),1);
  for i=1:length(N)

D2(i)=(1/x(3)*log((x(3)*T1(i+1)+1)/(x(3)*T1(i)+1)))*(1-
(N(i)^2)/(Nhat(i)^2));
  end
  g(2)=sum(D2(:,1));
  D3=zeros(length(N),1);
  for i=1:length(N)
      D3(i)=(-
((x(2)/(x(3)^2))*log((x(3)*T1(i+1)+1)/(x(3)*T1(i)+1)))+(x(2
)/x(3))*((x(3)*T1(i)+1)/(x(3)*T1(i+1)+1))*((T1(i+1)*(x(3)*T
1(i)+1)-T1(i)*(x(3)*T1(i+1)+1))/((x(3)*T1(i)+1)^2)))*(1-
(N(i)^2)/(Nhat(i)^2));
  end
  g(3)=sum(D3(:,1));
```

ACKNOWLEDGEMENTS

My first grateful thanks go to my parents. I've to thank them for several reasons, but in this case I wish to especially thank them for the outstanding support they gave me during this academic experience and for their readiness to help me in getting through hard times.

I would also like to thank my sweet darling Marianna for her endless love and encouragement.

I would like to express my deepest gratitude to my supervisors, Professors Massimiliano Giorgio and Pasquale Erto, for their insightful advice and unreserved support. Every discussion with them always clarified confusions and inspired new ideas. In particular, I'd like to thank Professor Massimiliano Giorgio for being always ready in helping me and providing precious suggestions and encouragement to overcome some troubles I found along my way.

Other grateful thanks go to Ansaldo STS S.p.A. for providing me a PhD Scholarship and the possibility of entering an industrial reality operating at the leading edge of transportation solutions. In particular, I would like to thank the Ansaldo STS RAMS department. My special thanks go to Pietro Marmo and Ilaria Stanzione, for providing me data and suggestions based on their extensive practical experience. Every discussion with them allowed me to gain information and ideas to improve practical usability of methodologies presented in this Thesis. In this sense, also Titti Lamberti helped me a lot. I would also thank my officemates for the wonderful working environment they created.

Other special thanks go to my PhD colleagues, both former and present: Andrea, Ciro, Caterina, Mariangela, Flaviana, Gianluca. I want to thank them all for the beautiful times we spent in and outside the University and for all conversations we had. Maybe without their suggestions and support I would not have been able to cope with some difficult situations.

My final thanks go to all my sincere friends and to everyone I could learn something from.