On Compact Modeling of Coupling Effects in Maintenance Processes of Complex Systems

Vitali Volovoi[a,*], Rene Valenzuela Vega[a]

[a]*School of Aerospace Engineering, Georgia Institute of Technology, Atlanta GA, 30332, USA*

Abstract

While the complexity of modern engineering systems is significantly mitigated by modularization, the number of modules (*e.g.*, line-replaceable units) can still be large enough to pose a challenge of developing a coordinated maintenance policy that accounts for the coupling among individual maintenance schedules for each module. This paper focuses on opportunistic maintenance, induced failure, and other coupling mechanisms caused by competing risk phenomena. Modeling the maintenance process of an individual module or component, even if it includes modern condition-based considerations, can be described by a relatively small number of distinct states. In contrast, creating a system-level model that captures all relevant coupling leads to a state-space explosion; an implementation of such models is either very expensive or not feasible at all. To address this issue, the present paper explores the idea of developing component-level models that incorporate the aggregate effects of other components by providing a compact statistical representation of the combined influence on a given component of all other system components. This approach is somewhat analogous to the mean-field theory used in physics to avoid explicit description of pair-wise interactions. An analytical method based on asymptotic considerations is developed for combining the effects of multiple components into a single Weibull distribution (inspection intervals are assumed to be smaller than the failure scale). The accuracy of this approach is demonstrated by successfully representing the combined effect of two competing Weibull distributions as well as the combined effect of two competing lognormal distributions. In particular, it is shown that the proposed method provides a superior match for the combined distribution in the relevant time range as compared to standard methods of approximating a distribution (*e.g.*, matching moments or using the maximum likelihood estimate). The accuracy of representing opportunities using exponential distributions is explored as well.

1. Introduction

Historically, the reliability and safety of engineering systems has been approached from the opposite direction in at least two distinct dimensions. On the one hand, there is white-box vs. black-box dichotomy [1] where the distinction is based on whether the failure process of an entity is modeled with or without the explicit recognition of individual constituents (components) that comprise the entity. Here "component" refers

*Corresponding author

Email addresses: vitali@gatech.edu (Vitali Volovoi), rene.valenzuela@gatech.edu (Rene Valenzuela Vega)

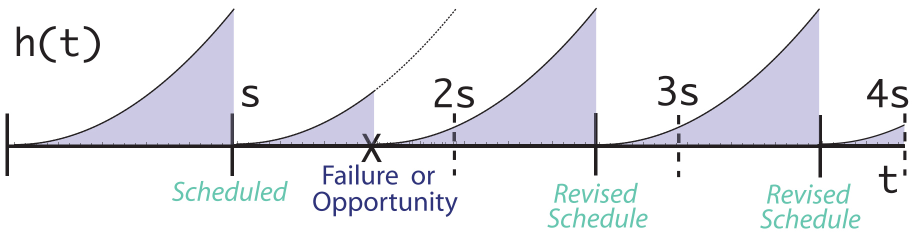

Figure 1: Age-based replacement maintenance schematics

to an elementary building block of a white-box (system) model, which can correspond to a lower-level entity if models are constructed hierarchically, or to the lowest level of the hierarchy, as determined by practical considerations (*e.g.*, individual modules, such as line-replaceable units or LRU). On the other hand, there is non-repairable vs. repairable dichotomy [2] with the former approach dealing with a single failure event of an entity, and the latter addressing repeated failure events, which assumes the possibility of partial of full recovery from failures. While a non-repairable entity is characterized by its lifetime distribution (*e.g.*, its cumulative distribution function of time to failure), repairable entity behavior is described by a point process, and so must be characterized differently, *e.g.*, using the rate of occurrence of failures (ROCOF), or the expected number of failures for a given time period. Any permutation of those choices translates into appropriate set of tools: for example, selecting the black-box direction can lead to accelerated testing techniques for non-repairable entities, and to modeling repairable entities by means of stochastic processes.

To make things more difficult (or interesting), the white-box approach entails selecting either the repairable or non-repairable option both at the system (output) and component (input) levels. Boolean algebra methods (*e.g.*, fault trees and reliability block diagrams) assume that both systems and the components comprising those systems are non-repairable. This symmetry between the inputs and outputs characterization (and associated simplicity and clarity) is perhaps one of the reasons for the popularity of those tools. In contrast, the use of superimposed processes [2] implies that both inputs and outputs are repairable entities. In this context, the state-space models that are subject of this paper can be classified as selecting repairable outputs and non-repairable inputs within the white-box (system) approach. This state-space representation is attractive in the context of modeling maintenance processes, as the impact of individual changes to the system (*e.g.*, change of the maintenance interval, or introducing a more reliable module) can be captured directly.

To illustrate these options, let us consider a system consisting of n identical components that exhibit an increasing hazard rate (as specified below) and are subject to age-based replacement [3]: all parts are replaced after a specified interval s. A failed component causes a total renewal of the system: all components (not only the failed one) are replaced with brand new ones; furthermore, we assume that the replacements occur instantaneously. Note that failures can shift the original schedule of replacements (see Figure 1). Here the outputs of interest are the expected numbers of scheduled replacements and failures. The number of opportunistic maintenance actions for a given component is simply the sum of the failures for all other components. A system three-state diagram is depicted in Figure 2 A: either scheduled maintenance takes place and all n components are replaced, or one of the components fails (and is replaced), while the

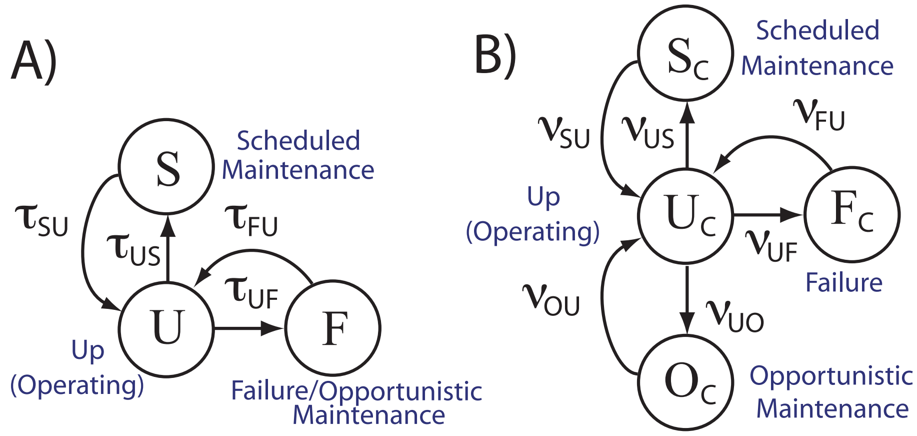

Figure 2: State-space representation of a simple system with identical components : A) Global (system) B) local (component)

rest (*i.e.,* $n-1$ components) are replaced as a result of opportunistic maintenance. Out of four transitions depicted, only one, τ_{UF}, is random: transitions τ_{FU} and τ_{SU} are instantaneous, while τ_{US} has a fixed delay or holding time. For a single component system ($n=1$) no opportunistic maintenance is possible, and the transition τ_{UF} occurs when the component fails. The timing of transition τ_{UF} is fully defined by the cumulative distribution function $F(t)$ for the failure of the component: once a new part is put in service t_1, the failure time is independent identically distributed in accordance with $F(t-t_1)$, as the replacement component has the same properties as the original one. The resulting model corresponds to a semi-Markov process where the time distribution of a transition to a new state depends on the current state together with the time the system spent in the current state (so-called holding time), but not on the previous history. The failure intensity, or hazard rate, is evaluated as $h(t) = f(t-t_1)/(1-F(t-t_1))$, where $f(t)$ is the probability density function corresponding to $F(t)$. An increasing hazard rate implies $h(t_3) > h(t_2), \forall\, t_3 > t_2 > t_1$.

As the number n of components in the system increases, the transition distribution τ_{UF} needs to be adjusted to represent failure of any of n components instead of one, but the structure of the model (Figure 2 A) remains unchanged. However, this is rather an unusual situation, as the real-life systems are likely to consist of distinct types of components with nonsymmetric interrelationships, which will cause the size of the system model to grow dramatically, as the system states are represented explicitly (while the state of each component can be inferred from the system state). We will refer to such models as "global" (Markov chains also fall under this category), as opposed to "local" models that describe explicitly the states of components, so that the state of the system can be inferred. Stochastic Petri Nets (SPN) represent an example of local modeling that can provide more compact models with large number of components [4, 5], and a version of SPN will be utilized in the next section to create such system-level models. However, even for local models, a description of the fully coupled behavior with hundreds (or thousands) of components is challenging, if not impossible.

The importance of multi-unit maintenance has been well recognized, and extensive literature surveys on the subject [6] make it clear that, as the number of components in the system increases, the level of details captured by the models decreases in order to make modeling overall complexity tractable. There is a compelling evidence that in both natural and engineering domains complex systems are unlikely to be fully coupled, as modular architecture provides clear advantages in developing desirable systems properties. The evolutionary advantage of so-called nearly decomposable systems has been demonstrated for biological systems [7], while similar processes were identified in the history of steam engine development [8]. These concepts are also explicitly employed in the design principles of computer systems [9] (including structured design [10]). It is therefore logical to take full advantage of the modular structure of the systems in modeling system failures. While in some situations a fully decoupled modeling of each unit is possible, coupling mechanisms can significantly impact the results. Nevertheless, single-unit models are still widely used in practice due to the overwhelming complexity of alternative methods of modeling, leading to sub-optimal selections of maintenance policies.

To this end, constructing decoupled (component) models with accurate representation of the coupling among components provides an intriguing alternative (*cf.* [11]). In the context of state-space modeling this implies specifying any appropriate additional states of the component, which is relatively easy to do if the coupling mechanisms are well understood. However, compact representation of the transitions between the component states due to coupling effects can be significantly more challenging, and addressing this challenge is the main focus of this paper. It is reasonable to assume that this representation can depend on the type of coupling, so a brief overview of the possible types of coupling is described next.

First, we note that in the context of state-space reliability modeling, coupling is closely related to dependency. It can be argued that one of the purposes of white-box modeling is explicit modeling of dependency mechanisms, so that the assumption of independence can be reasonably applied to individual state transitions, even if, from the black-box perspective, the corresponding system-level transitions appear to be statistically dependent. For example, there are several general sources of dependent state transition [12, 6]:

- Common-cause failures, either due to common environment (*e.g.*, cold temperatures during the Challenger launch that impacted both O-rings), or a common defect (*e.g.*, if components were obtained from the same batch made by the same manufacturer). In either case it is possible to explicitly model the common-cause failure by providing appropriate state transitions that impact several components simultaneously, while considering the remaining causes of failures for the components to be independent.

- Shared-load configuration, where the failure of a component can cause redistribution of the load for the components that remain in operation, and therefore change their failure distributions. From the white-box modeling perspective, this implies that several component states must be distinguished, each having distinct failure distribution associated with the transition to the failed state (in this context care needs be taken to account for aging processes in each state [13]). In addition, the consequences of the failure of the modeled component can be different as well from the system's perspective: shared-load can be considered a special case of redundant configurations, if a failure of another component does not alter the "load" for the component, and therefore its failure distribution.

- Opportunistic maintenance: if one of the components fails and needs a replacement or a repair, other components of the system can be serviced at the same time (an opportunity is created for conducting maintenance actions on those parts). For a variety of reasons (*e.g.,* the system has been taken off line or partially disassembled, technicians are available, etc.), servicing several components simultaneously is more efficient economically than servicing each component in isolation (hence the name "economic coupling").

- Induced (cascading) failures, when the failure of a component causes other failures (for example, a liberated blade in a gas turbine can break neighboring vanes).

- Logistics coupling usually refer to the logistics constraints where repair of a component is dependent on what happen to other components that might be competing for the same resources (spare parts, labor, or facilities). However, there is also the possibility of a positive coupling, where failure of another similar system provides an additional repair resource (*e.g.*, cannibalization of parts).

From the modeling perspective, these sources can divided into two scenarios: in the first (relatively simple) scenario the occurrence of an additional condition is time-independent (as it either happens or does not). Effectively, the relevant event occurs prior to the modeled time period. For example, components can belong to a higher-risk subpopulation, and a single value of the corresponding probability is sufficient (combined with the definitions of the lifetime distributions for both higher-risk subpopulation and for the rest of the component population). In contrast, the timing of another event is important in the second scenario so, strictly speaking, the full description of the timing of that event needs to be specified. It can be observed that the majority of the listed above coupling types follow this second scenario, where effectively there is a "race" between the internal and external events in the component model. We will refer to those scenarios as competing risks as this term is commonly used [14], but in the context of maintenance modeling the competition is not limited to risk-related transitions (*i.e.* some type of failure), but instead refers to any possible transitions for a given state (*e.g.*, timing of the first available spot among several repair queues).

Figure 2 B depicts a single component view of the described process. Here, parameters of a single component model are used for the failure transition ν_{UF}, while a new state corresponding to the opportunity is introduced along with the corresponding transition ν_{UO} associated with the failure of the other $n-1$ components. Parametric distributions are preferred in state-space modeling from the compactness perspective, assuming that their accuracy is assured. Significant further simplification can be achieved if constant transition rates are used: first, each transition is fully characterized by a single parameter, its constant rate λ; and second, if all state transitions have constant rates, the holding times are not affecting the chances of transition to the next state, and the resulting process is Markov (i.e., the chances of transitioning to a new state are fully determined by the current state). The analysis for Markov processes is significantly easier than for semi-Markov processes. A transition with constant rate λ follows an exponential distribution, whose cumulative form is given by $F_e(t) = 1 - e^{-\lambda t}$. In the case of failure transitions, λ is the reciprocal of the mean time to failure. Fixing the choice of the type of distributions for transitions for Markov models also facilitates hierarchical model construction and aggregation of states and transitions [15, 16].

Steady-state results often depend only on the mean parameters of the distributions associated with the state transitions, justifying the use of exponential distribution even if the underlying distributions are different (see, for example, the extensions of the Palm-Khinchin theorem, especially in the context of logistics [17]).

An additional argument for using exponential distribution is based on what can be characterized as the central limit theorem for repairable systems: in a complex repairable system with multiple components, failures form a homogeneous Poisson process [3] at the system level. The latter argument, however, assumes that there is no coordination among component failures. In practice, for many systems with clear aging or degradation patterns (*e.g.*, gas turbine engines), major inspections and overhauls impose an overall structure, and within each maintenance cycle the failure rate is generally increasing. Opportunistic maintenance has been extensively studied [18], [19] and [20], and in all of those studies the opportunities were assumed to follow exponential distribution [21], [22]. Opportunities can be caused by extraneous events that can be considered random, in which case exponential distribution is quite appropriate. However, the opportunities can be also caused by failure of components, which provides an impetus for investigating the performance of exponential distributions for such cases, and establishing the requirements for a distribution to adequately represent the combined effect of multiple components. Specifically, if the exponential distribution is inadequate, we seek to find an alternative compact representation of a given combined distribution. To this end, the natural step is approximate the targeted distribution with two-parametric distributions, and ensure that both first two moments (mean and standard deviation, respectively) of the targeted distribution are matched. Similarly, the targeted distribution can be sampled and the maximum-likelihood estimate (MLE) can be used to obtain the most appropriate distribution parameters [23].

The hypothesis explored in this paper is that an alternative selection of parameters can be more advantageous in the specific context of modeling for coupled maintenance scenarios that involve competing risk type of coupling. The goal of this paper is not only to predict the expected number of maintenance events for multiple components, but also to represent the effect of multiple components for component simulation in a compact fashion. The resulting representation can be used as modeling blocks for larger models (for example to evaluate the durations of outages and other relevant system-level effects).

In the context of reliability, Weibull distributions $F_w(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^\beta}$ are often used due to their flexibility of representing rates that can be either increasing or decreasing with time. The former correspond to the shape parameter $\beta > 1$ (*e.g.*, failures in deteriorating systems), while the latter correspond to the shape parameter $\beta < 1$. Conveniently, for $\beta = 1$, an exponential distribution is recovered, with the scale parameter θ representing the reciprocal of the transition rate. There are additional reasons for using Weibull distribution in system reliability, including ease of integration for finding moments, and the relationship to the "weakest link" mode of failure. The latter property is particularly pertinent: the Fisher-Tippett-Gnedenko theorem [24, 25] states that for a large number of identically distributed functions, the competing risk (i.e., the minimum of failure times) will converge to one of the three families of extreme value distributions (Weibull, Gumbel, or Fréchet). We explore whether in the case of non-identical distributions and when the number of distributions is finite (and even quite small) a Weibull distribution can provide a good approximation for the described combined effect. The justification for the developed method stems from the asymptotic considerations with respect to the small parameter $s/\theta \ll 1$, where s is the replacement interval and θ is scale of the failure distribution. Asymptotic considerations have been successfully used [26, 27, 28] to make approximations for reliable systems where the first terms of the Taylor series have been used. In contrast to that previous work, the present approach relies on the second order approximation.

2. Computational Methods for System Analysis

There are no closed-form solution existing for finite time horizons, but two numerical options exist: numerical integration of the corresponding renewal integral equations, or discrete-event simulation. Both options are briefly discussed next.

2.1. Renewal equations

For a single component with no scheduled maintenance (when the part is replaced only upon failure) the corresponding renewal process is well studied. While the integral equation For our purposes the following form of the corresponding integral equation is useful:

$$m(t) = f(t) + \int_0^t m(\tau) f(t-\tau) d\tau \tag{1}$$

Here the renewal density $m(t) = \frac{dM(t)}{dt}$, $M(t)$ is the expected number of renewals or renewal function, and $f(t)$ is probability density function [3, 29] (here we assume that derivative of renewal function exists). Efficient methods for the numerical solution of renewal equations exist using, for example, finite differences [30, 31]. The form of Eq. 1 provides a natural interpretation that is amenable to generalization: renewal at time t can occur either due to the first failure at that time with the probability density $f(t)$, or due to the repeated failure, where the previous renewal took place at time τ with the corresponding renewal density $m(\tau)$, and the chances of the failure $f(t-\tau)$ for the renewal time τ.

Let us now introduce scheduled replacements at interval s, so the renewal can be caused either by a failure or by scheduled replacement (if no failures occurred during interval s). Therefore, we can separate renewal density $m(t)$ into the two distinct parts: $m(t) = u(t) + w(t)$, where $u(t)$ and $w(t)$ represent renewal density due to failures and scheduled replacements, respectively. For the first cycle $0 < t < s$ Eq. 1 remains unchanged with $m(t) = u(t)$ and the renewal density due to scheduled replacements does not contribute ($w(t) = 0$). Noting that $R(s) = 1 - F(s)$ represents the chances that no failures will occur throughout the segment s, we conclude that in the vicinity of the first scheduled replacement $t = s$, the renewal density due to scheduled replacement can be described using Dirac delta function $w(t) = R(s)\delta(t-s)$. This is equivalent to stating that the expected number of renewals due to scheduled maintenance is zero for $t < s$ and is equal to $R(s)$ for $t = s$. For $s \leq t$, the renewal density $m(t) = u(t) + w(t)$ can be obtained from the following system of equations:

$$w(t) = R(s)m(t-s) = R(s)\left[u(t-s) + w(t-s)\right] \tag{2}$$

$$u(t) = \int_{t-s}^t m(\tau) f(t-\tau) d\tau = \int_{t-s}^t \left[u(\tau) + w(\tau)\right] f(t-\tau) d\tau \tag{3}$$

Here the Equation 2 states that in order for the scheduled maintenance to occur at time $t > s$ two conditions must be met: there was a renewal at time $t - s$ and there were no failures during interval s. Similarly, Equation 3 states that in order for the failure to occur at time t previous renewal must occur at some time $t-s < \tau < t$. In general, for n distinct components that can all cause renewal, we can introduce separate renewal densities to to the failure of each type of the component $u_i(t)$, $i \dots n$, and the corresponding equations have the following form:

$$w(t) = R(s)m(t-s) = R(s)\left[w(t-s) + \sum_{i=1}^{n} u_i(t-s)\right] \quad (4)$$

$$u_i(t) = \int_{t-s}^{t} m(\tau) f_i(t-\tau)\tilde{R}_i(t-\tau)d\tau = \int_{t-s}^{t} \left[\sum_{i=1}^{n} u_i(\tau) + w(\tau)\right] f_i(t-\tau)\tilde{R}_i(t-\tau)d\tau \quad (5)$$

Here $\tilde{R}_i(t-\tau)$ is the reliability of all other components, provided by the following expression:

$$\tilde{R}_i(t) = \prod_{i \neq j}^{n} R_j(t)$$

Effectively, in the presence of multiple components we use disjoint sets of possible outcomes, which allows us to sum the probability densities. As a result, we can interpret Eq. 5 as stating that renewal due to failure of component i occurs if the following three conditions are met:

1. Previous renewal took place at time τ, hence the term $m(\tau)$
2. Component i has failed at time t, while it has not failed in the interval $]t-\tau, t[$, hence the term $f_i(t-\tau)$
3. No other components failed during interval $]t-\tau, t[$, hence the term $\tilde{R}_i(t-\tau)$

Solving Eqs. 4,5 using finite differences can lead to highly accurate results as long as the selected time step is small enough. In this paper, s is discretized into 1000 segments, which leads to the results with the relative error usually not exceeding 10^{-5}.

2.2. Simulation

For a small system, a custom Monte Carlo simulation model can be easily developed, but using standardized graphical representation provides advantages for creating larger models, in particular for verification purposes. To this end, local or component-based representation of the state space is more convenient: instead of each state representing the system as a whole (as in Markov chains), states of individual components are described along with their interactions, so that the system state can be inferred from its component states. This is the essence of Stochastic Petri Nets (SPNs) [32],[33], where individual components are denoted with small circles (called tokens) and their transitions between states (called places) are denoted with solid rectangles. In the specific version of SPNs used in this paper [13], each token can change states independently of the others, unless their behavior is explicitly coupled by means of inhibitors or enablers (denoted as an arc terminated at a transition with a hollow or solid circle, respectively). A transition is disabled by an inhibitor if there are enough tokens in the place where the inhibitor originates, while an enabler acts in the opposite way (the corresponding transition is disabled *unless* there are enough tokens in the corresponding place).

Figure 3 depicts an SPN model for $n = 3$ components. On the left, the operating configuration is shown, while on the right is the situation when the failure of one of the components occurs: the corresponding token is moved from the place "Operating"

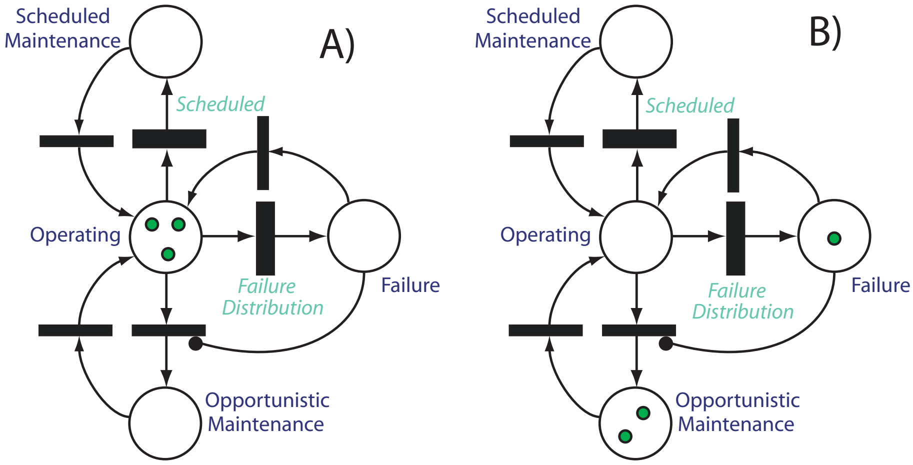

Figure 3: SPN coupled model with three components: A) system is in operating state B) system is in a state when one of the components failed and the rest are undergoing opportunistic replacements

to "Failure," thus enabling the transition of the rest of the tokens to the place "Opportunity." Due to the fact that the state transitions are local, SPNs are not restricted conceptually to constant transition rates, although in practical terms this implies that the solutions are obtained using Monte Carlo simulations. Following a common SPN convention, immediate transitions (without time delay) are depicted with narrow rectangles. Note that both the global state representation, Figure 2, and the local one, Figure 3, depict the same process: Figure 3, A represents the operating state of the system, while B corresponds to failure/opportunistic maintenance in Figure 2. The compactness of the model shown in Figure 2 is somewhat deceptive: generally global models do not scale well with the number of components, and local models are more compact [13] (but still very large for a system with a large number of components).

2.3. Decoupled modeling

Further simplification can be achieved by creating separate models for each component (see Figure 4 for an SPN model and compare for a generic state-space representation Figure 2 B) and then combining the results. Again, the simplification becomes apparent only as the complexity of the problem increases. Indeed, in the considered simple case, the coupled model (Figures 2, 3) has three and four total states, respectively, as opposed to three separate models (one per component) that each have four states (Figure 4). It is important to note, however, that the situation is reverse in practical situations where components are distinct (and so the failures of each component need to be counted separately, each requiring a separate state). In general, coupled representation implies explicit modeling of $n^2 - n$ pair-wise coupling for n components (coupling in maintenance does not have to be symmetric). In contrast, there will be only n separate component models, with only the combined effect of all other components being represented. This is an attractive strategy, provided that the coupling among the components behavior is either negligible, or adequately captured. In the case of identical components due to symmetry only one such model is required, and the main question is how to represent the opportunity distribution (see the corresponding transition in Figure 4).

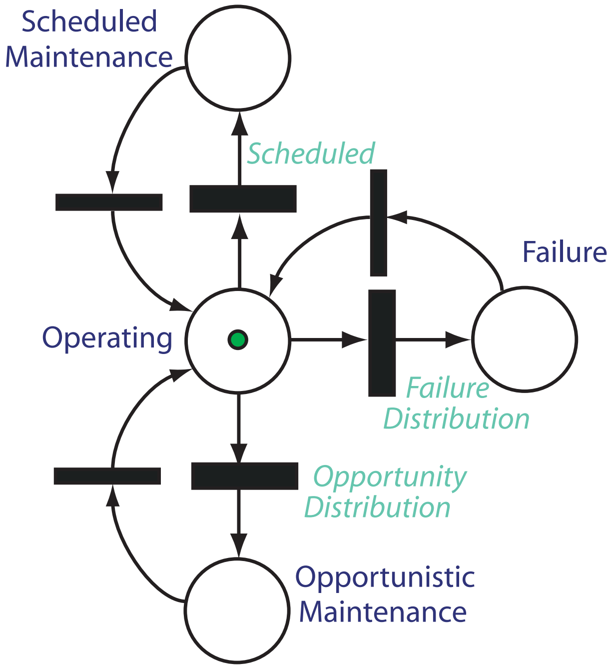

Figure 4: SPN Model for a single component with opportunistic maintenance

2.4. Components with the same shape

The main goal of the paper is developing a procedure for a compact representation of a cumulative distribution representing the opportunity for renewal due to failure of multiple parts. First, let us consider a situation where all n components $X_1 \ldots X_n$ of a given system follow Weibull distributions with the same shape parameter $\beta_i = \beta$, while scale parameters are allowed to be different: θ_i. Then the opportunity for the first component $i = 1$ stems from the failure of components $i = 2 \ldots n$, and the corresponding cumulative distribution function for the opportunity can be calculated as

$$\begin{aligned}\tilde{F}_1(t) = 1 - \tilde{R}_1(t) = 1 - \prod_{i=2}^{n} R_i(t) = \\ = 1 - \exp\left[-t^\beta \sum_{i=2}^{n} \frac{1}{\theta_i^\beta}\right]\end{aligned} \tag{6}$$

Here $R(t) = 1 - F(t)$ denotes respective reliability functions. As a result, the combined effect of the opportunity is represented by a Weibull distribution with the following the scale and shape parameters:

$$\tilde{\beta}_1 = \beta; \qquad \tilde{\theta}_1 = \frac{1}{\left[\displaystyle\sum_{i=2}^{n} \frac{1}{\theta_i^\beta}\right]^{\frac{1}{\beta}}} \tag{7}$$

So, if all components have the same shape, then the combined failure distribution has the same shape, and the scale is uniquely defined analytically.

2.5. Example 1: Three identical components

The simplest case where the opportunity representation is non-trivial is for systems that consist of three components, so the combined effect of failure of two components needs to be represented. Let us consider a system that operates for time $T = 4$. Each component follows a Weibull distribution with the same shape $\beta = 3$ and scale $\theta = 1$. Based on Eq. 7 we can conclude that the opportunity for each component will follow a Weibull distribution with the same shape parameter $\beta = 3$ and scale $\theta = 1/\sqrt[3]{2}$. We consider a full range of fixed replacement intervals s that span the values from very conservative s that are significantly smaller than the mean time to failure, to the large values of s where the effect of scheduled maintenance is negligible, as the failure will always occur first. As shown in [3], the values of s that are integer fractions of time T correspond to the discrete jumps of the total cost function that combines the cost of failures and scheduled replacements (the jumps are associated with an extra scheduled replacement right before the end of the system life T), while for all other values the cost function is a smooth function of s. As a result, those values provide good characterization of the overall cost function, and so they will be utilized in this and the following examples. There is no replacement taking place at time T, as the system has reached the end of its life. This last point is usually irrelevant for systems that operate for long periods of time, *i.e.,* when the scheduled replacement interval s is much less than T, but it can be significant when the time horizon T is of the same order as s.

As expected, if the opportunity is provided by the calculated Weibull function, solving both coupled and component-based models by discretizing Eqs. 4, 5 leads to identical results within the considered accuracy of 10^{-5} (here and below, the results from finite-difference solutions are based on 1000 time steps for a replacement interval s). Next, we use exponential distribution to represent the opportunity instead of using the Weibull distribution with the shape parameter $\beta = 3$. Figure 5 shows the results for expected failures, opportunities, and scheduled replacements for various numbers of scheduled maintenance segments s. Two methods are compared for estimating the scale of the opportunity:

1. Balanced: scale parameters of the opportunity distribution are not pre-calculated, but rather obtained by means of iteration, to ensure that the number of opportunities matches the corresponding number of failures (since each failure entails two opportunities). The convergence of this procedure is assured due to the fact that the involved mapping is monotonically decreasing. Indeed, the more opportunities, the more frequent the maintenance actions, and the fewer failures occur for components with an increasing failure rate. As a result of this convergence, an optimum scale parameter is obtained with respect to failures and opportunities: any improvement in the prediction of opportunities will incur the deterioration of the prediction of failures, and *vice versa*. Note that the while it is relatively easy to obtain this solution for identical components (as this a one-dimensional problem), the scalability of this approach is far from being straightforward, as the system of nonlinear equations needs to be solved.
2. Scale-matched (SM): A more computationally efficient and scalable approach consists of evaluating the combined chance of opportunity considering failures of other components for the relevant replacement interval (independent of the failures of the component itself) and then finding the opportunity scale so that combined number of failures during the replacement interval is matched.

Figure 6 shows the relative errors associated with those approximations. One can observe that as long as the failures are relatively infrequent in comparison to the maintenance intervals, both exponential approximations of the opportunity work reasonably well. When replacements are less frequent, the balanced approach generally over-predicts

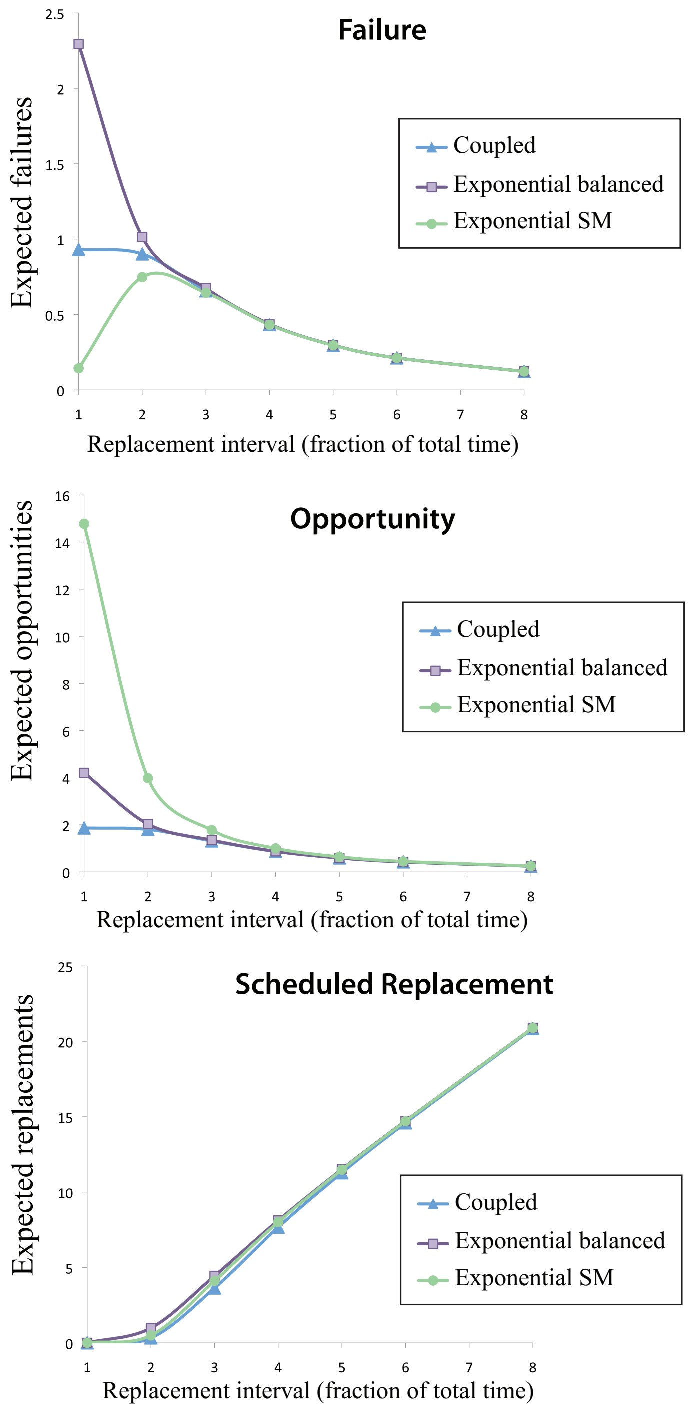

Figure 5: Expected failures (top), opportunities (middle), and scheduled replacements (bottom) for three identical components following Weibull with $\beta = 3$ and $\theta = 1$ for time $T = 4$ as functions of the number of scheduled replacements. The coupled model is compared to the sum of three component models with opportunities approximated by exponential distributions. The scale is found either by balancing total opportunities and failures ("balanced"), or by matching the scale of the distribution with the total number of expected failures during the full replacement interval ("SM.")

both failures and opportunities, while the scale-matched method actually under-predicts failures while over-predicting opportunities. As a result, there is no assurance that the prediction of the scheduled replacements by the balanced method is optimal: one can observe in Fig. 6 that scale-matching actually provides a slightly smaller error. Next, we explore the possibility of selecting Weibull shape parameters for modeling opportunities in order to further improve the accuracy of the approximation, effectively trying to generalize Eq. 7 for the components with different shapes as well as other distributions.

3. Proposed approach

3.1. Winning race ratio

We can still calculate the opportunity for the first component (that is, the combined chances that one of other components fails) using the general formula for the cumulative distribution $\tilde{F}_1(t)$:

$$\tilde{F}_1(t) = 1 - \prod_{i=2}^{n} R_i(t) \tag{8}$$

Taking a derivative of Eq. 8 we can obtain the corresponding probability density function $\tilde{f}_1(t)$ and evaluate the chances of the opportunity as

$$O_1(s) = \int_0^s R_1(t)\tilde{f}_1(t)dt \tag{9}$$

A direct use of this formula can be complicated by the fact that for a large number of components the integration can be somewhat involved. However, the main difficulty is that the formulae provide only the odds of the first action. As renewals take place, the schedule of replacements can change (see Figure 1), so to calculate the mean number of failures or opportunistic maintenance for some interval T_w (say, the warranty period), either the renewal Eqs. 4, 5 need to be solved, or a simulation needs to be used since the number of maintenance intervals that would “fit” into T_w is unknown *a priori*.

There are two distinct possibilities for an opportunity to occur for a given interval s (see Figure 7):

A: The component would not fail on its own during the interval s, while some other component does fail during this interval. This probability is easily obtained using Boolean operations if failures are independent and individual distributions are known.

$$A_1(s) = R_1(s)\tilde{F}_1(s) \tag{10}$$

B: The component would fail on its own during interval s, but some other component also fails earlier during this interval. In other words, the component loses the race to failure to another component (in the first scenario the component does not enter the race at all). If we denote the odds of this component’s “losing” this race to the other component with γ_1, the following relationship can be written:

$$B_1(s) = O_1(s) - A_1(s) = \tilde{F}_1(s)F_1(s)\gamma_1(s) \tag{11}$$

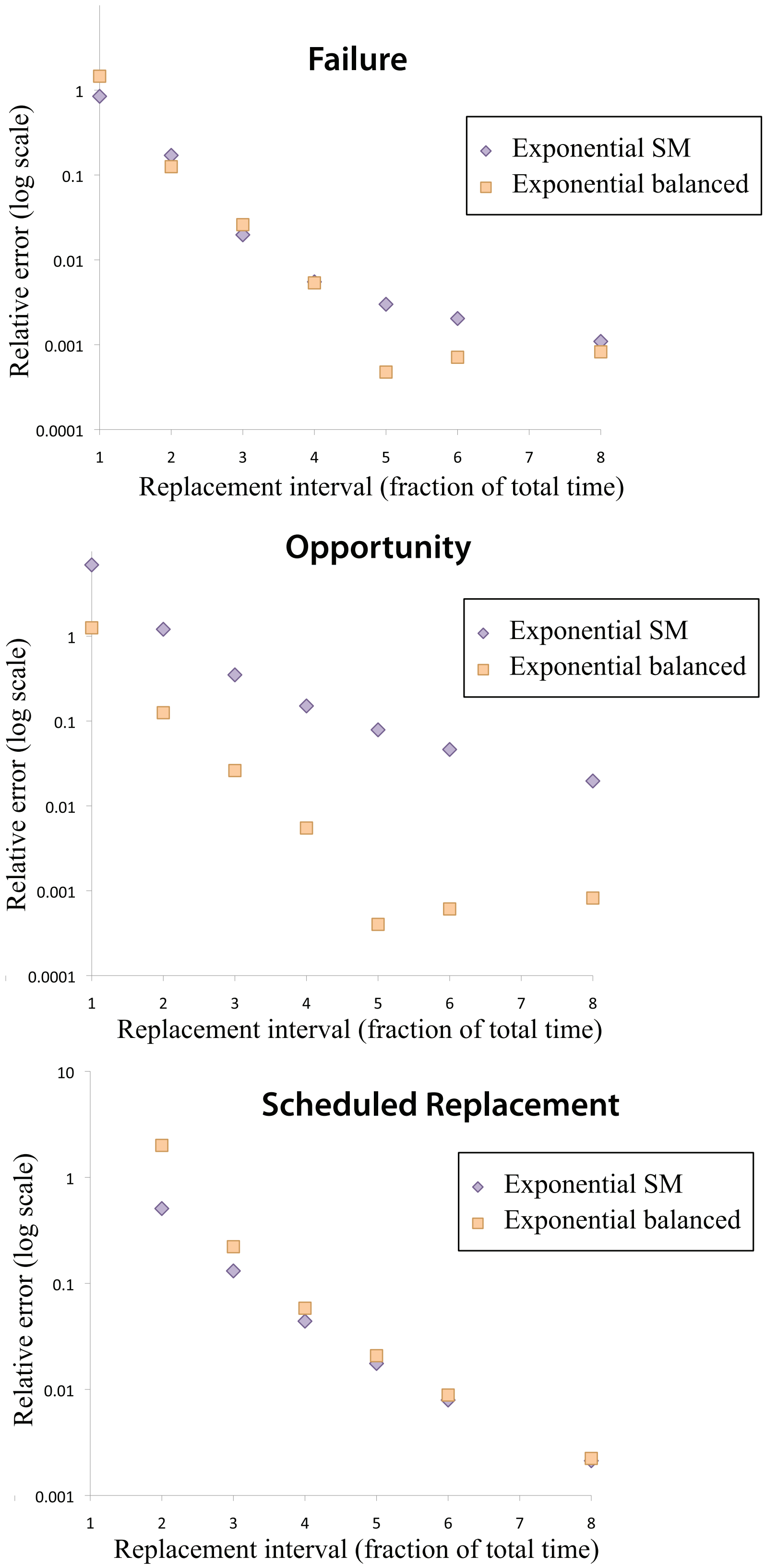

Figure 6: Absolute values of relative errors due to approximation of opportunities using exponential distribution. Logarithmic scale is used for the errors due to large variability. Failures (top), opportunities (center), and scheduled replacement (bottom) are evaluated for both balanced and scale-matched (SM) approximations.

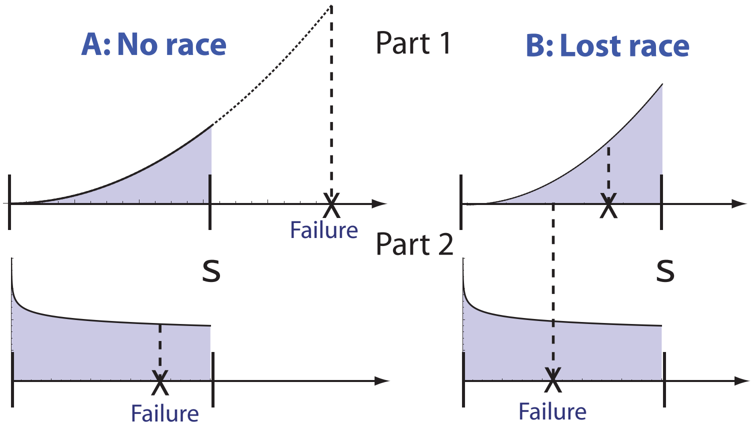

Figure 7: Two cases for part 1 losing a “race”. A - Part 1 does not enter the race, B - Part 1 enters the race but loses it to some other part (Part 2)

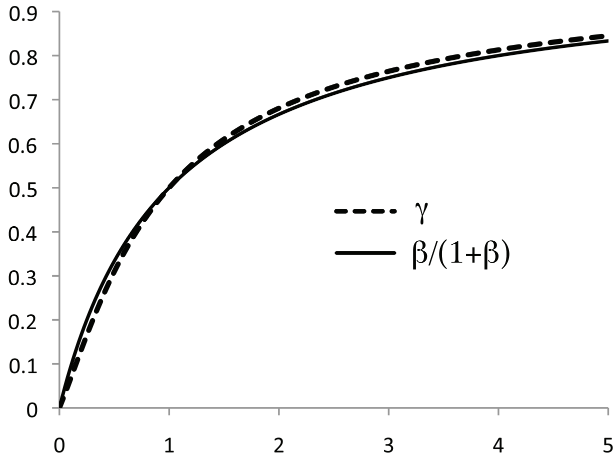

Figure 8: Shape sensitivity of γ_{12} for one of the functions with $\beta_1 = 1$ and varying β_2 of the other function. The scale $\theta = 5$ for both functions. Asymptotic approximation (Eq. 16) is shown by the solid line.

The second scenario is less likely to occur for the practical range of values, but its probability is non-negligible. If general rules regarding the chances of losing or winning the race are developed, then a good approximation of those chances will facilitate finding an equivalent function that represents the action of many components simultaneously. Since the second scenario is relatively less likely, and the odds of first scenario are easy to compute, the overall accuracy is expected to be quite good.

3.2. Odds of “winning” for Weibull functions

One can observe that for Weibull distributions, the smaller the shape function, the more likely the race will be “won.” This dependency can be captured parametrically: Figure 8 shows γ, the chances of “winning” the race, for the distribution where the first component has the shape parameter $\beta_1 = 1$, while the shape of the opportunity varies (the scale parameter is fixed). Effectively, if competing failures both occur during a given interval, the more distribution is skewed toward the beginning of the interval, the larger the chances that this distribution will win the race.

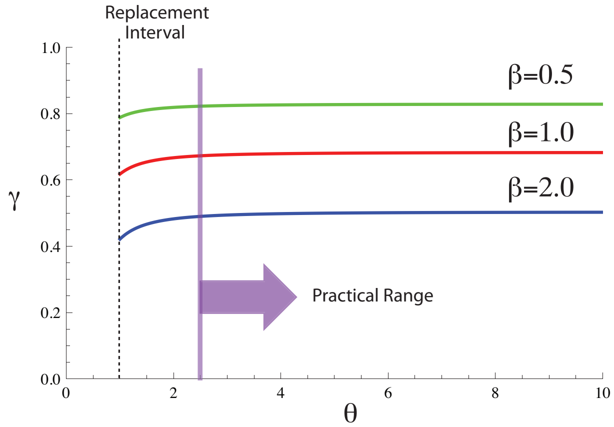

Figure 9: Scale sensitivity for race ratio γ when the first component follows Weibull distribution with shape $\beta = 2$ and scale $\theta = 5$ for different shape values. The replacement interval is unity.

What is more remarkable is that, for the practical range of scale parameters (which is unlikely to be less than $2.5s$, as this would lead to too many failures), γ is quite insensitive to the scale parameter, and can be well approximated by a constant value that is a function only of the respective shape parameters. In particular, this implies that in that range γ, as defined in Eq. 11, can be considered independent of s. In fact, this scale independence for Weibull distributions can be confirmed using asymptotic considerations, as shown next.

3.3. Asymptotic Considerations

Let us consider s to be small as compared to the the failure scales θ_i, so we can introduce small parameters $\epsilon_i = \left(\frac{t}{\theta_i}\right)^{\beta_i} \ll 1$ for $0 \le t \le s$ and evaluate γ:

$$\gamma = \frac{-\int_0^s R_1(t)\frac{dR_2(t)}{dt}dt - R_1(s)(1-R_2(s))}{(1-R_1(s))(1-R_2(s))} \tag{12}$$

The first-order terms were studied for highly reliable systems [26, 27, 28], but they are not sufficient to evaluate γ. Let us include the second-order terms with respect to the introduced small parameters and evaluate the ratio γ. First, we can observe that $R_i = e^{-\epsilon_i} \approx 1 - \epsilon_i + \frac{\epsilon_i^2}{2} \ldots$. The integral in the nominator can be evaluated with accuracy up to the second order terms:

$$-\int_0^s R_1(t)\frac{dR_2(t)}{dt}dt \approx \int_0^s \left(1 - \epsilon_1 + \frac{\epsilon_1^2}{2}\right)\frac{\beta_2}{t}\left(\epsilon_2 - \epsilon_2^2\right)dt \approx$$
$$\approx \int_0^s \frac{\beta_2}{t}\left(\epsilon_2 - \epsilon_2^2 - \epsilon_1\epsilon_2\right)dt = \int_0^s \frac{\beta_2}{t}\left(\frac{t^{\beta_2}}{\theta_2^{\beta_2}} - \frac{t^{2\beta_2}}{\theta_2^{2\beta_2}} - \frac{t^{\beta_1+\beta_2}}{\theta_1^{\beta_1}\theta_2^{\beta_2}}\right)dt =$$
$$= \frac{s^{\beta_2}}{\theta_2^{\beta_2}} - \frac{s^{2\beta_2}}{2\theta_2^{2\beta_2}} - \frac{\beta_2}{\beta_1+\beta_2}\frac{s^{\beta_1+\beta_2}}{\theta_1^{\beta_1}\theta_2^{\beta_2}} = \epsilon_2 - \frac{\epsilon_2^2}{2} - \frac{\beta_2}{\beta_1+\beta_2}\epsilon_1\epsilon_2 \tag{13}$$

Similarly, neglecting the terms that are higher than the second order in the remaining terms in Eq. 12:

$$R_1(s)(1 - R_2(s)) \approx \epsilon_2 - \epsilon_1\epsilon_2 - \frac{\epsilon_2^2}{2} \tag{14}$$

$$(1 - R_1(s))(1 - R_2(s)) \approx \epsilon_1\epsilon_2 \tag{15}$$

Substituting these expressions into Eq. 12, and simplifying, we obtain the following asymptotic expression:

$$\gamma \approx \frac{\beta_1}{\beta_1 + \beta_2} \tag{16}$$

Figure 8 shows the numerical accuracy of this approximation.

This observation leads to the following strategy for deriving an equivalent Weibull distribution that represents the opportunity:

1. Evaluate the chances of "winning" the race for one of the other components:

$$\hat{F}_1(s) = 1 - \prod_{i=2}^{n}(1 - \gamma_{1i}F_i(s)) \tag{17}$$

   Here pair-wise winning odds, γ_{1i}, are calculated based on some reference scale parameters, and so can be pre-calculated (if only Weibull functions are involved, then Eq. 16 can be used).
2. Divide the chances obtained in the previous step by $\tilde{F}_1(s)$ (see Eq. 8) to yield an estimate of the combined winning ratio $\hat{\gamma}_1(s) = \dfrac{\hat{F}_1(s)}{\tilde{F}_1(s)}$.
3. Find the appropriate shape parameter $\hat{\beta}_1(s)$ given $\hat{\gamma}_1(s)$ (for Weibull distributions Eq. 16 can be used).
4. Determine the scale parameter by matching the chances of failure for the total interval $\tilde{F}_1(s)$ given the shape $\hat{\beta}_1(s)$.

The key consideration is that if renewal of the system occurred, then the remaining interval is even smaller as compared to the scales of the failure functions, so the asymptotic assumptions hold.

4. Results

4.1. Testing the procedure: Example 2

We can note that the smallest non-trivial number of components is three, as for two-component systems opportunity is provided by another component, and there is no need to combine the distributions. As described in example 1, if all three components are identical then the combination is trivial. Let us therefore consider a system where two components are identical, but the third component is distinct: $\beta_1 = \beta_2 = 4$, $\theta_1 = \theta_2 = 3$; $\beta_3 = 2$, $\theta_3 = 5$. Then the opportunity for the third component follows Weibull distribution with $\beta_{1o} = 4$ and calculated using Eq. 7 $\theta_{1o} \approx 2.5227$. For the other two components we use the developed method, and by then compare the results for the first component by evaluating the coupled model, subtracting the results for the third component, and dividing the results by two (since the first two components are identical). For the total

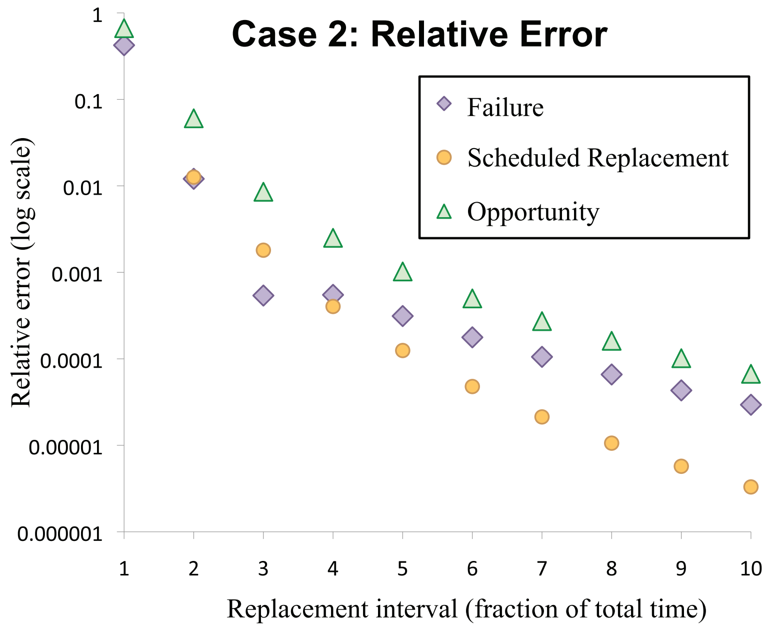

Figure 10: Relative errors for Case 2 as a function of the number of replacement intervals per time $T = 5$

time $T = 5$, the relative errors are shown in Fig. 10, demonstrating that for a broad range of replacement intervals, $s < T/2$ are not exceeding 1%, which for most of the application is sufficient. The natural question arises regarding the relative importance of the shape selection, as one can envision the possibility that for any shape parameter, the accuracy will be reasonable as long as the scale-matching is performed, and perhaps other shape parameters might provide even better accuracy. The results shown in Figure 11 directly address this question by varying the shape parameter parametrically for a fixed replacement interval $s = T/5 = 1$, and matching the scale for every shape. One can observe that the shape selected based on the proposed procedure $\beta \approx 4.2$ is indeed quite close to optimal, and the sensitivity with respect to the shape parameter is not trivial. We note that the perfect match would require that all three curves in Figure 11 intersect in a single point with the zero ordinate. It is also interesting to observe a non-linear dependence of the failure prediction errors (as exponential distribution can provide a smaller error that $\beta = 1.5$.)

4.2. Example 3: Weibull distributions with different shapes

Let us consider a three-component system where all three components have different shapes, and check that the developed procedure provides reasonable accuracy. The first component follows Weibull distribution, with $\beta_1 = 3$ and $\theta_1 = 4$, while the maintenance interval $s = 1$. The opportunity is provided by the failures of two other components that have the following Weibull parameters: $\beta_2 = 2$, $\theta_2 = 5$, $\beta_3 = 4$, $\theta_3 = 3$.

We can test the procedure by estimating the total number of events of interest for the time interval $T = 5$. The SPN models described in Figures 3, 4 have been used to independently verify the procedure, but the results presented here are obtained using the finite-difference method due to their superior accuracy.

One can observe (see Figure 12) that up to $s = T/3$ the accuracy of the method is quite good, and provides about an order of magnitude of improvement in terms of error with respect to the use of exponential distribution, so for smaller replacement intervals s

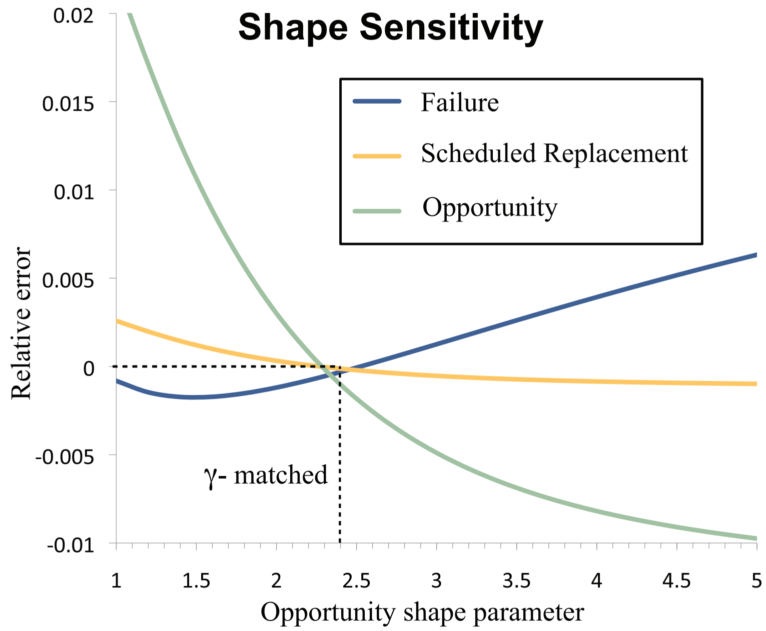

Figure 11: Sensitivity of the errors for $s = T/5 = 1$ for Case 2 as a function of shape parameter β for opportunity. Dashed lines represent the shape selected in accordance with the matched γ ratio.

exponential distribution can be sufficient as well.

Let us now compare the proposed model with more traditional fits, such as matching the first two moments of the target distribution. Left sides of Figures 13, 14, and 15 show the corresponding probability density functions, and one can observe that the proposed distribution *does not* match the target distribution over the whole range of values as well as the more traditional fits. However, the situation is reversed when we focus our attention on the portion of the distribution that is most relevant to maintenance (see the right sides of Figures 13, 14, and 15). In other words, one can conclude that the described procedure is effectively a tail-fitting one, and some potential for similarities with the second theorem in extreme value theory [34] can be further explored. The practical implications of the difference are not negligible. Indeed, for the interval $s = 1$, while using the distribution with the parameters obtained by matching the moments provides a reasonable prediction of scheduled maintenance and failures, it leads to a 4.89% underestimation of the number of opportunities for $T = 100$.

4.3. Example 4: Lognormal distributions

Finally, let us consider a situation where three components follow lognormal distribution with the following parameters: $\mu_1 = \log 2.5$, $\mu_2 = \log 4$, and $\mu_3 = \log 5$; $\sigma_1 = 0.4$, $\sigma_2 = 0.6$, and $\sigma_3 = 0.8$. As in the previous examples we considered the importance of estimating the Weibull shape, and compared the developed procedure with the exponential representation of the opportunity obtained by scale-matching (similar to the case with Weibull distribution, about an order of magnitude improvement is obtained by matching the winning ratio). The total time horizon interval $T = 4$ and use finite-difference solutions for Eqs. 4, 5. The summary of the results is shown in Figure 16, demonstrating the relative importance of the shape selection. Note that the shape was selected based on $s = 1$ which is equivalent to four replacement intervals for the lifetime T of the system, and explains why the accuracy is relatively better for that particular segment. One could

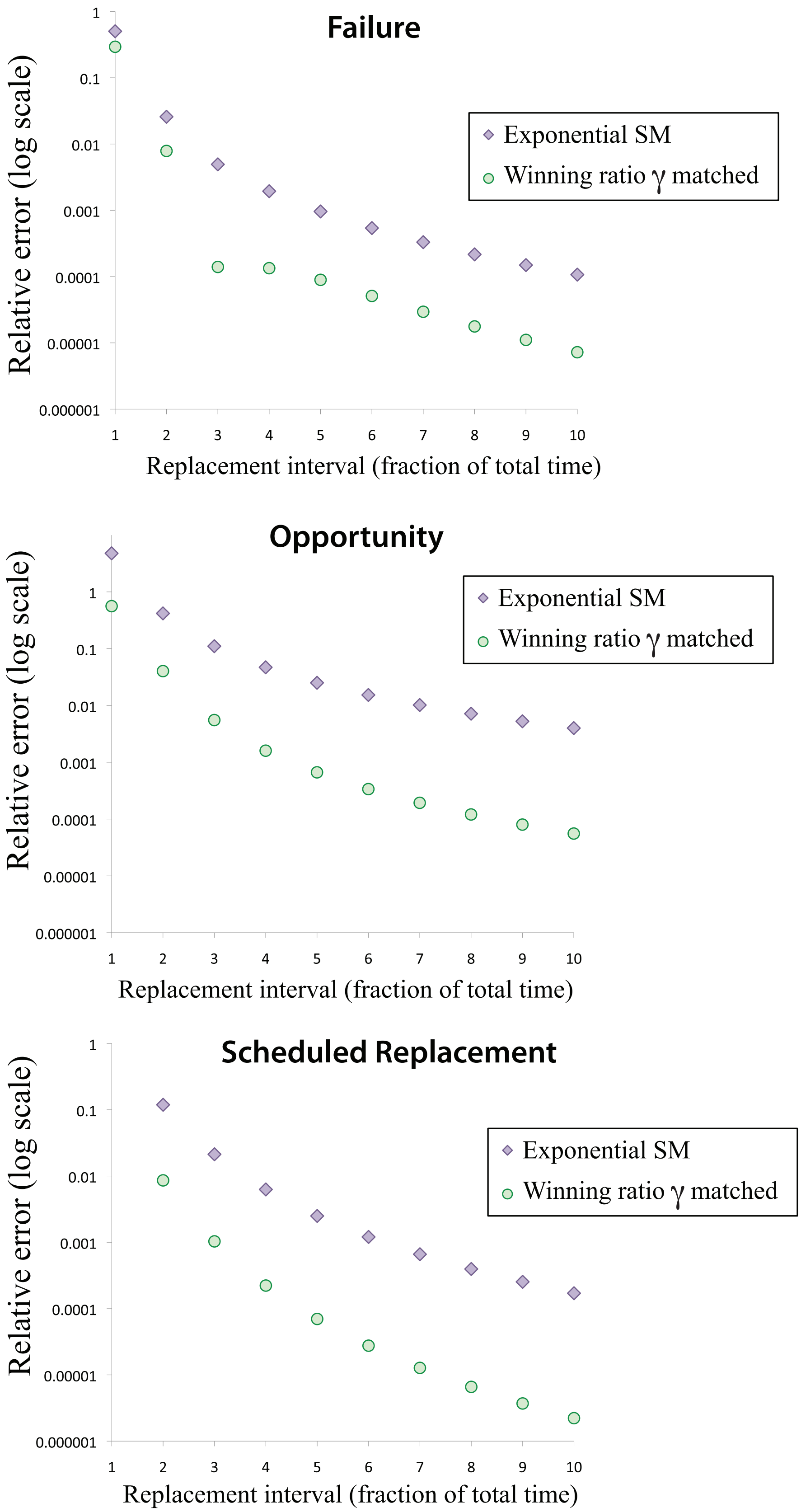

Figure 12: Example 3: failures follow Weibull distributions (all three components are different). Comparison of relative errors (in logarithmic scale) obtained using component models with opportunities represented using Weibull distribution with matched γ and exponential distribution with the matched scale. Total time $T = 5$.

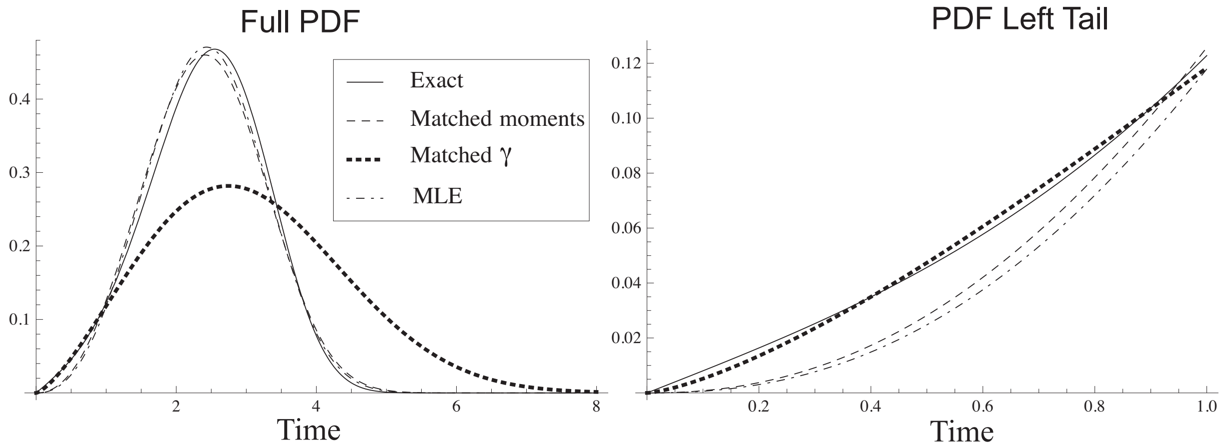

Figure 13: Probability density function (PDF) of opportunity for the first component, and its approximations using matching the first two moments, maximum likelihood estimate, and the current procedure that matches weighted γ. Full PDF (left), the left tail (right) $0 \leq t \leq 1$.

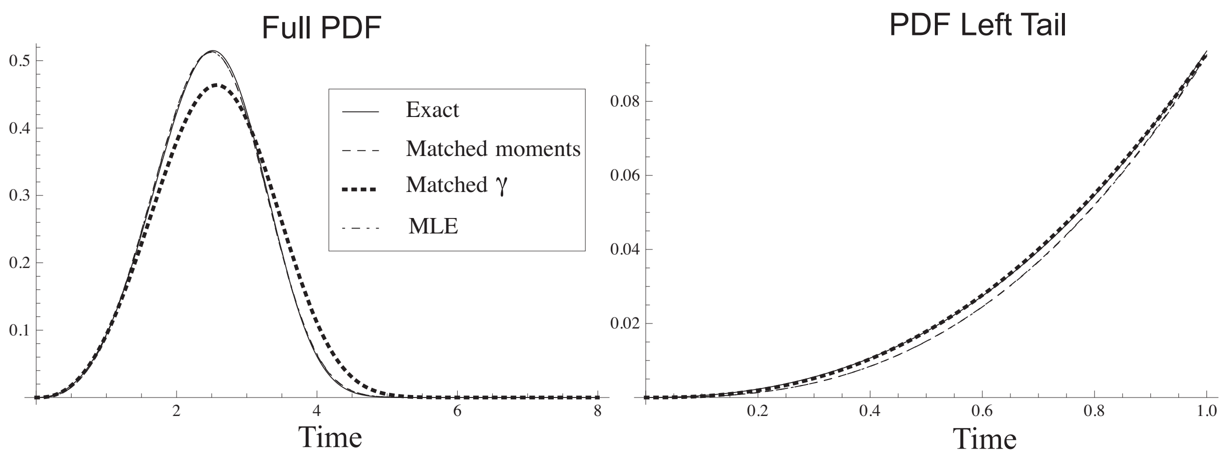

Figure 14: Probability density function (PDF) of opportunity for the second component, and its approximations using matching the first two moments, maximum likelihood estimate, and the current procedure that matches weighted γ. Full PDF (left), the left tail (right) $0 \leq t \leq 1$.

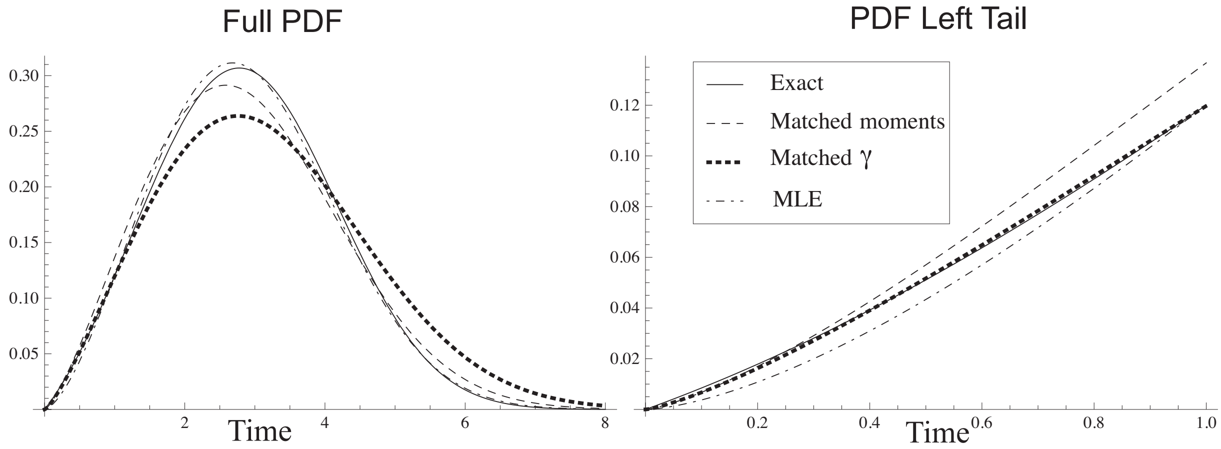

Figure 15: Probability density function (PDF) of opportunity for the third component, and its approximations using matching the first two moments, maximum likelihood estimate, and the current procedure that matches weighted γ. Full PDF (left), the left tail (right) $0 \leq t \leq 1$.

adjust the corresponding Weibull shape for each segment s, but that is clearly not necessary as the accuracy is quite good already for a wide range of replacement intervals. As expected, for small replacement intervals, exponential representation of the opportunity suffices.

5. Conclusions

An analytical method for the compact characterization of competing-risk coupling of a component with the rest of the system has been developed and tested on small scale state-space models. While the models considered in the paper were specific to modeling opportunistic maintenance combined with the age-based replacement policies, the described phenomena is pertinent to a broad range of coupling scenarios that involve several competing transitions from the same state. It has been well known previously (*e.g.*, Ref. [35]) that the use of exponential distributions in such situation can sometimes lead to significant errors (if the underlying distribution is non-exponential, including deterministic delays). However, it was not clear as to which (compact) properties of a distribution (in addition to the mean) impact the system-level results. While the standard deviation seems to be a natural candidate, the present work argues that the use of so-called "winning ratio" γ, provides superior accuracy. Specifically, it was shown that for Weibull distributions matching the winning ratio leads to more precise results than the use of a distribution that matches the two first moments of the targeted distributions or an approximation obtained using MLE. For Weibull distributions, the developed procedure is also justified on asymptotic considerations that demonstrate that, as the inspection interval gets significantly smaller than the scale of the distribution, the winning ratio tends to a simple ratio determined by the shape parameters of Weibull distributions. Instead of providing a good global match for the whole range of the distribution, the resulting approximation targets the left tail of the distribution, which is the most relevant for realistic maintenance scenarios. It was further shown that a combination of lognormal distribution can be well approximated by a Weibull distribution with matched winning

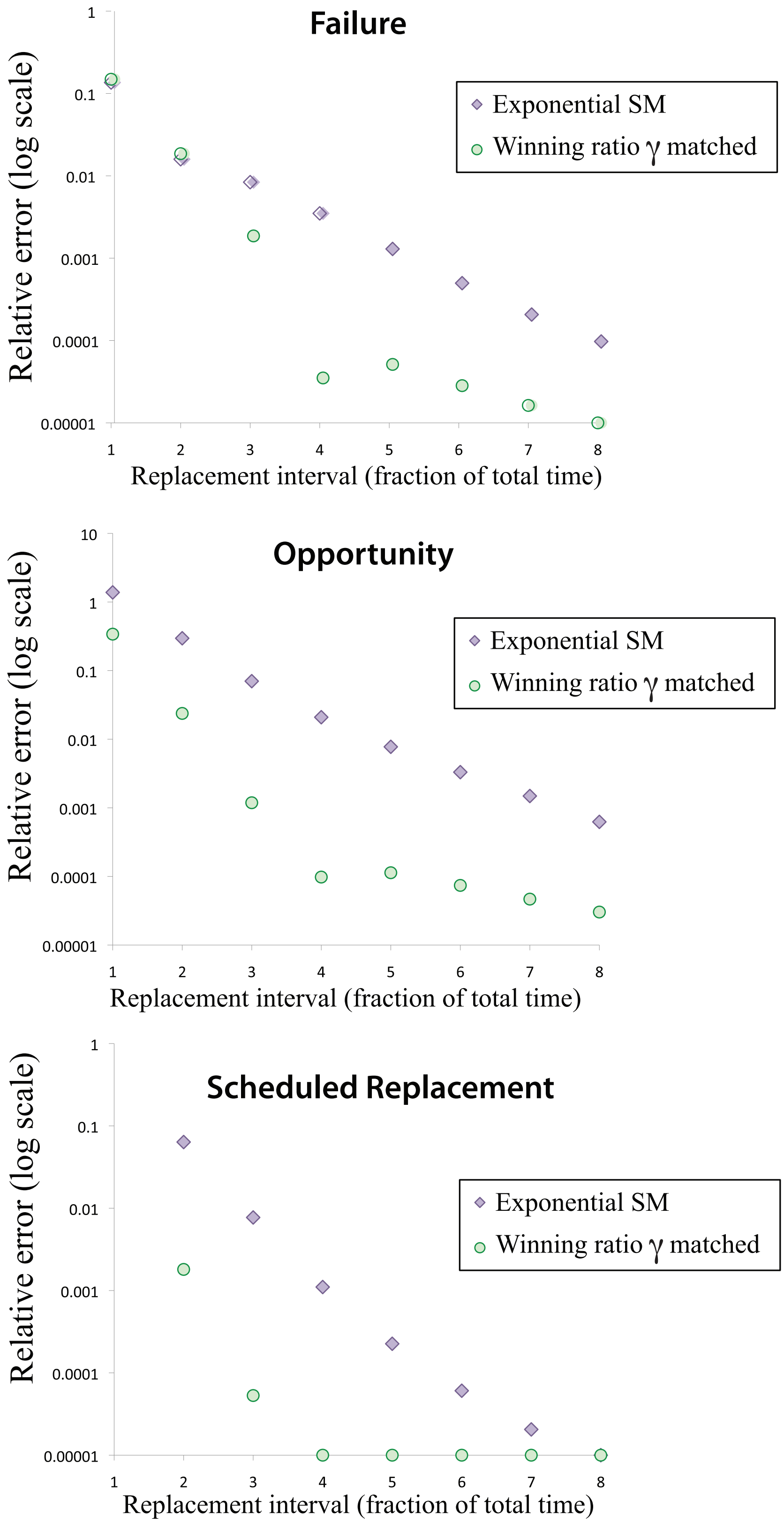

Figure 16: Example 4: failures follow log-normal distributions. Comparison of relative errors (in logarithmic scale) obtained using component models with opportunities represented using Weibull distribution with matched γ and exponential distribution with the matched scale. Total time $T = 5$.

ratio as well. It is hoped that the developed procedure will facilitates the compact representation of maintenance policies for complex systems by enabling the application of component-wise representation of maintenance processes that accurately represent inter-component couplings.

References

[1] W. R. Blischke, D. N. P. Murthy, Reliability. Modeling, Prediction, and Optimization, John Wiley and Sons, 2000.

[2] J. Thompson, W. A., On the foundations of reliability, Technometrics 23 (1) (1981) 1–13.

[3] R. Barlow, F. Proschan, Mathematical Theory of Reliability, John Wiley and Sons, New York, 1965.

[4] P. J. Haas, Stochastic Petri Nets. Modelling, Stability, Simulation, Springer, New York, 2002.

[5] Y. Dutuit, E. Châtelet, J.-P. Signoret, P. Thomas, Dependability modelling and evaluation by using stochastic Petri nets: Application to two test cases, Reliability Engineering and System Safety 55 (1997) 117–124.

[6] D. I. Cho, M. Parlar, A survey of maintenance models for multi-unit systems, European Journal of Operational Research 51 (1) (1991) 1 – 23.

[7] H. A. Simon, Near decomposability and the speed of evolution, Industrial and Corporate Change 11 (3) (2002) 587—599.

[8] K. Frenken, A. Nuvolari, The early development of the steam engine: an evolutionary interpretation using complexity theory, Industrial and Corporate Change 13 (2) (2004) 419—450.

[9] P. J. Courtois, Decomposability: queueing and computer system applications, Academic Press, New York, 1977.

[10] W. Stevens, G. Myers, L. Constantine, Structured design, IBM Systems Journal 13 (2) (1974) 115–139.

[11] P. Lollini, A. Bondavalli, A decomposition-based modeling framework for complex systems, IEEE Transactions on Reliability 58 (1).

[12] C.-D. Lai, M. Xie, Stochastic Ageing and Dependence for Reliability, Springer, New York, 2006.

[13] V. V. Volovoi, Modeling of system reliability using Petri nets with aging tokens, Reliability Engineering and System Safety 84 (2) (2004) 149–161.

[14] T. Bedford, B. M. Alkali, Competing risks and opportunistic informative maintenance, Proceedings of the Institution of Mechanical Engineers Part O-Journal of Risk and Reliability 223 (O4) (2009) 363–372.

[15] K. S. Trivedi, R. M. Geist, Decomposition in reliability analysis of fault-tolerant systems, IEEE Transactions on Reliability R-32 (5) (1983) 463–468.

[16] A. Bobbio, K. S. Trivedi, An aggregation technique for the transient analysis of stiff Markov chains, IEEE Transactions on Computers 35 (9) (1986) 803–814.

[17] M. J. Carrillo, Extensions of Palm's theorem: A review., Management Science 37 (6) (1991) 739 – 744.

[18] R. Laggoune, A. Chateauneuf, D. Aissani, Opportunistic policy for optimal preventive maintenance of a multi-component system in continuous operating units, Computers & Chemical Engineering 33 (9) (2009) 1499–1510.

[19] J. Dagpunar, Maintenance model with opportunities and interrupt replacement options, Journal of the Operational Research Society 47 (11) (1996) 1406 – 1409.

[20] N. Fard, X. Zheng, Approximate method for non-repairable systems based on opportunistic replacement policy, Reliability Engineering and System Safety 33 (2) (1991) 277 – 288.

[21] R. Dekker, M. Dijkstra, Opportunity-based age replacement: exponentially distributed times between opportunities, Naval Research Logistics 39 (2) (1992) 175 – 90.

[22] R. Dekker, E. Smeitink, Opportunity-based block replacement, European Journal of Operational Research 53 (1) (1991) 46 – 63.

[23] W. Q. Meeker, L. A. Escobar, Statistical Methods for Reliability Data, Wiley Series in Probability and Statistics, Wiley and Sons, 1998.

[24] R. Fisher, L. H. C. Tippett, Limiting forms of the frequency distribution of the largest and smallest member of a sample, Proc. Cambridge Phil. Soc. 24 (1928) 180–190.

[25] B. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire, Annals of Mathematics 44 (3) (1943) 423–453.

[26] B. Gnedenko, I. Ushakov, Probabilistic Reliability Engineering, John Wiley and Sons, New York, 1995.
[27] I. Gertsbakh, Reliability Theory with applications to Preventive Maintenance, Springer, Berlin, Germany, 2000.
[28] Y. Burtin, B. Pittel, Asymptotic estimates of the reliabilit of a complex system, Engineering Cybernetics 10 (3) (1972) 445–451.
[29] T. Nakagawa, Maintenance Theory of Reliability, Springer, London, 2005.
[30] M. Xie, On the solution of renewal-type integral equations, Communications in Statistics - Simulation and Computation 18 (1) (1989) 281–293.
[31] T. Dohi, N. Kaio, S. Osaki, Renewal processes and their computational aspects, in: S. Osaki (Ed.), Stochastic Models in Reliability and Maintenance, Springer-Verlag, Heidelberg, Germany, 2002, Ch. 1.
[32] S. Natkin, Les réseaux de petri stochastiques et leur application a l'evaluation des systemes informatiques, Ph.D. thesis, Conservatoure National des Arts et Metier, Paris, France (1980).
[33] M. K. Molloy, On the integration of delay and throughput measures in distributed processing models, Ph.D. thesis, Department of Computer Science, University of California, Los Angelos, CA, USA (1981).
[34] A. Balkema, L. de Haan, Residual life time at great age, Annals of Probability 2 (1974) 792—804.
[35] J. A. Faria, A. Matos, Manuel, An analytical methodology for the dependability evaluation of non-Markovian systems with multiple components, Reliability Engineering and System Safety 74 (2) (2001) 193–210.