Universal Failure Model for Multi-Unit Systems with Shared Functionality

Vitali Volovoi[a,*]

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

Abstract

A Universal Failure Model (UFM) is proposed for complex systems that rely on large number of entities for performing a common function. Economy of scale or other considerations may dictate the need to pool resources for the common purpose, but the resulting strong coupling precludes the grouping of those components into modules. Existing system-level failure models rely on modularity for reducing modeling complexity, so the UFM will fill an important gap in constructing efficient system-level models. Conceptually, the UFM resembles cellular automata (CA) infused with realistic failure mechanisms. Components' behavior is determined based on the balance between their strength (capacity) and their load (demand) share. If the load exceeds the components' capacity, the component fails and its load share is distributed among its neighbors (possibly with a time delay and load losses). The strength of components can degrade with time if the load exceeds an elastic threshold. The global load (demand) carried by the system can vary over time, with the peak values providing shocks to the system (*e.g.*, wind loads in civil structures, electricity demand, stressful activities to human bodies, or drought in an ecosystem). Unlike the models traditionally studied by CA, the focus of the presented model is on the system reliability, and specifically on the study of time-to-failure distributions, rather than steady-state patterns and average time-to-failure characteristics. In this context, the relationships between the types of failure distributions and the parameters of the failure model are discussed.

1. Introduction

As a part of the design process of complex systems, elaborate mapping is usually developed between tasks (functions) and the components that accomplish those tasks. A parallel process should simultaneously provide the mapping between the failures of those components to accomplish those tasks and the resulting consequences to the system. This parallel process is named differently depending on the design phase and the type of industry (functional failure mode and effect analysis, functional hazard analysis, etc). Ensuring the efficiency and continuity of this process throughout the design phases and operation provides one of the most important opportunities for improving the safety and reliability of complex systems. To this end, the use of universal modeling language (UML) [1] and, in particular, system modeling language (SysML) [2] and similar standardized tools for expressing functional interdependence provide an important means for

*Corresponding author

Email address: vitali@gatech.edu (Vitali Volovoi)

Preprint submitted to Reliability Engineering and System Safety *Accepted for Publication: 2013*

describing and tracking failure modes in a consistent fashion throughout the system's life cycle.

There is fundamental complexity issue related to the vastness of the state space that corresponds to failure scenarios of complex systems and the most efficient way of dealing with the resulting complexity is developing hierarchical models. However, the modularity that is required for those models is often broken by conflicting objectives, including the economies of scale that favor pooling resources together, and hence introduce coupling among the large number of components. The coupling is induced by shared resources and functionalities. This provides the main impetus for the development of the Universal Failure Model (UFM) that is specifically focused on situations where modularity is violated, thus providing a complementary building block for constructing comprehensive system reliability models. The rest of the paper is organized as follows: first, UFM is introduced and brief overview of relevant literature is provided, next UFM is discussed in the context of modeling failures of complex systems with the focus on the need to compress the information about individual subset of entities that comprise the system. To this end, the use of parametric distributions for time to failure is advocated, and an example of analyzing UFM from this perspective is provided. Finally, conclusions are provided and future research directions are discussed.

2. The Proposed Universal Failure Model (UFM)

The UFM is intended to provide a middle ground or an interface between compact parametric representation of failures used in system reliability and design on the one hand, and detailed domain-specific failure models on the other. In this model, a single functionality can be supported by a very large number of components, and this common purpose, as well as reliance on common resources, provides coupling mechanisms precluding the grouping of those components into modules that can be independently analyzed. The main focus of this paper is the failure dynamics of this coupled behavior. The components need not be identical, but they all are assumed to serve the same purpose. At the component level, behavior is determined based on the balance between the strength (capacity) of the component and its load (demand) share.

2.1. Background

In spirit, the proposed model is similar to the significant body of work accumulated in recent years in the area of large-scale networks [3], and specifically, their robustness to failure. The initial focus of those studies was on random failures (*e.g.*, nodal removal). More recently however, dynamic failure scenarios have attracted more attention, especially in the aftermath of the Northeast Blackout of 2003. In particular, capacity constraints and propagating failures as a result of shared load have been studied [4], as well as applications to power grids [5, 6], aviation [7], and congested networks [8]. The prevalence of certain network topologies in nature, including those that follow power law (scale-free) distribution of links, and their susceptibility to various failures in comparison to deliberately designed topologies such as "highly optimized topologies" (HOT) [9], provide important insights into the nature of failures for networked systems. Specifically, it is important to note the lack of robustness of HOT with respect to conditions that are significantly different from the ones that the networks were originally designed for.

While it is clear that network models (where nodes and links are distinguished) are relevant to the understanding of failures in complex systems, even simpler architecture

effectively consisting of nodes only (with links implied by geometric proximity) has been selected as a starting point for the current model, as it provides sufficient flexibility for exhibiting a broad range of patterns while being simple enough for exhaustive analysis of relevant statistical properties. The ultimate objective of UFM is to facilitate classification of failure dynamics, and in particular, identify distinct patterns of failure propagation as functions of the input parameters and the "tipping points", as well as the most efficient ways of delaying the occurrences of those tipping points, or preventing them all together. The resulting formalism can be classified as Cellular Automata (CA) with some similarity to sandpile models [10]. Unlike traditional CA models that assume that the behavior of individual cells is purely local, global variations of the load ("shock models" [11]) can be of significant importance and therefore provisions are made to include these global variations in the modeling. In addition, specific memory of past states (accumulated damage in a given cell, or unserviced demand) can be also introduced (this is analogous to research on the use of cell memory [12] in CA models).

In recent years there were multiple applications of CA to provide detailed domain-specific models, including the durability of concrete in aggressive environments [13], multi-pit corrosion [14], wind damage in forest planning [15], rock failures [16], and creep rupture [17]. Among the relevant general research in CA, connections to self-organized critical behavior models used to model landslides, forest fires, and earthquakes [18] must be noted, as well as models that extend the notions of damage in CA, such as the introduction of damaging agents [19]. Those and similar resources can be used to map the properties of the UFM to specific domains. To this end, relevant detailed damage propagation models (not CA-based) can be utilized as well, including the work on semicrystalline polymer fiber [20] and models of composite damage propagation [21]. In general, CA is mostly concerned with steady-state patterns, and in terms of the failure propagation, only the averaged property, *e.g.,* expected transient time, is usually assessed. In contrast, specific shapes of distributions are studied in this work, thus relating the study to existing statistical reliability models [22, 23] as well as work that related physics-based models of interactions of several specific failure modes to the time-to-failure distributions (e.g., investigation of coupling between the pitting and corrosion [24]).

2.2. UFM description

In the following description we consider a two-dimensional case, although both one- and three-dimensional cases might be of interest as well (different shapes of cells might be also explored). Let us consider n nodes, each having an initial strength $s_i(0), i = 1 \dots n$ that is independent and identically distributed (with distribution $s(x)$) and subject to the total load $L(0)$ that is assumed to be uniformly applied to each cell (initially, before the system incurred any damage), so the each cell initially carries the load $l_i(0) = L(0)/n$. The total load can vary with time in a random fashion $L(t)$. The strength of a component consists of elastic (reversible) and plastic (irreversible) phases: for the load below a certain threshold $\beta s_i(t)$, where $0 \leq \beta \leq 1$ elastic response takes place (there is no damage accumulation). For values of the load that exceed the threshold $\beta s_i(t)$ damage accumulates and the strength of the component is reduced. While different damage accumulation models can be considered [11], a simple accumulation rule is used in the following example:

$$s_i(t + \Delta t) = s_i(t) - \frac{\tau}{\Delta t} \frac{l_i(t) - \beta s_i(t)}{1 - \beta} \tag{1}$$

Here τ is the characteristic time scale of the damage accumulation. When the load exceeds the strength (demand exceeds the supply), the component fails. If failure occurs, the load is redistributed (shared) by the neighboring cells/nodes. This load redistribution has a characteristic neighborhood radius r (here defined as Chebyshev distance, that is the minimum of the distances along the coordinate axes) and time delay ϕ. The global failure of the system can be defined in different ways: either "holes" of a certain size are developed, there is a continuous "cut" of failed components that connects non-adjacent boundaries, or a certain portion $0 < \eta \leq 1$ of the population has failed. It is interesting to explore the relationships among these different global failure states, and a hypothesis can be tested that when the system reaches a critical state, it becomes unstable and several failure criteria are satisfied more or less at the same time (cf. percolation).

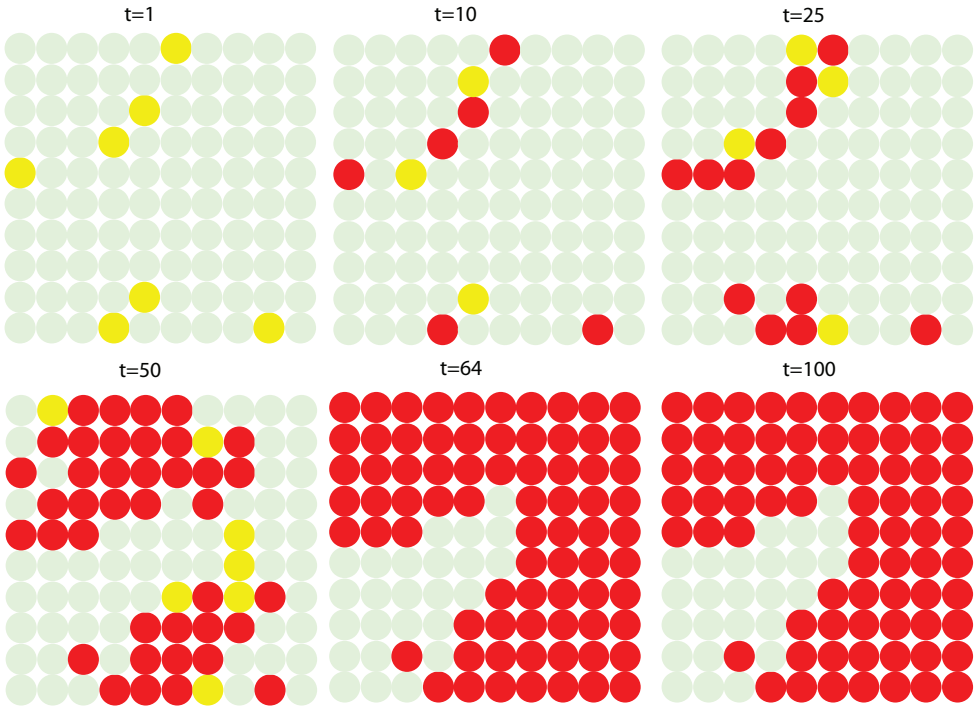

Figure 1: A snapshot of the system state. Light green cells correspond to normal loading, red cells are failed units, while yellow represents the cells that experience permanent damage

Another interesting question relates to modeling edge effects. Specifically, one can consider mirror-boundary conditions in terms of load redistribution, so that if an element at the edge fails, part of the load is being redistributed "outside" of the boundary. This option seems attractive if the considered model represents only a portion of the actual system, however, the assumption that the outside boundary elements all stay intact will obviously lead to overly benign conditions for the edge elements. A more balanced approach consists of randomly assigning failures to the "imaginary" elements at the border at the same rate as the observed elements in the system, thus mimicking load redistribution for the boundary elements of an "open" system.

An alternative approach to modeling edge effects is to model the finite size of the system while not allowing any load redistribution outside of the model. In this case, scaling issues can be investigated [25], and in general the edge elements can be subject to higher loads. The analogy with the crack growth in materials that are initiated at the

surface can be explored in this context. We will refer to the model as a "closed system" (since the load is not transferred over the boundary).

2.3. Example of the UFM implementation

Let us consider a closed 10×10 system in terms of load redistribution (as described above) with the Chebyshev radius 1 (only immediate neighbors are affected, a so-called Moore neighborhood) and a steady global load $L(t) = L(0)$. The failure propagation follows Figure 1; in the figure the following parameters are used: plastic threshold $\beta = 0.5$ and time scale $\tau = 0.3$, the initial load is $l_i(0) = 0.35$; the initial strength is distributed in accordance with the normal distribution with the mean value $\mu = 1.0$ and standard deviation $\sigma = 0.25$; the load is redistributed immediately $\phi = 0$. Time step is $\Delta t = 1$, and simulation is until $t = 100$. We note that in the beginning only a few cells are stressed; we can see how the damage gradually develops. However, in the case of this run, a stable state is reached by the time $t = 64$, and no further damage occurs.

3. Interfacing the system-level model

As described in the introduction, the goal of UFM is provide building blocks for constructing system models. In order to develop the interface of the UFM with the system-level failure analysis, let us recall the main challenges of modeling the failures of complex systems [26]:

3.1. Complex failures in complex systems

- *Large State Space*: System failures can be caused not only by the components' failure to perform their intended functions (errors of omission) but also by performing unanticipated actions (errors of commission). This greatly increases the state space needed to capture systems' behavior: instead of a binary choice for a component (it either functions or it does not), additional dimensions of the component's state must be taken into account. Furthermore, discovering these additional dimensions of the component's state is an extremely challenging process. As the design progresses, those discovered dimensions should be incorporated into the functional requirements for the component.

- *Difficulty in assessing failure mode priority*: System failures do not have to be caused by a single component failure, and instead could be the result of combination of several deviations by components from their nominal states. These deviations could stem from partial degradation of the component's performance, providing the need for further increases of the underlying dimensionality of the state space (as those degradation levels must be distinguished).

- *Disparate set of domain-specific failure mechanisms* leads to elaborate patchwork of tools and methods that are used by the experts to evaluate the reliability of components and systems. In particular, two distinct approaches to reliability prediction can be identified: on the one side of the spectrum, there is a domain-specific physics-based models of failure mechanisms for the components taken in the context of their environment (e.g., corrosion). On the other side of the spectrum, data-driven statistical analysis of field or test data leads to reliability predictions. In all practical cases, the data-driven approach includes some physics-based considerations, although they might

not be explicitly stated. In particular, the use of parametric distributions (*e.g.* Weibull or Lognormal) at some point were motivated by the physics of failure, or historically performed well, which implies that the new designed system is deemed sufficiently similar from the physics of failure perspective to justify the use the same type of distribution. As will be discussed below, the type of distribution can dramatically influence the quality of predictions. Accelerated testing and system-level analysis both combine physics-based and data-driven approaches. The former relies on statistical representation of time to failure under a controlled environment, and then rely on physics-based considerations to predict the timing of failures in the field. The latter relies on statistical information about component failures (that can be obtained either from past experience or based on physics-of-failure modes) and the information about the interrelationships among the components to infer the relevant reliability characteristics of the system. In the case of repairable systems, those characteristics are not limited to simple time to failure, but might also include availability, expected number of failures, etc. The fidelity of modeling interrelationships among components can very from simple logical "and" and "or" (*e.g.*, fault trees) to discrete event-representation where the timing and order of events is taken into account, but the state space is discrete (e.g., Markov chains and stochastic Petri nets (SPN) [27]), and to models where both time and spatial description is continuous (e.g., agent-based simulation)[28].

Given these challenges, the means to improve system safety and reliability and potential place of UFM are discussed next.

3.2. Coping with the complexity of failures

- First, if one accepts the notion that complexity can lead to system failures, it is important to measure the complexity. The complexity of the system is related to the amount of information needed to describe the system (following a general definition of Kolmogorov complexity expressed as entropy [29]), and, specifically, to the size of the state space representing distinct states of the system. The simplest proxy for this parameter is the number of components that comprise the system, and this measure of system reliability implies that the system is designed so that the failure of any of its components results in the failure of the whole system (*i.e.,* the system does not have any redundancy). Under these (not very practical) assumptions, the reliability of the system is simply the product of the reliability of the individual components [30].

- Modularity: Alternative measures of complexity take into account not only the number of entities and the size of state space describing those entities, but also some measures of the amount of interrelationships among those entities (couplings). In particular, using graph-theoretical setting allows to represent individual components as nodes and use the links between pairs of nodes to represent dependence between the corresponding components' states (*i.e.,* the couplings. Several measures of complexity rely on these concepts, including a measure based on branching diversity for graphs that can be represented as trees (no cycles) of collection of trees (forests) [31] and cyclomatic complexity (related to the number of linearly independent loops) [32]. Modularity (*cf.* the related principles of encapsulation and information hiding that are formalized in object-oriented programming) provides a well-accepted means of reducing complexity by deliberately designing the system's architecture in a hierarchical fashion and minimizing inter-modular coupling. However, efficiency (due to economies of scale) and

other practical considerations often lead to a situation where a single functionality is supported by a very large number of components, and this common purpose, as well as reliance on common resources, provides inevitable tight coupling mechanisms. As a result, for such tightly-coupled systems it is impossible to group the components into modules that can be independently analyzed.

At the same time, there is 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 [33], while similar processes were identified in the history of steam engine development [34]. These concepts are also explicitly employed in the design principles of computer systems [35] (including structured design [36]). In this context the importance of so-called "weak links" has been explored in various domains, including biological systems [37].

Deviation from modular design can lead to serious consequences, as argued by Perrow [38] who postulated that a high degree of complexity and level of coupling inevitably lead to accidents in complex systems by inducing a conflict between centralized and decentralized modes of control. In this setting, coupling, rather than representing a single aspect of complexity has a distinct meaning and is measured in terms of the time required for the disturbances to propagate among entities. This is an important parameter since the propagation time is related to the time available for reaction to the disturbances in the system. In some (but not all) circumstances it is reasonable to assume that those two measures of coupling are related, as a large number of links are expected to facilitate faster disturbance propagation.

The relative merits of two opposing trends (economies of scale vs. tight coupling) could be fundamentally hard to evaluate due to the inherent differences in the frequencies of the associated feedback (see the discussion on the importance of the feedback in avoiding failures [26, 39]). Indeed, while the benefits of economies of scale are immediately obvious, the increased vulnerability to catastrophic failures due to tight coupling could take years to exhibit itself (when the decision-makers responsible for the selection of the system's architecture are long gone). This shortsightedness effect is well known in public policy [40], and providing at least partial remedy for this myopia by developing credible models that explore future scenarios is a very attractive goal. It is hoped that the current work will lead to some small steps toward that goal. From more technical perspective, it is also interesting to note the parallels with the recent popularity of copulas in modeling dependencies in the context of financial risks [41]. Indeed, unlike standard correlations, copulas provide the means for modeling scenarios where under normal circumstances components of the system appear to be independent, while in stressed conditions those components exhibit highly coupled dynamics. In this context it is interesting to see whether UFM eventually allows to demonstrate similar phenomenon endogenously. Finally, it is interesting to explore how the presence of two opposing forces can lead to a self-organized configuration in terms of the amount of coupling, *cf.* [42].

- Redundancy is a fundamental principle of design for reliability [43] recognized since the time of Von Braun and implemented under various names: "no single-point failure" (in aerospace), "damage tolerance" (in structures), and "defense-in-depth"(nuclear

plants) [44]. Redundancy increases the system complexity, but also alters the relationship between reliability and operational costs. If the reliability of the system is driven by component reliability, then lower reliability implies more frequent maintenance (and increased demand for spare parts), leading to increased operational cost. However, this relationship between reliability and maintenance costs can be reversed if redundancy is used to improve system reliability, as component failures do not result in system failures, yet require maintenance actions [26]. One of the interesting potentials of UFM is facilitating a quantitative measure of effective redundancy (*e.g.,* developing a coherent metric expressed as a scalar that is not necessarily an integer).

- Component Reliability: One of the simplest methods of increasing component reliability is derating [43], where the components are rated for a higher stress environment than they experience in operation. Plotting the probability distribution of both the load, $l(x)$, and the strength, $s(x)$ (also known in other applications as demand and available resources or supply, respectively) provides visual representation of the chances of failure (see Figure 2). The intersection of the two probability density functions (shown in red) is indicative of the possibility of failure, although, contrary to common belief, this area does not numerically represent the probability of failure. Effectively, derating implies increased safety margins when the mean corresponding strength (capacity) is further separated from the load (demand). One can also interpret derating as internal redundancy (as the component has a spare capacity that provides protection against variability, see also the discussion on the effective redundancy above).

  Probability of failure can also be reduced by decreasing the variance of each of the distributions, in particular that of the strength, which provides one of the motivations for statistical quality control [45] and similar concepts. Indeed, improving manufacturing tolerances and reducing other variabilities during the manufacturing process leads directly to the reduction of the variability of the strength distribution, and indirectly (e.g., via the influence on the surrounding components) to similar effect for the load variability. In structural applications, the distance between the nominal values of strength and load can be associated with safety factors [46]. Qualitatively, it is clear that for a structural component that is subject to degradation, both distributions will move toward each other as time progresses, leading to increased probability of failure. While the leftward movement of the strength curve is attributed to the degradation of the component itself (at the local level, some portion of the component), the rightward movement of the load curve is caused by the load redistribution due to the degradation of the component environment (or at the local level, the degradation of the adjacent portions of the same component). Quantification of the dynamic relationship between the load and the strength is significantly more challenging and constitutes another important goal of UFM.

3.3. Motivational example

Let us consider stress-rupture failure mode for composite overwrapped pressure vessels (COPV) that are used to store fuel in space vehicles and gas in other applications. COPVs are designed to store gas under the pressure p_0,which is a fraction of the ultimate pressure that would blow up COPVs. Manufacturing processes for COPVs have significant variability (in particular, due to the strength variability of the fibers used in the composite). A proof test is used to screen out weak vessels: for a short period of

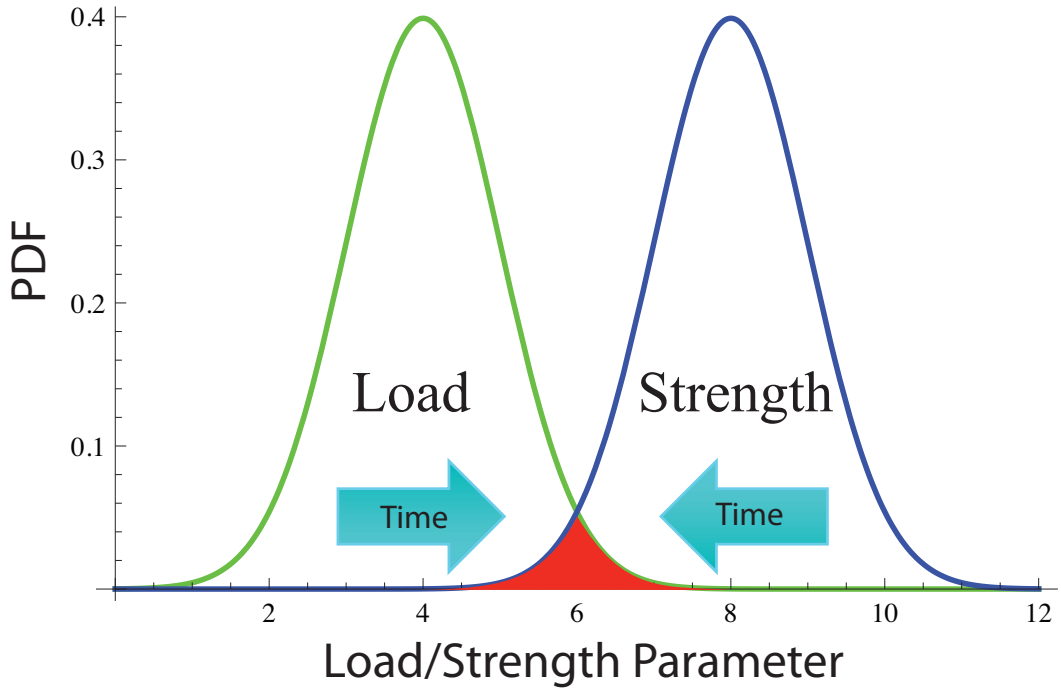

Figure 2: Dynamic Relation between the strength and the load.

time (usually measured in minutes), COPVs are tested under pressure that is significantly higher than p_0. This leads to the "to proof or not to proof" question: on the one hand, the damaged vessels are discovered; on the other hand, the stronger ones that survive the test might have been weakened. The underlying phenomenon is fully analogous to observer effect in physics and the original formulation of the Heisenberg uncertainty principle (but distinct from the modern interpretation of the latter that involves inherent fluctuations of the quantities of interest regardless of the observation [47]). Is the resulting population actually better (*i.e.,* has higher reliability) than the original population? The question can be related to the shape of the time-to-failure distribution: if the population failure rate decreases with time, then proofing makes sense (as the effective age of the system is increased by proofing). It can be observed that the heterogeneity of the the population leads to a decrease in failure rate with time (as only the stronger members of the population survive). At the same time, effective redundancy acts in the opposite direction (as redundancy degrades due to random failures at the component level that do not cause system failure). Similarly, degradation mechanisms at the component level also lead to the failure rate increasing with time. Therefore one needs to understand which of the opposing trends dominates, and the UFM aims at addressing such trade-offs. To this end, general mapping between input parameters of UFM and the types of parametric distributions for time to failure can provide an effective interface with system-level models.

4. Parametrization of time to failure

State-space based models for evaluating system safety and reliability rely on compact representation of state transitions (e.g., failures and repairs, or recoveries from intermittent faults). Parametric distributions are preferred from the compactness perspective,

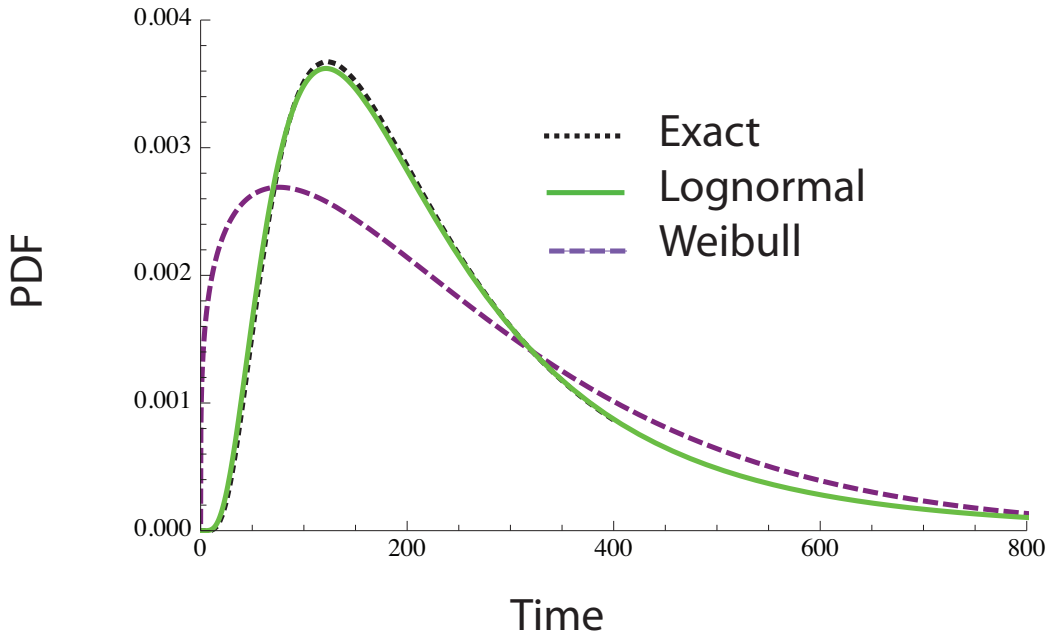

Figure 3: Probability density function (PDF) for failure of a component with variability driven by the operational conditions (temperature following a Gaussian distribution)

assuming that their accuracy is assured. The question of selecting appropriate distribution might seem obsolete in the modern world where non-parametric representation can be easily stored on a computer; however, selection of the distribution effectively implies infusing the statistical process with physics-based knowledge, and significantly reduces the need for experimental data about the system. Successful application of parametric distributions is closely related to taking advantage of the underlying general physical processes, just like the blind use of parametric distributions can lead to serious modeling flaws. The central limit theorem assertion that the sum of large number of independent random variables follows normal distribution is the best-known case, but in the context of reliability, the importance of several types of distributions is similarly clear, as briefly discussed next.

4.1. Commonly used parametric distributions

Exponential Distribution A failure transition with the constant rate λ follows an exponential distribution, whose cumulative form is given by $F_e(t) = 1 - e^{-\lambda t}$. λ is the inverse of the mean time to failure. State-space models with constant transition rates are particularly convenient: first, each transition is fully characterized by a single parameter, λ; and second, the resulting process is Markov (i.e., the chances of transitioning to a new state are fully determined by the current state), which drastically simplifies the analysis. In the context of repairable systems, steady-state results often depend only on the mean parameters of the distribution, justifying the use of exponential distribution even if the underlying distributions are different (see, for example, the extensions of the Palm-Khinchin theorem for queues in logistics applications [48]). "The central limit theorem for repairable systems" provides another argument for resorting to exponential distributions: in a complex repairable system with multiple components, failures form a homogeneous Poisson process [49]. This theorem is only valid, however, if there is no

coordination among component failures. In practice, for many systems with clear aging or degradation patterns, major inspections and overhauls impose an overall structure, and within each maintenance cycle the failure rate can vary significantly.

Weibull Distribution $F_w(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^{\beta}}$ 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 $\beta = 1$, Weibull distribution becomes exponential, with the scale parameter θ representing the inverse of the transition rate. An additional reason for using Weibull distribution in system reliability is its relationship to the "weakest link" mode of failure. The Fisher-Tippett-Gnedenko theorem [50, 51] 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).

Lognormal Distribution The lognormal model of time to failure is justified when a process moves towards failure based on the cumulative effect of many small "multiplicative" shocks. Specifically, if at any instant in time a degradation process undergoes a small increase in the total amount of degradation that is proportional to the current total amount of degradation, then the time to failure (i.e., reaching a critical amount of degradation) is expected to follow a lognormal distribution [52].

There are other situations where the use of lognormal distributions can be justified as well. For example, let us consider the scenario where the sole source of uncertainty is the operating temperature (while uncertain, it stays constant over time): for a fixed operating absolute temperature T the failure time t is uniquely determined. The Arrhenius equation is used for the rate of chemical reactions, $k(T) = a \exp\left(-\frac{E_a}{RT}\right)$, where a is a pre-exponential factor, E_a activation energy, and R gas constant. If for an operation with (constant) temperature T_0 the time of failure is t_0, then the time of failure for other temperatures can be uniquely determined as $t = g(T) = \alpha \exp\left(\frac{\beta}{T}\right)$, where the following notations are used: $\alpha = t_0 \exp\left(-\frac{E_a}{RT_0}\right)$ and $\beta = \frac{E_a}{R}$. Next, if the operation temperature is normally distributed with the mean value T_0 and the standard deviation σ, one can obtain the corresponding failure distribution in analytical form using the fact that there is a one-to-one relationship between the operation temperature and failure time. Indeed, the application of the formulae for the function of a random variable to calculate the probability density function for the time failure yields

$$f_t(t) = f_T(g^{-1}(t)) \left| \frac{dg^{-1}(t)}{dt} \right| = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{\left(\frac{\beta}{\ln\left(\frac{t}{\alpha}\right)} - T_0\right)^2}{2\sigma^2} \right] \frac{\beta}{t\left[\ln\left(\frac{t}{\alpha}\right)\right]^2} \tag{2}$$

As can be observed in Fig. 3, this function (black dotted curve) can be approximated quite well using log-normal distribution (green solid curve), while Weibull distribution is poorly suited in this case (dashed purple curve).

Other distributions Other distributions can also be applicable, including Gamma distribution [53] and Birnbaum Saunders distribution, which is often used to model fatigue life [54]. In addition, as discussed later, time shifts can be introduced to a distribution to capture dormant phase of failure development. One of the intriguing possibilities is to investigate applicability of a wider range of parametric distributions that found ap-

plications in finance in describing values of relevant parameters (rather than time to failure), especially in relation to financial crashes, including stretched-exponential distributions [41] and log-Lévy distributions [55]. A systematic justification of the use of a given distribution can be provided by relating the most appropriate distribution to fit the universal failure model for a given set of parameters (providing a mapping between the UFM parameters and appropriate distributions); at the same time a complementary mapping between existing domain-specific models and the appropriate range of parameters of the universal failure model can be developed as well.

4.2. Selecting failure distributions using the UFM

Next, let us evaluate the statistics of time to failure for a slightly different configuration: plastic threshold $\beta = 0.55$ and time scale $\tau = 0.2$. The initial load is $l_i(0) = 0.45$, and the load is redistributed immediately $\phi = 0$. The initial strength is distributed in accordance with the normal distribution, with the mean value $\mu = 1.0$ and standard deviation $\sigma = 0.2$. In addition, let us consider Weibull distribution individual cell strength, instead of the normal distribution, since Weibull often better matches the strength distribution (for example of fibers in composites). To facilitate the comparison, we match the first two moments of the normal distribution by appropriately selecting Weibull shape parameter $\beta = 5.7974$ and the scale $\theta = 1.08$ 100 steps are conducted and 100, 000 Monte Carlo runs are used. The failure criteria for the system is based on on the fraction of failed states $\eta = 0.9$.

There is about a 33% chance that the system will not fail at all by the end of simulation (this chance is the same for both distributions used for the initial strength distribution). This is a somewhat unusual situation from the classical reliability perspective, where the time to failure is usually assumed to be given by a continuous distribution. In general, as can be observed in Figure 4 that shows both histograms of the time to failure, the difference between the use of two distributions to model the strength in this case is quite minimal (one can observe a very minor effect of a slightly heavier left tail for the Weibull distribution).

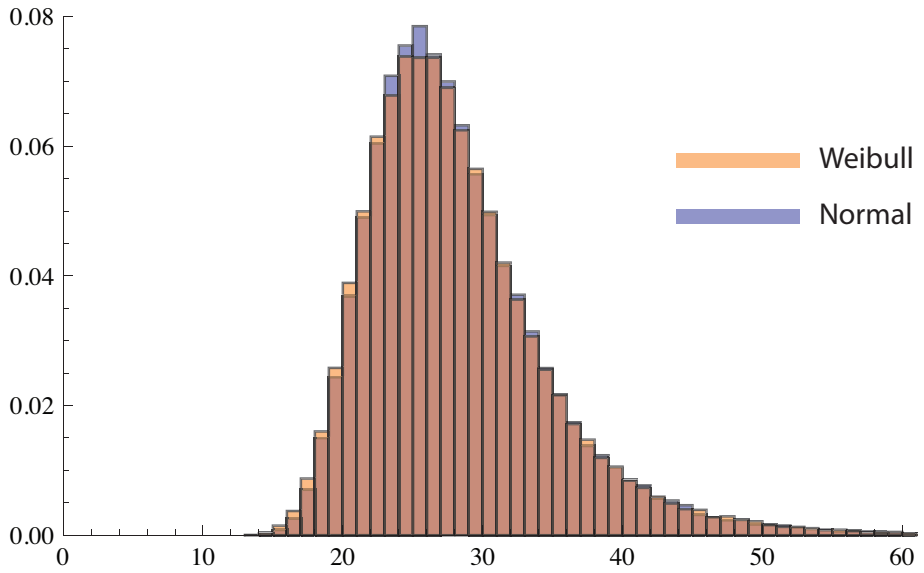

Figure 4: Distribution of failure time for the initial strength following normal and Weibull distributions

Next, we will further analyze the data obtained for the case where the strength was normally distributed at $t = 0$. Let us try to fit several types of distributions that are commonly used to describe time to failure. Here it suffices to say that none of the four distributions fits the data too well (Figures 5, 6 demonstrate the fit based on matching mean and standard deviation, and the fit using the Maximum Likelihood Estimate (MLE), respectively. We can note that the data is clearly skewed (so that the "fatter" or "heavier" is to the right); this corresponds to the positive value of skewness. The value for the considered data is $\gamma_d = 1.485$. For the considered range of the shape parameters of Weibull distribution has negative skewness, and one can observe that the match to the data is poor indeed, especially if MLE is used. The results can be overly conservative: out of $100,000$, the first failure took place at time $t_1 = 13$ (out of $100,000$, there are only two such cases), while Weibull distribution (based on MLE) would predict 3505 failures by the time $t = 13$. The best fit is provided by lognormal and Birnbaum-Saunders distributions: predictions for the number of failures is 43 and 36 out of $100,000$, respectively.

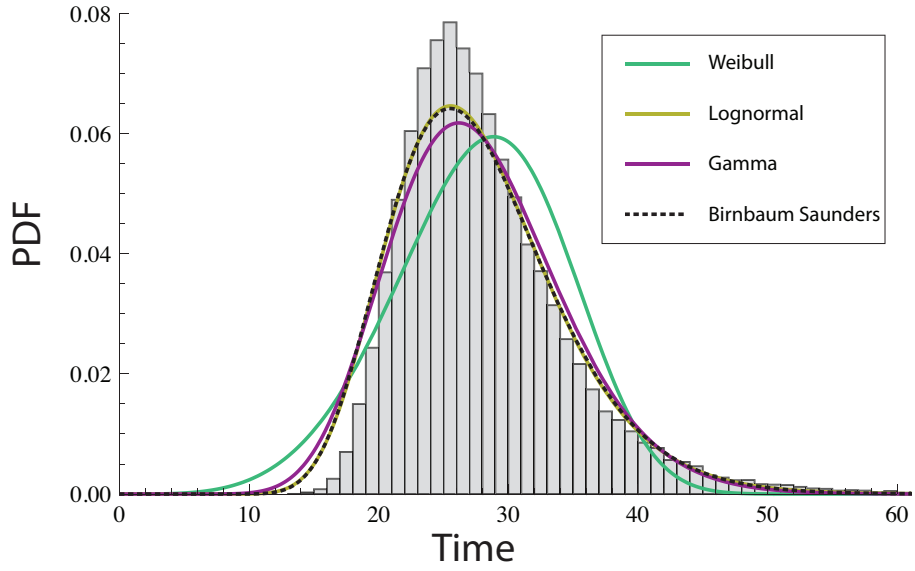

Figure 5: Matching failure time distribution using common parametric distribution: first two moments are matched with the data;

However, we note that even those two distributions significantly over-predict the relative weight of the tails. In fact, the kurtosis (peakedness) of the data $\kappa_d = 7.39$ is almost twice the value of both parametric approximations ($\kappa_l = 3.83$ and $\kappa_b = 3.73$ for MLE approximations of lognormal and Birnbaum-Saunders, respectively). Similarly, the skewness is off as well ($\sigma_d = 1.485$ vs. $\sigma_l = 0.68$ and $\sigma_b = 0.66$). There are two possibilities here: the underlying distribution is indeed different (and highly short-tailed), or there is a time shift that exists in the system. Checking the second hypothesis, and matching the skewness of the data, we can determine effective time shifts $t_{0l} = 13.44$, $t_{0b} = 15.06$. The resulting distributions are shown in Figure 7, recalling that there were no observed failures for $t < 13$, and that shifted kurtosis estimates are much closer to the data $\tilde{\kappa}_l = 7.16$ and $\tilde{\kappa}_b = 6.58$; we can conclude the there there is a possibility of effective time delay before the failures start to occur. The possibility of latent time in the systems is of great potential importance, and certainly merits further investigation.

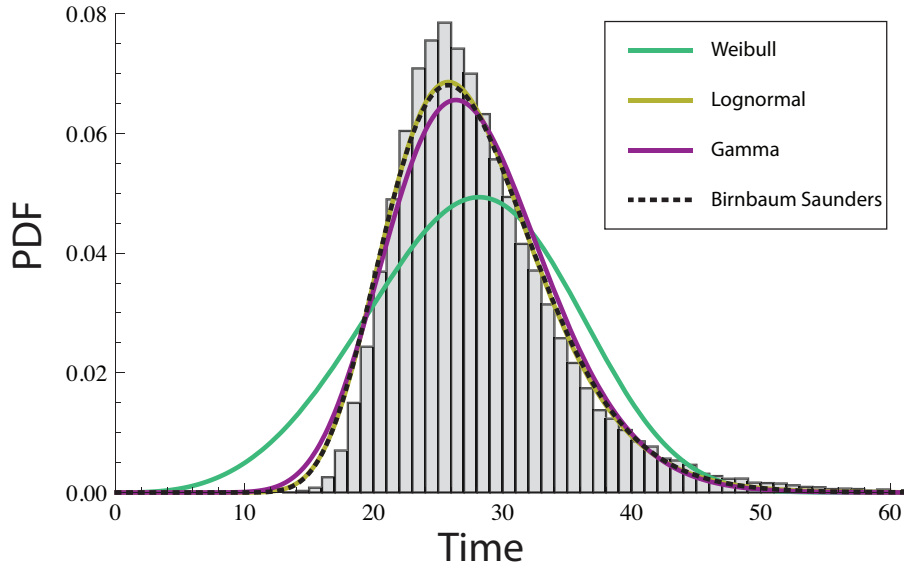

Figure 6: Matching failure time distribution using common parametric distribution using Maximum Likelihood Estimate (MLE)

5. Conclusions and Future Work

A Universal Failure Model (UFM) is introduced for complex systems that rely on large number of entities to accomplish a common function. It provides fundamental building blocks for the modeling of failures of complex systems by capturing the failure dynamics of a very large number of coupled entities (components) supporting a single functionality. The resulting strong coupling precludes the grouping of those components into modules as required for hierarchical model construction. Existing system-level failure models rely heavily on modularity for reducing modeling complexity, so the UFM can fill an important gap in constructing efficient system-level models. Such models can be useful in addressing the challenges of modeling interdependent infrastructures [56]. Conceptually, the UFM resembles cellular automata (CA) supplemented with realistic failure mechanisms. Components' behavior is determined based on the balance between the strength (capacity) of the component and their load (demand) share. If the load exceeds the components' capacity, the component fails and its load share is distributed among its neighbors (possibly with a time delay and load losses). The size of the neighborhood that assumes the load of the failed component determines how local the load redistribution is. The strength of components can degrade with time if the load exceeds a certain elastic threshold. While individual features of the UFM appear in various contexts (shock models in reliability, balance of supply and demand in economics, balance of strength and load in structures, the visual nature of CA), they have not been previously combined into a single model. The interplay of those features provides a "sandbox" where the dynamics of complex systems can be systematically explored. As a result, system design trade-offs among the effective redundancy, strength variability (as related to manufacturing tolerances and therefore costs), and healing capabilities can be made. Important distinctions of the UFM as compared to existing CA models include external global shocks (so that behavior is not purely local), assigning memory to the states to account for accumulated damage, and explicit emphasis on the interface with the system-level reliability mod-

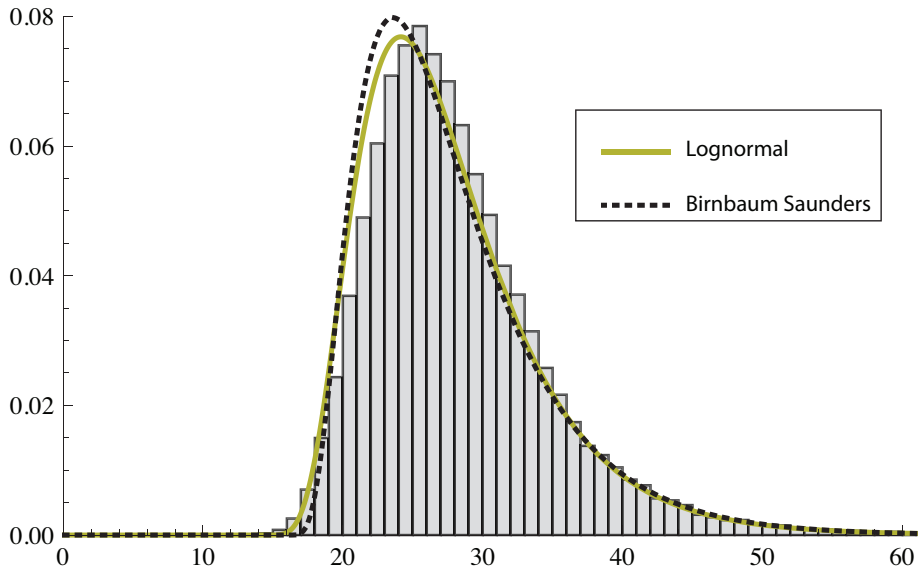

Figure 7: Introducing the time shift for lognormal distribution (shift $t_{0l} = 13.44$) and Birnbaum Saunders (shift $t_{0b} = 15.06$)

els (in particular, stochastic Petri nets). The latter require a focus on specific patterns of time-to-failure distributions, rather than steady-state patterns and average time-to-failure characteristics that are traditionally studied in CA. The missing details specific to particular systems (i.e., anisotropy of load redistribution) can be captured indirectly by adjusting the parameters of the model based on more detailed domain-specific damage models, providing predictive capabilities that are superior to purely data-driven models. The simplicity and visual nature of the proposed models can facilitate a broad understanding of failure mechanisms in complex systems not only by the experts in failure analysis but by a broader audience, including the designers of those systems.

Future work will be focused on systematic investigation of mapping between the different input parameters and extensions of UFM on the one hand, and the observed patterns of time-to-failure and other relevant measures of failure processes (including identifying system-failure precursors) on the other. In particular, various patterns of global load (demand) variations over time will be investigated, with the peak values providing shocks to the system (*e.g.*, wind loads in civil structures, electricity demand, stressful activities to human bodies, drought in an ecosystem, or panic in a stock market). Introduction of alternative to lattice arrangements of the neighbors (thus providing connection to coarse-scale network models) is another natural direction. Finally, investigating optimal strategies for restoring/healing failed components to prevent system failures is also of great interest.

References

[1] S. ISO/IEC, Information technology – open distributed processing – unified modeling language (UML) version 1.4.2, Tech. rep., International Organization for Standardization (2005).

[2] S. Friedenthal, A. Moore, R. Steiner, A Practical Guide to SysML: The Systems Modeling Language, Morgan Kaufmann and OMG, Amsterdam, Netherlands, 2009.

[3] S. Boccaletti, V. Latora, Y. Morenod, M. Chavezf, D.-U. Hwang, Complex networks: Structure and dynamics, Physics Reports 424 (2006) 175–308.

[4] A. E. Motter, Cascade control and defense in complex networks, Physical Review Letters 93 (9) (2004) 098701 – 1–098701–4.
[5] R. Kinney, P. Crucitti, R. Albert, V. Latora, Modeling cascading failures in the North American power grid, European Physical Journal B 46 (1) (2005) 101–107.
[6] I. Dobson, B. A. Carreras, D. E. Newman, A loading-dependent model of probabilistic cascading failure, Probability in the Engineering and Informational Sciences 19 (2005) 15–32.
[7] L. Lacasa, M. Cea, M. Zanin, Jamming transition in air transportation networks, Physica A-Statistical Mechanics and its Applications 388 (18) (2009) 3948–3954.
[8] D. De Martino, L. Dall'Asta, G. Bianconi, M. Marsili, Congestion phenomena on complex networks, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 79 (1).
[9] J. Carlson, J. Doyle, Highly optimized tolerance: Robustness and design in complex systems, Physical Review Letters 84 (11) (2000) 2529 – 2532.
[10] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality, Phys. Rev. A (1988) 364–374.
[11] T. Nakagawa, Shock and Damage Models in Reliability Theory, Springer, London, 2007.
[12] R. Alonso-Sanz, J. P. Cárdenas, Effect of memory on boolean networks with disordered dynamics, International Journal of Modern Physics C: Computational Physics & Physical Computation 18 (8) (2007) 1313–1327.
[13] F. Biondini, P. G. Malerba, F. Bontempi, D. M. Frangopol, Cellular automata approach to durability analysis of concrete structures in aggressive environments, Journal of Structural Engineering 130 (11) (2004) 1724–1737.
[14] R. Pidaparti, L. Fang, M. Palakal, Computational simulation of multi-pit corrosion process in materials, Computational Materials Science 41 (3) (2008) 255–265.
[15] H. Zeng, T. Pukkala, H. Peltola, S. Kellomaki, Optimization of irregular-grid cellular automata and application in risk management of wind damage in forest planning, Canadian Journal of Forest Research 40 (6).
[16] D. V. Alekseev, G. A. Kazunina, Simulation of damage accumulation kinetics with a probabilistic cellular automaton, Physics of the Solid State 48 (2) (2006) 272–278.
[17] M. Chrzanowski, K. Nowak, Cellural automata in damage mechanics: Creep rupture case, Archive Mechanics 59 (4-5) (2007) 329–339.
[18] D. L. Turcotte, B. D. Malamud, Landslides, forest fires, and earthquakes: examples of self-organized critical behavior, Physica A 340 (2004) 580–589.
[19] C. Ferraz, H. Herrmann, Strange man in random networks of automata, Physica A-statistical Mechanics and Its Applications 387 (23) (2008) 5689—5695.
[20] H. M. Taylor, A model for the failure process of semicrystalline polymer materials under static fatigue, Probability in the Engineering and Informational Sciences 1 (02) (1987) 133–162.
[21] S. Mahesh, S. Phoenix, Lifetime distributions for unidirectional fibrous composites under creep-rupture loading, International Journal of Fracture 127 (4) (2004) 303–360.
[22] Z. W. Birnbaum, S. C. Saunders, A probabilistic interpretation of miner's rule, SIAM Journal on Applied Mathematics 16 (3) (1968) 637–652.
[23] A. J. Lemoine, M. L. Wencour, On failure modeling, Naval Research Logistic Quarterly 22 (1985) 497–508.
[24] M. Chookah, M. Nuhi, M. Modarres, A probabilistic physics-of-failure model for prognostic health management of structures subject to pitting and corrosion-fatigue, Reliability Engineering and System Safety 84 (2) (2004) 149–161.
[25] Z. P. Bažant, Scaling theory for quasibrittle structural failure, Proceedings of the National Academy of Sciences of the United States of America 101 (37) (2004) 13400–13407.
[26] V. Volovoi, System reliability at the crossroads, ISRN Applied Mathematics 2012 (2012) Article ID 850686.
[27] V. V. Volovoi, Modeling of system reliability using Petri nets with aging tokens, Reliability Engineering and System Safety 84 (2) (2004) 149–161.
[28] G. Calanni, V. Volovoi, A. Colon, M. Blake, Novel air traffic procedures: Investigation of off-nominal scenarios and potential hazards, AIAA Journal of Aircraft 48 (1) (2011) 127–140.
[29] D. Braha, A.A., Minai, Y. Bar-Yam (Eds.), Complex engineered systems: science meets technology, Springer, Berlin, 2006.
[30] M. M. Rausand, Høyland, System Reliability Theory. Models, Statistical Methods, and Applications, 2nd Edition, John Wiley and Sons, New York, 2004.
[31] B. Huberman, T. Hogg, Complexity and adaptation, Physica D 22 (1-3) (1986) 376–384.

[32] T. J. McCabe, A complexity measure, IEEE Transactions on Software Engineering SE-2 (4) (1976) 308–320.
[33] H. A. Simon, Near decomposability and the speed of evolution, Industrial and Corporate Change 11 (3) (2002) 587—599.
[34] 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.
[35] P. J. Courtois, Decomposability: queueing and computer system applications, Academic Press, New York, 1977.
[36] W. Stevens, G. Myers, L. Constantine, Structured design, IBM Systems Journal 13 (2) (1974) 115–139.
[37] P. Csermely, Weak Links: The Universal Key to the Stability of Networks and Complex Systems, The Frontiers Collection, Springer, 2006.
[38] C. Perrow, Normal accidents: living with high-risk technologies, Princeton University Press, Princeton, N.J., 1999.
[39] T. Harford, Adapt: Why Success Always Starts with Failure, Farrar, Straus and Giroux, New York, 2011.
[40] A. Healy, N. Malhotra, Myopic voters and natural disaster policy, American Political Science Review 103 (2009) 387–406.
[41] Y. Malevergne, D. Sornette, Extreme Financial Risks (From dependence to risk management), Springer-Verlag, Berlin Heidelberg, 2006.
[42] P. Krugman, The Self Organizing Economy, Wiley-Blackwell, 1996.
[43] D. Crowe, A. Feinberg (Eds.), Design for reliability, CRC Press, Boca Raton, Fla., 2001.
[44] J. Reason, Managing the risks of organizational accidents, Ashgate, Brookfield, Vt., 1997.
[45] W. A. Shewhart, Economic control of quality of manufactured product, D. Van Nostrand Company, 1931.
[46] H. O. Madsen, S. Krenk, N. C. Lind, Methods of Structural Safety, Dover Publications, 2006.
[47] A. Furuta, One thing is certain: Heisenberg's uncertainty principle is not dead, Scientific American.
[48] M. J. Carrillo, Extensions of Palm's theorem: a review, Management Science 37 (6) (1991) 739 – 744.
[49] R. Barlow, F. Proschan, Mathematical Theory of Reliability, John Wiley and Sons, New York, 1965.
[50] 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.
[51] B. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire, Annals of Mathematics 44 (3) (1943) 423–453.
[52] A. N. Kolmogorov, On the log-normal distribution of particles sizes during break-up process, The Proceedings of the USSR Academy of Sciences (Dokladi Adademii Nauk) XXXI (2) (1941) 99–101.
[53] J. van Noortwijk, A survey of the application of gamma processes in maintenance, Reliability Engineering and System Safety 94 (1) (2009) 2–21.
[54] Z. W. Birnbaum, S. C. Saunders, A new family of life distributions, Journal of Applied Probability 6 (2) (1969) 319—327.
[55] B. B. Mandelbrot, The (mis)behavior of markets, Basic Books, 2004.
[56] M. Ouyang, L. D. nas Osorio, An approach to design interface topologies across interdependent urban infrastructure systems, Reliability Engineering and System Safety 96 (2011) 1462–1473.