Chapter 2
System Reliability and Risk Analysis

2.1 System Reliability Analysis

This introduction to system reliability analysis is based on [1]. Historically, it seems that the word reliability was first coined by the English poet Samuel T. Coleridge, who along with William Wordsworth started the English Romantic Movement [2]:

“He inflicts none of those small pains and discomforts which irregular men scatter about them and which in the aggregate so often become formidable obstacles both to happiness and utility; while on the contrary he bestows all the pleasures, and inspires all that ease of mind on those around him or connected with him, with perfect consistency, and (if such a word might be framed) absolute reliability.”

These lines were written by Coleridge in the year 1816, in praise of his friend the poet Robert Southey. From this initial ‘familiar’ use, the concept of reliability grew into a pervasive attribute worth of both qualitative and quantitative connotations. In fact, it only takes an internet search of the word ‘reliability’, e.g., by the popular engine Google, to be overwhelmed by tens of millions of citations [3].

From 1816 to today several revolutionizing social, cultural, and technological developments have occurred which have aroused the need of a rational framework for the quantitative treatment of the reliability of engineered systems and plants and the establishment of system reliability analysis as a scientific discipline, starting from the mid 1950s.

The essential technical pillar which has supported the rise of system reliability analysis as a scientific discipline is the theory of probability and statistics. This theory was initiated to satisfy the enthusiastic urge for answers to gaming and gambling questions by Blaise Pascal and Pierre de Fermat in the 1600s and later expanded into numerous other practical problems by Laplace in the 1800s [3, 4].

Yet, the development of system reliability analysis into a scientific discipline in itself needed a practical push, which came in the early 1900s with the rise of the concept of mass production for the manufacturing of large quantities of goods
from standardized parts (rifle production at the Springfield armory, 1863 and Ford Model T car production, 1913) [3].

But actually, the catalyst for the actual emergence of system reliability analysis was the vacuum tube, specifically the triode invented by Lee de Forest in 1906, which at the onset of WWII initiated the electronic revolution, enabling a series of applications such as the radio, television, radar, and others.

The vacuum tube is by many recognized as the active element that allowed the Allies to win the so-called ‘wizard war’. At the same time, it was also the main cause of equipment failure: tube replacements were required five times as often as all other equipments. After the war, this experience with the vacuum tubes prompted the US Department of Defense (DoD) to initiate a number of studies for looking into these failures.

A similar situation was experienced on the other side of the warfront by the Germans, where chief Engineer Lusser, a programme manager working in Peenemünde on the V1, prompted the systematic analysis of the relations between system failures and components faults.

These and other military-driven efforts eventually led to the rise of the new discipline of system reliability analysis in the 1950s, consolidated and synthesized for the first time in the Advisory Group on Reliability of Electronic Equipment (AGREE) report in 1957. The AGREE was jointly established in 1952 between the DoD and the American Electronics Industry, with the mission of [5]:

- Recommending measures that would result in more reliable equipment;
- Helping to implement reliability programs in government and civilian agencies;
- Disseminating a better education on reliability.

Several projects, still military-funded, developed in the 1950s from this first initiative [5–7]. Failure data collection and root cause analyses were launched with the aim of achieving higher reliability in components and devices. These led to the specification of quantitative reliability requirements, marking the beginning of the contractual aspect of reliability. This inevitably brought the problem of being able to estimate and predict the reliability of a component before it was built and tested: this in turn led in 1956 to the publication of a major report on reliability prediction techniques entitled ‘Reliability Stress Analysis for Electronic Equipment’ (TR-1100) by the Radio Corporation of America (RCA), a major manufacturer of vacuum tubes. The report presented a number of analytical models for estimating failure rates and can be considered the direct predecessor of the influential military standard MIL-HDBK-217F first published in 1961 and still used today to make reliability predictions.

Still from the military side, during the Korean war maintenance costs were found quite significant for some armed systems, thus calling for methods of reliability prediction and optimized strategies of component maintenance and renovation.
In the 1960s, the discipline of system reliability analysis proceeded along two tracks:

- Increased specialization in the discipline by sophistication of the techniques, e.g., redundancy modelling, Bayesian statistics, Markov chains, etc., and by the development of the concepts of reliability physics to identify and model the physical causes of failure and of structural reliability to analyze the integrity of buildings, bridges, and other constructions;
- Shift of the attention from component reliability to system reliability and availability, to cope with the increased complexity of the engineered systems, like those developed as part of military and space programs like the Mercury, Gemini, and Apollo ones.

Three broad areas characterized the development of system reliability analysis in the 1970s:

- The potential of system-level reliability analysis [8] motivated the rational treatment of the safety attributes of complex systems such as the nuclear power plants [9];
- The increased reliance on software in many systems led to the growth of focus on software reliability, testing, and improvement [10];
- The lack of interest on reliability programs that managers often showed already at that time, sparked the development of incentives to reward improvement in reliability on top of the usual production-based incentives.

With respect to methods of prediction reliability, no particular advancements were achieved in those years.

In the following years, the scientific and practicing community has witnessed an impressive increase of developments and applications of system reliability analysis, aimed at rationally coping with the challenges brought by the growing complexity of the systems and practically taking advantage of the computational power becoming available at reasonable costs [1].

The developments and applications of these years have been driven by a shift from the traditional industrial economy, valuing production, to the modern economy centered on service delivery: the fundamental difference is that the former type of economy gives value to the product itself whereas the latter gives value to the performance of the product in providing the service. The good is not the product itself but its service and the satisfaction of the customer in receiving it.

This change of view has led to an increased attention to service availability as a most important quality and to a consequent push in the development of techniques for its quantification. This entails consideration of the fact that availability is a property which depends on the combination of a number of interrelated processes of component degradation, of failure and repair, of diagnostics and maintenance, which result from the interaction of different systems including not only the hardware but also the software, the human, and the organizational and logistic systems.

In this scenario, we arrive at our times. Nowadays, system reliability analysis is a well-established, multidisciplinary scientific discipline which aims at providing
an ensemble of formal methods to investigate the uncertain boundaries around system operation and failure, by addressing the following questions [1, 11, 12]:

- Why systems fail, e.g., by using the concepts of reliability physics to discover causes and mechanisms of failure and to identify consequences;
- How to develop reliable systems, e.g., by reliability-based design;
- How to measure and test reliability in design, operation, and management;
- How to maintain systems reliable, by maintenance, fault diagnosis, and prognosis.

Operatively, the system reliability analysis which addresses the questions above is based on the quantitative definition of reliability in probabilistic terms: considering the continuous random variable failure time $T$, the reliability of the system at time $t$ is the probability that the system does not fail up to time $t$, i.e., the probability that $T$ takes on values larger than $t$.

Another quantity of interest is the system availability, which is used to characterize the ability of a system to fulfill the function for which it is operated. It applies to systems which can be maintained, restored to operation or renovated upon failure depending on the particular strategy adopted to optimally assure its function [1–6]:

- Off-schedule (corrective) maintenance, i.e., replacement or repair of the failed system;
- Preventive maintenance, i.e., regular inspections, and possibly repair, based on a structured maintenance plan;
- Condition-based maintenance, i.e., performance of repair actions upon detection of the degraded conditions of the system;
- Predictive maintenance, i.e., replacement of the system upon prediction of the evolution of the degradation conditions of the system.

The instantaneous availability is defined as the probability that the system is operating at time $t$. It differs from reliability, which is instead used to characterize the ability of the system of achieving the objectives of its specified mission within an assigned period of time, by the probability that the system functions with no failures up to time $t$.

Operatively, the time-dependent, instantaneous availability function of a system is synthesized by point values, e.g.:

- For units or systems under corrective maintenance, the limiting or steady state availability is computed as the mathematical limit of the instantaneous availability function in time as this latter grows to infinity. It represents the probability that the system is functioning at an arbitrary moment of time, after the transient of the failure and repair processes have stabilized. It is obviously undefined for systems under periodic maintenance, for which the limit does not exist;
- For systems under periodic maintenance, the average availability over a given period of time is introduced as indicator of performance. It represents the expected proportion of time that the system is operating in the considered period of time.
2.2 System Risk Analysis

This introduction to system risk analysis is based on [13]. The subject of risk nowadays plays a relevant role in the design, development, operation and management of components, systems, and structures in many types of industry. In all generality, the problem of risk arises wherever there exist a potential source of damage or loss, i.e., a hazard (threat), to a target, e.g., people or the environment. Under these conditions, safeguards are typically devised to prevent the occurrence of the hazardous conditions, and protections are emplaced to protect from and mitigate its associated undesired consequences. The presence of a hazard does not suffice itself to define a condition of risk; indeed, inherent in the latter there is the uncertainty that the hazard translates from potential to actual damage, bypassing safeguards and protections. In synthesis, the notion of risk involves some kind of loss or damage that might be received by a target and the uncertainty of its transformation in an actual loss or damage.

One classical way to defend a system against the uncertainty of its failure scenarios has been to: (i) identify the group of failure event sequences leading to credible worst-case accident scenarios \( \{s^r\} \) (design-basis accidents), (ii) predict their consequences \( \{x^r\} \), and (iii) accordingly design proper safety barriers for preventing such scenarios and for protecting from, and mitigating, their associated consequences [1].

Within this approach (often referred to as a structuralist, defense-in-depth approach), safety margins against these scenarios are enforced through conservative regulation of system design and operation, under the creed that the identified worst-case, credible accidents would envelope all credible accidents for what regards the challenges and stresses posed on the system and its protections. The underlying principle has been that if a system is designed to withstand all the worst-case credible accidents, then it is ‘by definition’ protected against any credible accident [14].

This approach has been the one classically undertaken, and in many technologies it still is, to protect a system from the uncertainty of the unknown failure behaviors of its components, systems, and structures, without directly quantifying it, so as to provide reasonable assurance that the system can be operated without undue risk. However, the practice of referring to ‘worst’ cases implies strong elements of subjectivity and arbitrariness in the definition of the accidental events, which may lead to the consideration of scenarios characterized by really catastrophic consequences, although highly unlikely. This may lead to the imposition of unnecessarily stringent regulatory burdens and thus excessive conservatism in the design and operation of the system and its protective barriers, with a penalization of the industry. This is particularly so for those high-consequence industries, such as the nuclear, aerospace, and process ones, in which accidents may lead to potentially large consequences.

For this reason, an alternative approach has been pushed forward for the design, regulation, and management of the safety of hazardous systems. This approach,
initially motivated by the growing use of nuclear energy and by the growing investments in aerospace missions in the 1960s, stands on the principle of looking quantitatively also at the reliability of the accident-preventing and consequence-limiting protection systems that are designed and implemented to intervene in protection against all potential accident scenarios, in principle with no longer any differentiation between credible and incredible, large, and small accidents [15]. Initially, a number of studies were performed for investigating the merits of a quantitative approach based on probability for the treatment of the uncertainty associated with the occurrence and evolution of accident scenarios [16]. The findings of these studies motivated the first complete and full-scale probabilistic risk assessment of a nuclear power installation [9]. This extensive work showed that indeed the dominant contributors to risk need not be necessarily the design-basis accidents, a ‘revolutionary’ discovery undermining the fundamental creed underpinning the structuralist, defense-in-depth approach to safety [14].

Following these lines of thought, and after several ‘battles’ for their demonstration and valorization, the probabilistic approach to risk analysis (Probabilistic Risk Analysis, PRA) has arisen as an effective way for analysing system safety, not limited only to the consideration of worst-case accident scenarios but extended to looking at all feasible scenarios and its related consequences, with the probability of occurrence of such scenarios becoming an additional key aspect to be quantified in order to rationally and quantitatively handle uncertainty [9, 17–24].

In this view, system risk analysis offers a framework for the evaluation of the risk associated to an activity, process, or system, with the final aim of providing decision support on the choice of designs and actions.

From the view point of safety regulations, this has led to the introduction of new criteria that account for both the consequences of the scenarios and their probabilities of occurrence under a now rationalist, defense-in-depth approach. Within this approach to safety analysis and regulation, system reliability analysis takes on an important role in the assessment of the probability of occurrence of the accident scenarios as well as the probability of the functioning of the safety barriers implemented to hinder the occurrence of hazardous situations and mitigate their consequences if such situations should occur [1].

### 2.2.1 The Framework of PRA

The basic analysis principles used in a PRA can be summarized as follows. A PRA systemizes the knowledge and uncertainties about the phenomena studied by addressing three fundamental questions [24]:

- Which sequences of undesirable events transform the hazard into an actual damage?
- What is the probability of each of these sequences?
- What are the consequences of each of these sequences?
This leads to a widely accepted, technical definition of risk in terms of a set of triplets [22] identifying the sequences of undesirable events leading to damage (the accident scenarios), the associated probabilities and the consequences. In this view, the outcome of a risk analysis is a list of scenarios quantified in terms of probabilities and consequences, which collectively represent the risk. On the basis of this information, the designer, the operator, the manager, and the regulator can act effectively so as to manage (and possibly reduce) risk.

In the PRA framework, knowledge of the problem and the related uncertainties are systematically manipulated by rigorous and repeatable probability-based methods to provide representative risk outcomes such as the expected number of fatalities (in terms of indices such as Potential Loss of Lives (PLL) and Fatal Accident Rate (FAR), the probability that a specific person shall be killed due to an accident (individual risk) and frequency-consequence (f-n) curves expressing the expected number of accidents (frequency f) with at least n fatalities.

In spite of the maturity reached by the methodologies used in PRA, a number of new and improved methods have been developed in recent years to better meet the needs of the analysis, in light of the increasing complexity of the systems and to respond to the introduction of new technological systems [1]. Many of the methods introduced allow increased levels of detail and precision in the modeling of phenomena and processes within an integrated framework of analysis covering physical phenomena, human and organisational factors as well as software dynamics (e.g., [25]). Other methods are devoted to the improved representation and analysis of the risk and related uncertainties, in view of the decision making tasks that the outcomes of the analysis are intended to support. Examples of newly introduced methods are Bayesian Belief Networks (BBNs), Binary Digit Diagrams (BDDs), multi-state reliability analysis, Petri Nets, and advanced MCS tools. For a summary and discussion of some of these models and techniques, see [1] and [20].

The probabilistic analysis underpinning PRA stands on two lines of thinking, the traditional frequentist approach and the Bayesian approach [19, 20]. The former is typically applied in case of large amount of relevant data; it is founded on well-known principles of statistical inference, the use of probability models, the interpretation of probabilities as relative frequencies, point values, confidence intervals estimation, and hypothesis testing.

The Bayesian approach is based on the use of subjective probabilities and is applicable also in case of scarce amount of data. The idea is to first establish adequate probability models representing the aleatory uncertainties, i.e., the variabilities in the phenomena studied, such as for example the lifetimes of a type of unit; then, the epistemic uncertainties (due to incomplete knowledge or lack of knowledge) about the values of the parameters of the models are represented by prior subjective probability distributions; when new data on the phenomena studied become available, Bayes’ formula is used to update the representation of the epistemic uncertainties in terms of the posterior distributions. Finally, the predictive distributions of the quantities of interest (the observables, for example
the lifetimes of new units) are derived by applying the law of total probability. The predictive distributions are subjective but they also reflect the inherent variability represented by the underlying probability models.

### 2.2.2 Uncertainty Analysis

Uncertainty is an unavoidable component affecting the behaviour of systems and more so with respect to their limits of operation. In spite of how much dedicated effort is put into improving the understanding of systems, components and processes through the collection of representative data, the appropriate characterization, representation, propagation and interpretation of uncertainty remains a fundamental element of the risk analysis of any system. Following this view, uncertainty analysis is considered an integral part of PRA, although it can also exist independently in the evaluation of unknown quantities.

In the context of PRA, uncertainty is conveniently distinguished into two different types: randomness due to inherent variability in the system (i.e., in the population of outcomes of its stochastic process of behavior) and imprecision due to lack of knowledge and information on the system. The former type of uncertainty is often referred to as objective, aleatory or stochastic whereas the latter is often referred to as subjective, epistemic, or state-of-knowledge [26–29]. Probability models are introduced to represent the aleatory uncertainties, for example a Poisson model to represent the variation in the number of events occurring in a period of time. The epistemic uncertainties arise from a lack of knowledge of the parameters of the probability models. Whereas epistemic uncertainty can be reduced by acquiring knowledge and information on the system, the aleatory uncertainty cannot, and for this reason it is sometimes called irreducible uncertainty.

In all generality, the quantitative analyses of the phenomena occurring in many engineering applications are based on mathematical models that are then turned into operative computer codes for simulation. A model provides a representation of a real system dependent on a number of hypotheses and parameters. The model can be deterministic (e.g., Newton’s dynamic laws or Darcy’s law for groundwater flow) or stochastic (e.g., the Poisson model for describing the occurrence of earthquake events).

In practice, the system under analysis cannot be characterized exactly—the knowledge of the underlying phenomena is incomplete. This leads to uncertainty in both the values of the model parameters and on the hypotheses supporting the model structure. This defines the scope of the uncertainty analysis.

An uncertainty analysis aims at determining the uncertainty in analysis results that derives from uncertainty in analysis inputs [29–31]. We may illustrate the ideas of the uncertainty analysis by introducing a model $G(X)$, which depends on the input quantities $X$ and on the function $G$; the quantity of interest $Z$ is computed by using the model $Z = G(X)$. The uncertainty analysis of $Z$ requires an assessment of the uncertainties of $X$ and their propagation through the model $G$ to produce a
characterization of the uncertainties of $Z$. Typically, the uncertainty related to the model structure $G$, e.g., uncertainty due to the existence of alternative plausible hypotheses on the phenomena involved, are treated separately \cite{27, 32–34}; actually, while the first source of uncertainty has been widely investigated and more or less sophisticated methods have been developed to deal with it, research is still ongoing to obtain effective and accepted methods to handle the uncertainty related to the model structure \cite{35}. See also \cite{36} which distinguishes between model inaccuracies (the differences between $Z$ and $G(X)$), and model uncertainties due to alternative plausible hypotheses on the phenomena involved.

The traditional tool used to express the uncertainties in PRA is (subjective) probabilities. In this context, the quantities $X$ and $Z$ could be chances representing fractions in a large (in theory infinite) population of similar items (loosely speaking, a chance is the Bayesian term for a frequentist probability, cf. the representation theorem of de Finetti \cite{37}, \cite{38}, p. 172). In this case, the assessment is consistent with the so-called probability of frequency approach, which is based on the use of subjective probabilities to express epistemic uncertainties of unknown frequencies, i.e., the chances \cite{22}. The probability of frequency approach constitutes the highest level of uncertainty analysis according to a commonly referenced uncertainty treatment classification system \cite{39}.

Recently, many researchers have argued that the information commonly available in the practice of risk decision making does not provide a sufficiently strong basis for a specific probability assignment; the uncertainties related to the occurrence of the events and associated consequences are too large. Furthermore, in a risk analysis context there are often many stakeholders and they may not be satisfied with a probability-based assessment expressing the subjective judgments of the analysis group: again a broader risk description is sought.

Based on the above critiques, it is not surprising that alternative approaches for representing and describing uncertainties in risk assessment have been suggested, which produce epistemic-based uncertainty descriptions and in particular probability intervals.

Work has also been carried out to combine different approaches, for example probabilistic analysis and possibility theory. Here the uncertainties of some parameters are represented by probability distributions and those of some other parameters by means of possibilistic distributions. An integrated computational framework has been proposed for jointly propagating the probabilistic and possibilistic uncertainties \cite{40}. This framework has been tailored to event tree analysis \cite{41} and Fault Tree Analysis (FTA) \cite{42}, allowing for the uncertainties about event probabilities (chances) to be represented and propagated using both probability and possibility distributions.
References


The Monte Carlo Simulation Method for System Reliability and Risk Analysis
Zio, E.
2013, XIV, 198 p., Hardcover