CHAPTER 1

RELIABILITY FOR THE 21ST CENTURY

Sallie A. Keller-McNulty and Alyson G. Wilson
Statistical Sciences Group
Los Alamos National Laboratory
P. O. Box 1663, D-1, MS F600
Los Alamos, NM 87545 USA
E-mail: sallie@lanl.gov

The sophistication of science and technology is growing almost exponentially. Government and industry are relying more and more on science's advanced methods to assess reliability coupled with performance, safety, surety, cost, schedule, etc. Unfortunately, policy, cost, schedule, and other constraints imposed by the real world inhibit the ability of researchers to calculate these metrics efficiently and accurately using traditional methods. Because of such constraints, reliability must undergo an evolutionary change. The first step in this evolution is to reinterpret the concepts and responsibilities of scientists responsible for reliability calculations to meet the new century's needs. The next step is to mount a multidisciplinary approach to the quantification of reliability and its associated metrics using both empirical methods and auxiliary data sources, such as expert knowledge, corporate memory, and mathematical modeling and simulation.

1. Introduction

By definition, *reliability* is the probability a *system* will perform its intended function for at least a given period of time when operated under some specified conditions. The 20th century solution to this problem[a] has been to define a reliability function as

$$R(t) = P(T > t) = \int_t^\infty f(x)dx = 1 - F(t) \tag{1}$$

[a]An excellent review of reliability theory and corresponding references can be found in Martz (2000).[1]

and to use the function as the basis of definition for other important concepts, such as failure rate and mean time between failures. Powerful parametric (e.g., binomial, Poisson, exponential, Weibull) and nonparametric statistical models have been developed to estimate reliability and its associated properties. These traditional reliability methods were developed for industrial, mass produced products such as electronics and consumer goods. Everything works quite nicely provided we have coherent system representations and clean, typically single, sources of quantitative data about the system. Problems today, however, are much more complex and include systems such as nuclear weapons, infrastructure networks, supercomputer codes, jumbo jets, etc. These systems demand more of reliability than our current methodology allows. In many instances it is not possible to mount vast numbers of full system tests, and frequently none are available.[2]

System assessment is complicated by the need to consider more than what has been traditionally considered as reliability, because a system's ability to perform is intertwined with other concepts such as its age, safety, and surety. In addition, our ability to do reliability assessments may be severely constrained by policy, cost, and schedule, particularly in problems dealing with the inherent reliability of an existing system. Therefore, we must expand our definition of the system to include all aspects that affect its performance and all constraints (e.g., test schedule) that affect the confidence we have in the assessment. The end result should be a reliability assessment that is an expression of our complete state of knowledge about the system.

Statisticians are frequently the scientists responsible for driving the reliability assessment process. Due to the demands stated above, their roles in this process must correspondingly (and significantly) broaden. This paper provides a broad overview of some of the concepts and research that need to be brought together to address the reliability challenges of the 21st century. Section 2 motivates the fact that these are decision, not simply analysis or modeling, problems. Section 3 outlines the diverse research areas needed, Section 4 gives motivating examples, and Section 5 provides conclusions.

2. Decision Context

Today's reliability problems are driven by the need to support decision-making at any point in the life of the system under study, using the broader definition of system given above. This requires continuous, over the life of the system, integration of all information and knowledge into the decision-

making process. Traditional decision analysis is based on objective functions and constraints that are well-defined and static. This is a serious flaw and may be the reason why formal decision-analytic projects frequently do not succeed. In addition, the incorporation of scientific knowledge, such as that derived from reliability assessments, into the decision-making process is weak. This is due to the inability to model evolving scientific discovery quickly and clearly and communicate the results in useful ways to decision-makers. The development of rigorous statistical and mathematical methods for this integration and evolution holds the promise for bridging the gap between narrowly defined reliability analyses, based on Equation 1, and these dynamic decision problems.

Once it is understood that the true *system* problem is a decision problem, this introduces the realization that diverse information sources, not simply test data, must be clearly understood and modeled. This is because diverse sources of information are what is used to guide prudent decision-making. More specifically, the diversity of scientific information that is used to support decision-making arises from:

- the sources of information, including theoretical models, test data, observational data, computer simulations, and expertise from scientists, field personnel, and decision-makers;
- the content of the information, including information about the system structure and behavior, decision-maker constraints, options, and preferences; and
- the multiple communities of practice that are the stakeholders in the decision process.

Figure 1 is a graphical view of the integrated reliability assessment that is needed to capture the full context of the decision problem, including the information sources listed above. Techniques from various disciplines (statistics, probability, mathematics, computer science, decision theory, graph theory, expert knowledge elicitation and representation, and simulation) must be merged to develop formal methods to integrate the multiple information sources relative to the content of the system evaluation (e.g., simultaneous assessment of performance, reliability, sustainability, dependability, safety, etc.). These assessments must support the needs of the stakeholders, such as complex resource allocation (e.g., mount a full-system test, build a new experimental facility, develop more scientific capability, implement a higher fidelity computer model) and continuous evaluation (i.e., supporting decisions that need to be made at, frequently unantici-

pated, multiple points in time). The methods must explicitly incorporate uncertainty and allow for dynamic changes in the evolution of knowledge and processes about the system being studied, thus enabling a flow of new information to support continuous decision-making. The solution and development of such methods is the 21st century challenge for reliability.

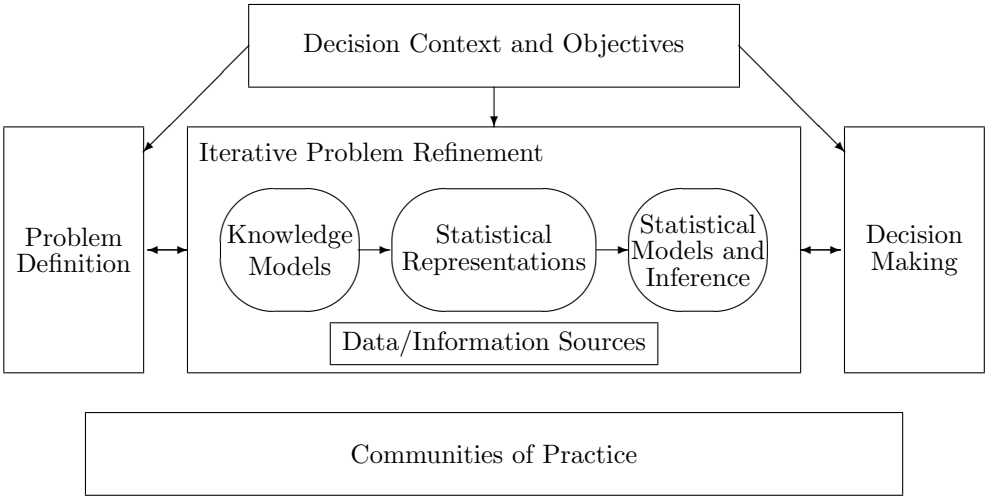

Fig. 1. Schematic view of integrated reliability assessment.

3. Overview of Research Areas

This section provides a nontechnical overview of several, possibly nonobvious, research areas that must be further developed and integrated if the 21st century reliability challenges are to be realized. Section 4 demonstrates the need for these research areas and points to gaps in what is well understood vis-*à*-vis specific examples.

3.1. *Knowledge Representation*

If the goal is to support continuous decision-making for highly complex and innovative systems, statisticians will be frequently faced with the dilemma of minimal data for the system under scrutiny. Complicating such situations is the increasing ubiquity of multidisciplinary and multinational research

teams. Statisticians often find themselves asked to contribute to complex, emergent projects that challenge their ability to build predictive models capable of integrating multiple types of data, information, and knowledge from a wide range of sources. Therefore, there is a need to develop a multidisciplinary approach to knowledge elicitation, representation, and transformation. This approach must mesh techniques from cultural anthropology, computer science, and statistics to address the complexities of multidisciplinary research.

Specifically, elicitation techniques derived from cultural anthropology can be used to elicit tacit problem-solving structures from the "natives"—generally, the scientists and engineers collaborating on difficult system problems. The elicited information, in turn, can be used to develop ontologies that represent the problem space in the "native language" of the research team, but which are more mathematically tractable to the computer science and statistical communities. Iterative cycles of representational refinement and quantification will lead to the emergence of predictive statistical models that make intuitive sense to all parties: the scientists, engineers, elicitation experts, knowledge modelers, and statisticians.[3] Important methodological challenges include advancing research on conceptual graphs, statistical graphical models, and the translation from qualitative to quantitative representations.

3.2. *Statistical Methods*

The need for aggressive research in statistical methods is clear. The research must address the treatment of heterogeneous and diverse information sources. Because pure, full-system testing approaches are frequently infeasible, methods that can parse the problem into subsystems and constituently fold together subsystem analyses to build full-system predictions must be developed. Mathematically, this is a complex, ill-posed problem. Solutions may easily take the statistician outside the comfort zone of traditional probability theory and into areas of fuzzy measurement, belief functions, and possibility theory. Methodological focuses will clearly include Bayesian inference, Bayesian hierarchical models, and computational methods for reliability and lifetime estimation; system reliability and lifetime analysis; computer model evaluation; methods for accelerated life testing; models for degradation data; and demonstration testing.[4,5]

Special challenges arise when the analyses must include expert judgment or the output from computer models. There is a body of work that looks at

various aspects of expert judgment, from the construction of priors for statistical models, to the development of utility functions and decision analyses within economics, to the anthropological and psychological work on elicitation strategies.[6,7] Research is needed on the development of a conceptual framework for utilizing "statistical" information elicited from experts. When presented with a dataset containing experimental data, statisticians have a "toolkit" of methods and a set of canonical examples to use when deciding how to analyze the data. This toolkit is still absent from the realm of expert judgment.

Another important source of data that must be leveraged into 21st century system reliability problems is that derived from computer simulations. These computer models are often highly complex themselves, and there are many research challenges associated with the calibration and validation of these models under the conditions of limited real world data. Simulation data are often limited as well, because the models require many hours, days, or even months to run. Another challenge to the evaluation of these simulation data is the high dimensionality of the output. New research in the areas of complex computer model evaluation, applying techniques from sensitivity analysis, Bayesian interpolation methods, and extensions of techniques from spatial statistics is beginning to formalize the use and limitations of such data.[8–13]

Traditional experimental design is concerned with allocating trials within a single experiment. Suppose, however, that data is available from many diverse kinds of experiments including different types of physical experiments (e.g., destructive or nondestructive tests) and runs of a computer code. The newly coined area of *hybrid experimental design*[14] considers the allocation of test resources across different types of experiments by trading off the costs of performing any particular trial with the information gained. Significant research on the extension and optimization of such methods is needed.

3.3. *Knowledge Management*

The way in which information is organized has a major influence on how that information is used. With the wide range of information (and knowledge) needed to solve the problems for continuous decision-making and evaluation, formal organization of information is critical. This will require a variety of tools to capture information, organize it, and make the results available for subsequent analysis to the distributed communities working on

a problem. Significant research is needed in knowledge management with a focus on the development of these tools both for the statistical and knowledge modeling researchers and for the stakeholders.

4. Motivating Examples

This section contains three current problems that illustrate the complexity of "system assessment" and the need for rigorous mathematical solutions. All three examples are defense related, which is fitting if one considers the role that defense agencies around the world have played in the development and adoption of reliability methods.

4.1. *Science-Based Stockpile Stewardship*

An example of the complexity facing scientists responsible for reliability assessment is Science Based Stockpile Stewardship (SBSS) at Los Alamos National Laboratory (LANL) and the history that has brought about this problem. From its earliest days, LANL has had a prominent role in the development and evaluation of the United States nuclear weapons stockpile, but the end of the Cold War brought significant changes to how this mission could be carried out. There have been significant reductions in the number of weapons, leading to a smaller, "enduring" stockpile. The United States is no longer manufacturing new-design weapons, and it is consolidating facilities across the nuclear weapons complex. In 1992, the United States declared a moratorium on underground nuclear testing; in 1995, the moratorium was extended, and President Clinton decided to pursue a "zero yield" Comprehensive Test Ban Treaty. However, the basic mission of LANL remains unchanged: LANL must evaluate the weapons in the aging nuclear stockpile and certify their safety, reliability, and performance even though the live test data that have traditionally been used for this evaluation can no longer be collected.

To complete this mission, a two-pronged approach of experiments and computational modeling was adopted. The experimental approach is exemplified by the Dual-Axis Radiography for Hydrotesting (DARHT) facility, which enables experimenters to better understand the nature of explosions. The computational modeling effort is exemplified by the Accelerated Strategic Computing Initiative (ASCI), which uses supercomputers to model the types of complex nuclear experiments that are no longer performed. At a fundamental level, though, the new experimental and computer technologies have not been developed to address SBSS; rather a "zero yield" policy

could be negotiated and implemented because advances in computer technology made it seem feasible that the sophisticated modeling could be done to realize SBSS. In short, the promise of the technology drove the policy. It created an expectation that certain tough questions could be answered with adequate justification.

Alongside the efforts at experimentation and modeling, statisticians and knowledge modelers have been working to integrate historical data and to quantify the vast resources of expertise at LANL in such a way as to facilitate their inclusion through Bayesian statistical methods. The challenge is to integrate experimental data, computational models, past tests, subsystem tests, and the expert judgment of subject-matter experts to provide a rigorous, quantitative assessment, with associated uncertainties, of the safety, reliability, and performance of the stockpile.

Without careful attention to the whole picture, or purpose of the system assessment, the accomplishments of individual scientists can become lost and detached. Figure 2 is a notional representation of several elements of the SBSS problem. Within parts of the figure, traditional methodology works well for various questions. For example, event-tree methods can be used to define the critical paths for successful completion on the physical experiments and the risks involved that could affect the schedule. But, what happens if an experiment that is needed to help resolve some of the equation-of-state parameterizations for the computational experiments cannot be done? The uncertainty that results must be propagated through the computational models and accounted for in our statements about confidence in our assessments. This in turn will affect the design of other computational experiments. This is not a standard problem addressed through traditional reliability analysis.

The engineering portion of certification depicted in Figure 2 can be thought of as a traditional engineering reliability problem based on coherent system representations. However, there is rarely direct data available on all parts of the system. Therefore, we must develop methodology that can integrate other, related information and be able to propagate information up and down throughout the system representation.[4,5,15] A major challenge is to then integrate the engineering reliability information with the physics performance assessment, material degradation models, etc. In contrast to the discrete nature of the engineering component condition representations of coherent systems, the physics is represented as continuous, time-dependent, integrated processes. It is these two elements, engineering and physics, in combination, that are needed to understand the condition

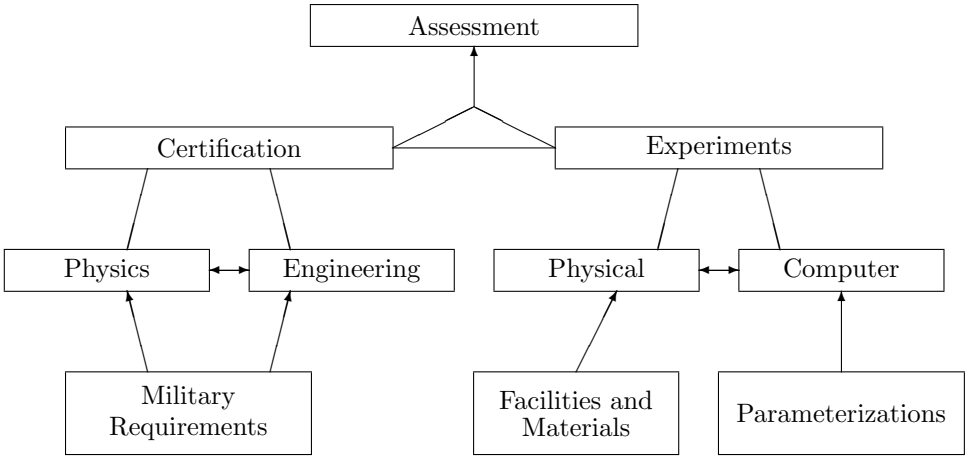

Fig. 2. Notional represesntation of an integrated assessment process for SBSS.

of the enduring stockpile. Once again our traditional reliability representations and treatments of problems do not address this integrated assessment.

4.2. *Conventional Munitions Stockpile Surveillance*

An extension of the SBSS problem is the development of methodology for assessing the reliability of aging stockpiles of conventional munitions. These problems are similar to the assessment of nuclear weapons stockpiles, but there is usually some full-system testing. Stockpile surveillance gathers many kinds of data, including full-system tests, component and subsystem information, and nondestructive evaluation. In addition, there are computer codes that can predict properties of materials aging. The challenge is the combination of information to support decisions about stockpile life extension programs. Figure 3 gives a notional representation of a system, its subsystems, and the type of information available at a snapshot in time for this system.

For this problem, a Bayesian hierarchal model has been developed,[15] where the hierarchy reflects the similar component behavior in Figure 3. The hierarchy was developed using knowledge modeling techniques and formal elicitation. With the model, information can be leveraged across the hierarchy to build reliability distributions for components with no test or

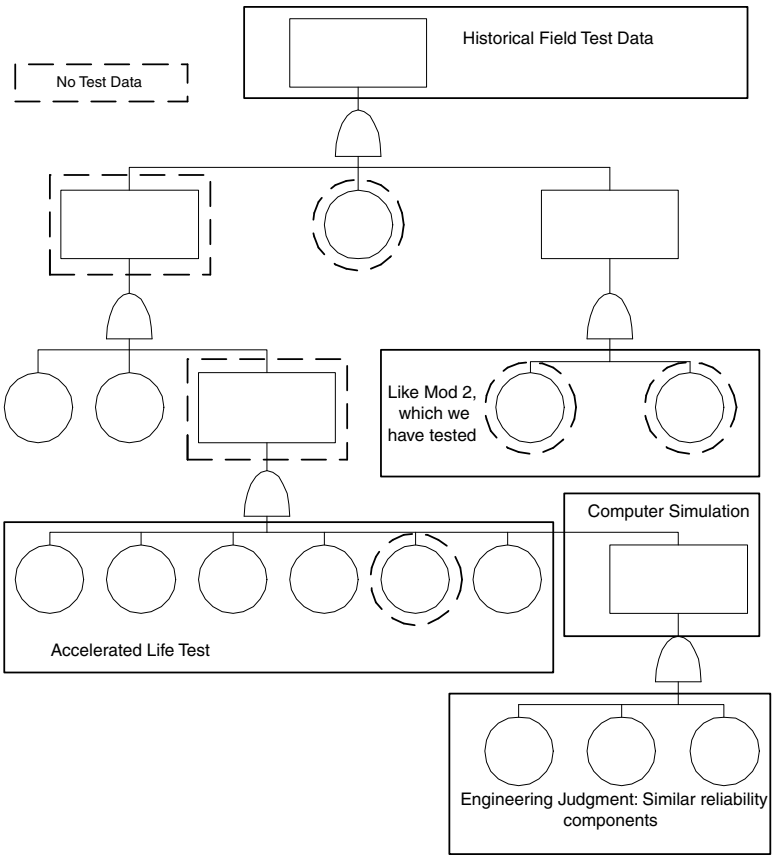

Fig. 3. Structural display of information sources for anti-aircraft missile problem.

computer simulation data, and improve the reliability distributions for components with data. As more data become available at the subsystem and component levels, the hierarchical model is used to do a downward propagation. This downward propagation results in a belief function representation of the subsystem and component reliabilities. Figure 4 shows the estimated reliability distribution for the full system and for two subsystems. For one of the subsystems, no test data existed. In all three cases, reliability distributions that included information for the full-system flight tests are displayed and estimated distributions without that information are displayed. As one would expect, more uncertainty exists when full-system test data are not used (available). The model is being extended to take into account time-

dependent changes in the system architecture and material degradation. The model is also being extended to account for covariates, e.g., storage conditions. The ability to solve the covariate and time-dependent problem will be what helps support the dynamic and continuous decisions that must be made regarding this missile system.

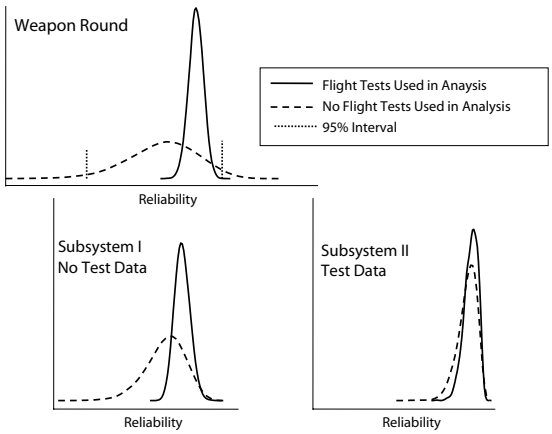

Fig. 4. Reliability estimation for anti-aircraft missile stockpile surveillance.

4.3. *Mission Success for a Ballistic Missile*

As a third example, consider a research and development program that is charged with developing, flying, and collecting data about ballistic missile systems that emulate the flight characteristics of "threat" missiles. These flights are expensive and politically visible, and technical risk mitigation is an important element of the program. These flight tests are also very difficult, because they are "one-of-a-kind" events with many complex factors. One organization is in charge of project management, cost controls, and scheduling; another is in charge of building the missile booster to send the rocket into the upper atmosphere; a third is in charge of building the missile payload.

The program managers approached LANL with a specific problem: how does one develop a predictive reliability model for an engineering system that is still in the design stages? Multiple concerns drove this question: the rocket development program is extremely expensive. Only one or two of the prototypes are built and flown and are usually destroyed in the

process; rarely are the engineers able to salvage subsystems for reuse in further iterations of the program. Because each system flown is unique, there is little direct performance or reliability data available for parts or subsystems on the test rocket. Hence the program managers had little idea how to make predictions or assess risk areas for the flights.

The goal is to develop an integrated, predictive reliability and performance model for an upcoming flight. Such a model will allow the stakeholders to pinpoint risk areas and to make clearly informed decisions about resource allocation to mitigate the risks and maximize the opportunity for mission success.

In developing the model, LANL developed a model framework that captured the critical interactions among the rocket's subsystems during flight. Figure 5 is the ontology developed for the problem—it maps the basic relationships and concepts within the problem and serves as a basis for subsequent knowledge system development. Many sources of data and information that the engineers used to build confidence in their rocket before flight were elicited and documented. This information was used to develop a Bayesian network (Figure 6) that forms the statistical basis for combining multiple sources of information in a rigorous, quantitative framework.

5. Conclusions

Traditional statistical science approaches to reliability based strictly on the reliability function given in Section 1 are no longer sufficient to address the reliability assessment process for multifaceted 21st century problems.[16] The complexities of big science problems such as SBSS, and the other examples in Section 4, demonstrate the impossibility of static system solutions. Today the overall assessment process is more about "decision-making" than "modeling." Many problems, such as these, are politically and economically charged. Therefore, even the best data collection design and corresponding statistical models for the problem at hand may not be feasible, or even allowed. Our 21st century reliability challenge is to be able to structure and overlay statistical models on integrated assessment processes, such as that represented in Figure 1. These models will need to be robust enough to support decision-making at various resolutions, (e.g., about a specific experiment, engineering component design, or facility resource allocation to support the overall assessment process). States of knowledge about the system will be a collection of heterogeneous and diverse sources of information. These sources of information will need to be integrated via tractable

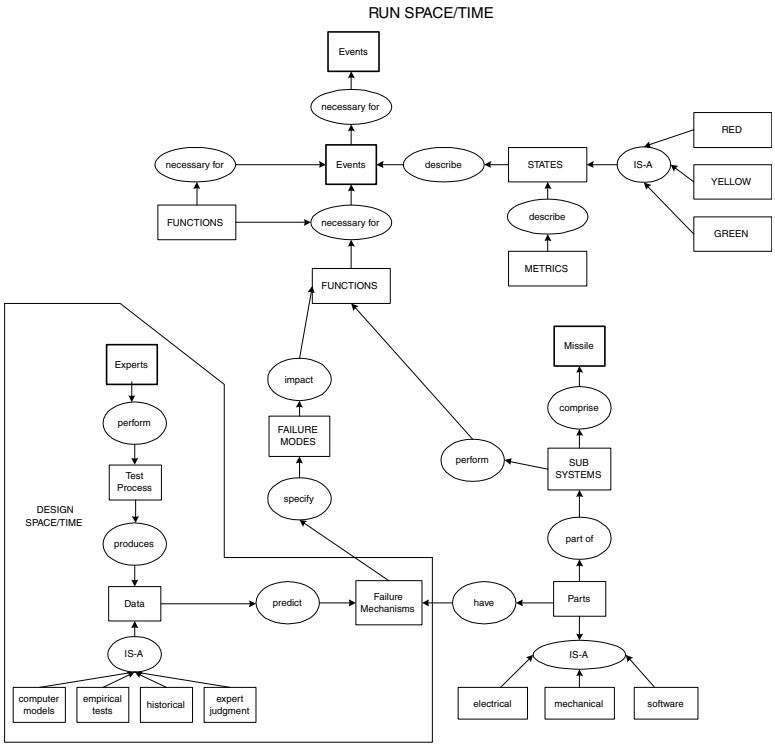

Fig. 5. Ballistic missile system ontology.

mathematical models. The information will be coming from very different disciplines, (e.g., physics, materials, chemistry, and engineering). Therefore, uncertainty quantification inherent in the statistical models will need to be flexible to account for natural ways to represent the information (e.g., probability, fuzzy measures, belief functions, possibility theory, etc). With these challenges come wonderful opportunities for the advancement of reliability analysis and the significant advancement of science.

Acknowledgments

The authors acknowledge several colleagues involved in the problems described in Section 4. These include Art Dempster, Todd Graves, Michael Hamada, Nicholas Hengartner, David Higdon, Valen Johnson, Deborah Leishman, Laura McNamara, Mark McNulty, Jerome Morzinski, Shane Reese, and Nozer Singpurwalla. Los Alamos National Laboratory report LA-UR-02-7283.

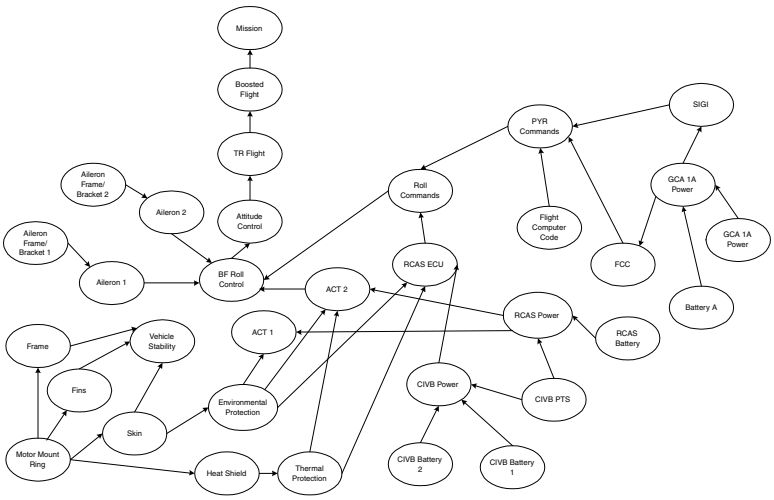

Fig. 6. Bayesian network for the ballistic missile system.

References

1. H. Martz, Reliability Theory. In *Encyclopedia of Physical Science and Technology*, Volume **14**, pp. 143-159 (2002).
2. T. Bement, J. Booker, S. Keller-McNulty, and N. Singpurwalla, Testing the Untestable: Reliability in the 21st Century, *IEEE Transactions on Software Reliability* (to appear).
3. D. Leishman and L. McNamara, "Interlopers, Translators, Scribes, and Seers: Anthropology, Knowledge Representation and Bayesian Statistics for Predictive Modeling in Multidisciplinary Science and Engineering Projects," Conference on Visual Representations and Interpretations 2002, Liverpool, UK (2002).
4. A. Wilson, C. S. Reese, M. Hamada, and H. Martz, "Integrated Analysis of Computational and Physical Experimental Lifetime Data," Mathematical Reliability: An Expository Perspective, eds. R. Soyer, N. Singpurwalla, T. Mazzuchi. Kluwer Press (2001).
5. C. S. Reese, A. Wilson, M. Hamada, H. Martz, and K. Ryan, "Integrated Analysis of Computer and Physical Experiments," Los Alamos National Laboratory report LA-UR-00-2915, submitted to Technometrics (2001).
6. J. Booker and L. McNamara, "Expertise and Expert Judgment in Reliability Characterization: A Rigorous Approach to Eliciting, Documenting, and Analyzing Expert Knowledge," Engineering Design Reliability Handbook, eds. D. Ghiocel and S. Nikolaidis. CRC Press (2002).
7. M. Meyer and J. Booker, Eliciting and Analyzing Expert Judgment: A Practical Guide, 2nd Edition. American Statistical Association and the Society for Industrial and Applied Mathematics (2001).

8. M. J. Bayarri, J. Berger, D. Higdon, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, and J. Tu, "Framework for Validation of Computer Models," Foundations 2002: Workshop on model and simulation verification and validation, Johns Hopkins University, Applied Physics Laboratory, Laurel, MD (2002).
9. R. Berk, P. Bickel, K. Campbell, R. Fovell, S. Keller-NcNulty, E. Kelly, R. Linn, B. Park, A. Perelson, N. Roupail, J. Sacks, and F. Schoenberg, "Workshop on Statistical Approaches for the Evaluation of Complex Computer Models," *Statistical Science* **17(2)**, 173-192 (2002).
10. D. M. Higdon, H. Lee, and C. Holloman, "Markov Chain Monte Carlo-based Approaches for Inference in Computationally Intensive Inverse Problems," Bayesian Statistics **7**, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. West. Oxford University Press (2002).
11. H. Lee, D. M. Higdon, Z. Bi, M. Ferreira, and M. West, "Markov Random Field Models for High-Dimensional Parameters in Simulations of Fluid Flow in Porous Media," *Technometrics* **44(3)**, 230-241 (2002).
12. M. D. McKay, "Nonparametric Variance-based Methods of Assessing Uncertainty Importance," *Reliability Engineering and System Safety* **57**, 267-279 (1997).
13. M. D. McKay, J. D. Morrison, and S. C. Upton, "Evaluating Prediction Uncertainty in Simulation Models," *Computer Physics Communications* **117**, 44-51 (1999).
14. M. Hamada, H. Martz, C. S. Reese, and A. Wilson, "Finding Near-Optimal Bayesian Experimental Designs via Genetic Algorithms," *The American Statistician* **55(3)** (2001).
15. V. Johnson, T. Graves, M. Hamada, and C. S. Reese, "A Hierarchical Model for Estimating the Reliability of Complex Systems," Bayesian Statistics 7, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, M. West. Oxford University Press (2002).
16. S. Keller-McNulty, G. Wilson, and A. Wilson, "Integrating Scientific Information: Reconsidering the Scientific Method as a Path to Better Policy Making," Los Alamos National Laboratory Report LA-UR-01-5739 (2001).