LA-UR-00-4737

Approved for public release; distribution is unlimited.

Title: PREDICT: A NEW APPROACH TO PRODUCT DEVELOPMENT AND LIFETIME ASSESSMENT USING INFORMATION INTEGRATION TECHNOLOGY

Author(s): Jane M. Booker, TSA-1
Thomas R. Bement, TSA-1
Mary A. Meyer, TSA-1
Los Alamos National Laboratory

William J. Kerscher, III
Delphi Automotive Systems

Submitted to: Statistics in Industry
September 29, 2000

PREDICT: A New Approach to Product Development and Lifetime Assessment Using Information Integration Technology

Jane M. Booker, Thomas R. Bement, and Mary A. Meyer
Statistical Sciences Group
Los Alamos National Laboratory

William J. Kerscher, III
Delphi Energy and Chassis Systems
Delphi Automotive Systems

Submitted to *Handbook of Statistics: Statistics in Industry*

September 29, 2000

1. Introduction

Information Integration Technology is a combination of processes, methods, and tools for collecting, organizing, and analyzing diverse information and for utilizing that information to guide optimal decision making. PREDICT (Performance and Reliability Evaluation with Diverse Information Combination and Tracking) is a highly successful example of information integration technology that has been applied in two parallel applications, automotive system development and stockpile physics packages in nuclear weapons. Specifically the PREDICT application is a formal, multidisciplinary process for estimating the performance of a product when test data are sparse or nonexistent. The acronym indicates the purpose of the methodology: to evaluate the performance or reliability of a product/system by combining all available (often diverse) sources of information and then tracking that performance as the product undergoes changes. PREDICT's calculations have been demonstrated to guide product development for automotive systems before, during, and after prototyping and production, and documents the product's performance through its lifetime from concept through customer use and maintenance.

PREDICT is a methodology that allows users to estimate reliability early in product development, before costly design and production decisions are made by making effective use of all available information: expert knowledge, historical information, experience with similar products, and computer model outputs. Until now, much of this information (especially expert knowledge) was not formally included in performance calculations because it was either implicit, undocumented or non-numeric. In PREDICT, all available information (with appropriate uncertainties attached) is collected and combined for estimating the reliability of the product at various stages in its lifetime.. The PREDICT methodology has been used to provide accurate reliability estimates for potential products while they were only engineering concepts. As the product undergoes changes during its development stage, or as conditions change, or new information becomes available, the reliability estimates are updated accordingly, providing a lifetime track record of the performance of the product or system.

PREDICT's philosophy and uniqueness arise from these aspects of the methodology:

- All available, and often diverse, sources of information along with their associated uncertainties are combined. Sources of information include expert knowledge, historical test data, data from of similar systems, parts, processes, etc., design specifications, production information, maintenance records, computer simulation model outputs, physical model / code outputs (both stochastic and deterministic), and test, experimental or observational data.
- The multidisciplinary methods used for integrating diverse information are formalized, especially regarding the use of expert knowledge (Meyer and Booker, 1991).
- As the product or system undergoes change (e.g., under development, or aging), the information integration methods "update" the performance estimates in light of the

information associated with this change. This updating can be in the form of traditional Bayesian updating, and can also be in the form of changes in the system itself or other conditions affecting the performance.

- The dynamics of the system are captured with updating analysis methods which not only track the changing state of knowledge but also provide planning tools for timely decision making. *What-if* questions can be posed to provide guidance for resource allocation and identify areas where reliability could be improved and/or uncertainty reduced by certain actions.
- The choices, modifications, and uses of the methods are customized to the application and its users. They are tailored to the community of practice's way of thinking about the product and its performance/reliability. For example, some engineers might think in terms of failures per million parts, while other would think if performance in terms of cycles per second.
- Emphasis is placed on the need for definitions and requirements for performance, and the importance of structuring the problem. Otherwise, all the information gathered has little purpose, and there is no rationale for its combination.
- Analysis is performed at all of the system's multiple levels and dimensions. The performance of most systems is not merely the successful operation of multiple pieces, but involves failure mechanisms and activities such as quality control, manufacturing processes, chemistry, physics, mechanics, etc.
- The analysis tools and methods are customized to the application and the ways that the community thinks about performance. That community then owns the analysis, the results and the tailored methods for their use. The PREDICT methodology becomes apart of the way they monitor reliability and performance.
- Documentation is a vital aspect of the methodology to provide a traceable record for updating and an understanding of the dynamic environment.

For non-statisticians, the statistics involved are not scary or difficult and are tailored to fit their "culture" or community of practice. For the engineer, PREDICT is a practical, logical, and useful methodology as reflected by comments such as "This can actually help me do my job." For the decisionmaker or program manager, PREDICT is a planning method, for determining the allocation of resources to improve performance and/or reduce uncertainty.

The sections that follow outline the applications, implementation steps, expert judgment, statistical tools, and decision making that make up the PREDICT methodology. The examples presented are notional for illustrative purposes only. The conclusion includes a discussion of research topics for continued development of the Information Integration Technology.

2. The Applications

PREDICT's success is evidenced by the benefits created in two applications where sparse data precluded traditional statistical reliability analysis (Mann, Schafer, Singpurwalla, 1974, Meeker and Escobar, 1998). Both systems were in dynamic environments: one system – fuel systems for Delphi Automotive Systems — undergoing changes because of development from concept to production, and the other system — stockpiled physics packages in nuclear weapons — already built to specifications but undergoing aging and maintenance changes. Each application presented new challenges for analysis. It was the success of PREDICT in these applications that attracted the attention of the R&D 100 Award judges in that national competition for the 100 top technologies of 1999 (AMSTAT News, 1999, Meyer, et. al, 1999).

While aging and maintainability issues are a traditional statistical problem, they require test data for the full system and its components. For the physics package systems in the nuclear stockpile, such tests are prohibited by test ban treaties and environmental concerns. Maintenance is complicated by the fact that many replacement parts and processes are no longer available, some materials can no longer be used due to regulations, and production plants are no longer operational. This lack of test data makes it essential to use all available information: historical records, surveillance information, expert knowledge, physical models, and test data from similar systems before test bans.

Delphi Automotive Systems (formerly General Motor's AC/Delco Parts and system's businesses) had a different dynamic environment and a newer statistical problem. When a new system is proposed, it is targeted for a certain model year that locks it into a tight schedule for development and production. Relying on statistical and engineering reliability methods, the Delphi engineers would build prototypes and run tests to gather data for reliability growth analysis. When test results indicated problems, it was possible that insufficient time for corrective action was available before the scheduled start of production, requiring delays in the schedule. In the extreme case, this could result in one of the industry's worst nightmares: product recalls. Delphi approached Los Alamos with the simple question "Can't we do better in estimating reliability before it's too late and avoid surprises?" "And how can we do this without test data?" At the same time, the weapons program at Los Alamos was asking the same questions.

The development of the PREDICT methodology was the result of beginning to answer those questions. With the parallel analyses in both applications, lessons and methods developed in one application cross-fertilized in another. The Delphi fuel systems lacked the test data during their early stages of development, when they were just engineering concepts. Eventually, tests were performed and customer use data became available so that the automotive experts could gauge the estimated performance during the early development stages. The weapons program will not have the luxury of obtaining such data. Successful predictions at Delphi provided corroboration for the methods applied to the weapons program. Conversely, the early developments and use of knowledge systems for the weapons program resulted in Delphi's interest in developing a similar approach.

We have tracked the development of five Delphi Automotive concept systems to date, and we are scheduled to analyze more. During those system studies, test data (for certain failure mechanisms, components, subsystems, and systems) became available to determine the accuracy of the experts' judgments. The positive results have been welcomed by the weapons program because they corroborate the PREDICT methodology. Such test data will not be available for the weapons program, so they view Delphi's successes as their own.

Delphi Automotive Systems has seen the value of *what-if* analysis as a planning tool, and has utilized it to avert "surprises" (e.g. problems found during production, or worst case, product recalls) and to effectively plan resources for test programs. In summary, Delphi has converted the way they perform reliability to this global view of combining all information and tracking changes. Their community of practice has changed, and the reactions of the experts involved has been that "PREDICT can help me do my job."

The successes PREDICT has had to date in both applications can be summarized by testimonies such as:

> "*PREDICT is an important tool that we will use to ensure the continued safety and reliability of our nuclear deterrent*." (Joseph Martz, Program Manager, Enhanced Surveillance and Materials R&D, Los Alamos National Laboratory)
>
> "*One of the most important and useful tools that I have used in new product definition and development*" (James Jeffers, Fuel Pump Program Manager, Delphi Automotive Systems)

3. Steps for Implementing PREDICT

The PREDICT methodology embodies a toolbox of multidisciplinary methods and techniques, applied using a framework of steps for the implementation and use of these methods. Figure 1 depicts these steps and illustrates the cyclical nature of tracking a complex system through its dynamic lifetime (whether in the loop from concept to in-use development, or in the loop due to aging/maintenance in the field). The steps followed to implement PREDICT are:

1) Define Requirements and Reliability/Performance Measures
   Because reliability is defined in terms of the system functioning according to specifications, those specifications must be carefully defined in terms that the community of experts understands. Performance can be measured in a number of different ways, cycles per second, output per shift, maximum stress limits, pressure ranges of operation, operational availability, reliability at 12 months, 10^{-9} probability of system failure, etc. There may not be a single definition of performance that fits all parts of the system. How to convert from one definition to another becomes important and is specified by the experts. Taking the time to carefully define terms in this step is important. And like the other steps, these definitions can change as the system changes and as new information becomes available.

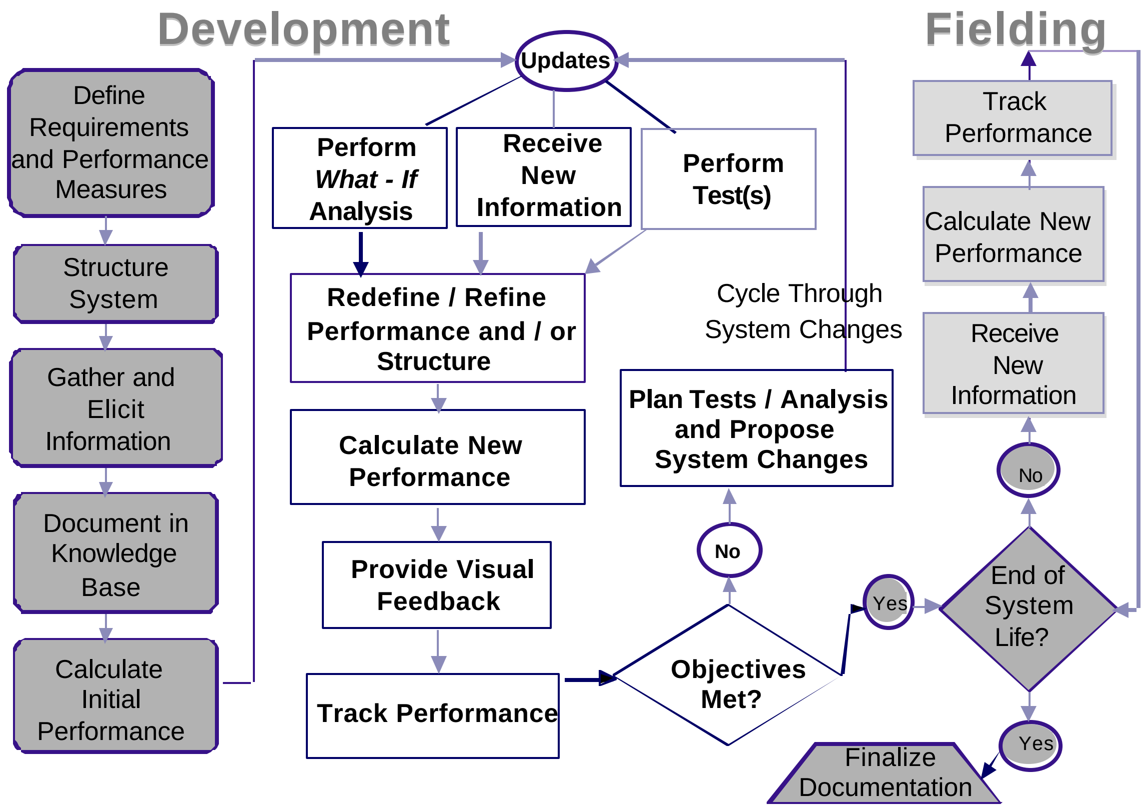

Figure 1. PREDICT implementation steps and flowchart

Traditional anthropological field techniques are used to elicit the insiders' ways of thinking about performance in their own words and to develop a framework which guides the subsequent elicitation and analysis of expert judgment on performance. This step ensures that the PREDICT process will make sense to the insiders, that the needed information is gathered in the appropriate form, and will fit their culture and community of practice. Pieces in this step draw from the fields of anthropology, sociology, expert judgment, decision analysis, and include identifying the key insiders who will provide an explanation of their culture. These key experts (or advisors) are individuals who are knowledgeable about their community of practice / culture, provide an "entree" into their culture of both other experts and management, explain its workings to analysts, provide guidance on the elicitation, and motivate wider participation by other experts.

2) <u>Structure the System—Create the Framework</u>
While no single step is the most important, this one is vital for the study to succeed. It is often neglected or over simplified, resulting in frustration by the experts and confusion over how to combine the sources of data and information. The system—all its parts, pieces, processes, activities, failure mechanisms, workings, environments, conditions, etc.—must be diagramed or

structured according to all these aspects affecting performance and in ways familiar to the community. There are various methods for establishing graphical representations of the system and its performance. The Probabilistic Risk Assessment (PRA) community has traditionally uses fault trees, event trees, failure modes and effects, and reliability block diagrams for the parts—components, subsystems, system. However, structuring the activities which can include manufacturing processes (e.g., assembly of parts), quality control / assurance activities (e.g., inspections), and physical processes (e.g., mechanics, chemistry) is not as obvious. Describing these processes using traditional tools is often cumbersome. If actual equations and models are lacking for establishing interactions and interrelations among the parts and processes, then perhaps a logic-based model is appropriate. Process trees, Bayesian networks (Jensen, 1996) and probability networks and directed graphic techniques are designed to handle complex and intricate relationships among parts and processes. This step includes formal elicitation (working with the experts to define the structure), knowledge and use of various structuring methods (e.g., logic diagrams, process trees), and formulating the interrelationships among the various parts and activities of the system. This step also includes the formulation of mathematical models and functional relationships that bind the parts, nodes and levels of the system structure together. For example, if the system is in series, a Weibull model might be chosen to calculate the reliability for each part/node and the product of those reliabilities would determine how to calculate the performance within and between levels.

3) Gather and Elicit All Sources of Data and Information
The above framework guides experts to identify sources of information that might be applicable to "populate" the various parts and processes of the structure and relating to the performance requirements. At this step uncertainties for all the sources are characterized according to the discussion in the previous section. This step also includes a formal elicitation exercise where experts provide their estimates in the absence of data and where experts provide their expertise about what sources of information are relevant to use and how they should be weighted (section 4.2).

4) Documentation (Knowledge System)
Documentation is an important step throughout the implementation of the PREDICT methodology. It begins by documenting the definitions of performance created in Step 1 and ends with the last bit of information acquired about the system's use. Elicitation methods, experts' qualifications, and how they arrived at their judgments (their sources of information, assumptions, caveats) are all recorded for traceability and later updating. One of the documentation techniques is to build an electronic repository, a knowledge system, which allows the user to readily store, access, and trace the expert judgments and the information arising from the below steps. Pieces in this step include elicitation and documentation techniques.

5) Calculate Initial Reliability (with Uncertainties Attached) from Experts

The framework in step 2 provides a formal structure and models for the system. The expert elicitation in step 3 provides performance estimates, uncertainty ranges, their reasons for these estimates, and the sources of future test/experimental data. This expert information is then combined with other sources of information (historical records, computer runs, etc.) to formulate an initial reliability (with uncertainties) of the system. If the system is a concept design, then most of the information will come from the experts. If the system is one already developed and fielded, then there is less emphasis on expertise, and it is used only when data are sparse. This initial reliability or performance estimate will be in the form of an uncertainty distribution. That distribution is documented and becomes the first snapshot in time of the existing knowledge about the system. Subsequent new information and analyses will change this estimate, beginning the tracking and updating cycle. Performance or reliability uncertainty distributions are calculated for all pieces and processes and propagated through higher system levels using Monte Carlo simulation. Pieces in this step include elicitation, statistical methods, and uncertainty analysis.

6) The Updating Cycle

The middle portion of the flowchart (figure 1) depicts a cycling set of steps that begins with the updating concept. Updating could occur for several different reasons: new information becomes available, new test data becomes available, or the experts ask *what-if* questions. After viewing the results from the initial reliability estimation of the concept product, the experts determine the ensuing courses of action based on these choices. The initial reliability results could indicate what parts or processes need improvement, what design changes might be beneficial, what tests or prototypes should be built, etc. Even before any (expensive) actions are taken (e.g. building prototypes), *what-if* cases can be calculated to predict the effects on reliability of such proposed changes or tests. Therefore, the experts may want to run several *what-if* cases before deciding on design changes, prototypes, or planning for tests.

Any new information that becomes available, such as design changes, test results, prototyping, manufacturing changes, is utilized to calculate new reliability and uncertainty estimates. Experts review the results of each calculation, using these as a basis for decisions about how to improve the reliability and reduce the uncertainty. With each subsequent change or addition of new information or new data, the reliability calculations are made again and again throughout the product's lifetime—design, prototyping, testing, production, and in-use phases. Iterations late in the system's lifetime will reflect reliability based on in-use data, coming from warranty data and customer-provided information. As part of the dynamics of the system, its performance requirements and structure may also change. Pieces in this step include elicitation, statistical methods, uncertainty analysis, and documentation techniques.

7) The Fielding Cycle

Once the requirements and objectives of the system are met, the system (and its analysis) goes into the fielding cycle. But here, new information, new use data, or other changes may still continue. To accommodate these, another cycle of update, reanalyze, document, and make decisions/plans occurs. This cycle continues through the system's lifetime until retirement.

8) Final Documentation

At the end of a system's lifetime, the implementation of the PREDICT methodology includes a complete, well-documented record—a knowledge system (section 4.3)—of the lifetime development and performance of this product. This can be used by others in the future, provide a learning tool, and contribute to corporate memory for the next new system.

4. Expert Judgment

4.1 Expert Judgment as Data

The formal use of expert judgment is at the heart of the PREDICT methodology and appears in many of its steps. For years, methods have been researched on how to structure elicitations so that analysis of this information can be performed statistically (Meyer and Booker, 1991). Expertise gathered in an *ad hoc* manner is not recommended for these purposes.

Expert judgments are the expressions of informed opinion, based on knowledge and experience, that experts make in responding to technical problems (Ortiz, et. al, 1991). Experts are individuals who have background in the subject area and are recognized, such as by their peers, as qualified to address the technical problems. Expert judgment is used in all technical fields—medicine, economics, engineering, risk/safety assessment, knowledge acquisition, decision sciences, pharmaceuticals, environmental studies, to name a few.

Because expert judgment is often used implicitly, it is not always acknowledged as expert judgment. It can also be obtained explicitly through the use of formal elicitation, the focus here.

Examples of expert judgment include:

- the probability of an occurrence of an event,
- a prediction of the performance of some product or process,
- the decision about what statistical methods to use and what variables enter into a statistical analysis,
- the decision about which data sets are relevant for use,
- the assumptions used in selecting a model,
- the decision concerning which probability distributions are appropriate to use,
- a description of experts' thinking and information sources in arriving at any of the above responses.

Expert judgment can be expressed in quantitative form—probabilities, ratings, odds, uncertainty estimates, weighting factors, and physical quantities of interest (e.g., costs, time, length, weight, etc.)—or in qualitative form—a textual description of the expert's assumptions in reaching an estimate, reasons for selecting or eliminating certain data or information from analysis, and natural language statements of physical quantities of interest (e.g., "the system performs well under these conditions.")

Quantitative expert judgment can be considered to be "data". And qualitative expert judgment can be quantified and then also be considered as data. Like "hard" data from test, experiments or physical observations, expert judgment must be handled according to the same kinds of considerations:

- Expert judgment is affected by how it is gathered. Elicitation methods take advantage of the body of knowledge on human cognition and motivation and include procedures for aiding memory and countering effects arising from the phrasing of the questions, response modes, the influence of the elicitor, and the expert's personal agenda (Meyer and Booker, 1991).
- Just as planning ahead for what to gather is important in experimental design, such planning is important for expert judgment.
- Expert judgment has uncertainty, which can be characterized and subsequently analyzed. Many experts are accustomed to giving uncertainty estimates in the form of simple ranges of values. In eliciting uncertainties, the analysts should be aware of experts natural tendency to underestimate uncertainty.
- Expert judgment can be conditioned on various factors. These factors include: the phrasing of the question (Payne, 1951), the information the experts considered, the experts' methods of solving the problem (Booker and Meyer, 1988), and the experts' assumptions (Ascher, 1978). A formal structured approach to elicitation gives analysts a better handle on conditioning effects.
- Expert judgment can be combined with other data. For example, in Bayesian updating analysis, an expert's estimate can be used as a prior distribution for an initial reliability. When test data become available, for the role of the likelihood, the expert's reliability estimates may be updated, using Bayesian methods (Kerscher et, al, 1998).

4.2 Formal Elicitation Phases and Steps

The formal steps for structuring and designing a formal elicitation are briefly outlined below. The details and techniques are available in Meyer and Booker (1991).

Phase 1: Determine whether expert judgment can be feasibly elicited. Questions that must be addressed include, "does the problem involve rapid response?", "can the potential experts 'think aloud'?", and "has there been prior use of expert judgment?".

Phase 2: Determine whether expert judgment can be better elicited in a probabilistic or alternative (e.g., fuzzy) framework. The answer depends on whether experts think in terms of (subjective) probability or not, what kinds of vagueness are involved, and how qualitative the information is.

Phase 3: Design the elicitation. This phase involves several detailed steps:

Step 1: Identify the advisor expert(s) who can provide reasons, goals, or motivations for championing the work. These individuals can be utilized to obtain and ensure the continued participation and good will of the insiders.

Step 2: Construct representations of the way that experts measure and forecast the performance/reliability of the system. This is begun by asking advisor experts how the community represents and thinks about the system. For example, experts may think in terms of a reliability block diagram.

Step 3: Draft the questions. Ask advisor experts to identify the phenomena (variables) of interest, how these are assessed, and what metrics or natural language terms are used.

Step 4: Plan the interview situation. Advisor expert(s) are asked what settings would be the best, groups/teams or individual interviews. Is it preferable to analytically aggregate multiple expert estimates or reach a consensus? Should estimates be anonymous?

Step 5: Select the experts. A selection strategy is developed with the expert advisor(s) considering the community of practice, experts' affiliations and publications, the diversity among the experts, and their availability.

Step 6: Motivate Experts' Participation. Ask advisor expert(s) for inhibitors and motivators to participation, and then mitigate and enhance these. Ask how the official request for experts' participation should be delivered (e.g., by whom, means, timing, and order of information). Identify factors that will help the experts do their jobs.

Step 7: Pilot test the questions and the interview setting. Pilot tests are conducted on advisor expert(s) and selected experts to test the "think aloud" protocol, and provide a last check on the elicitation design (i.e., question phrasing).

Phase 4: Perform the elicitation and document the results. Experts' estimates and their uncertainties may require some translation into uncertainty distributions, a common performance metric, or quantification. Whatever is done with the experts' judgments is fed back to them for review to minimize the chance of misrepresenting their knowledge.

4.3 Knowledge Systems

Knowledge is defined as what qualified individuals know with respect to their technical practices (e.g., problem solving). For example, it addresses questions such as, "how do you do x under circumstances y?" and "what is it you know?" It refers to the context in which information is used and, therefore, to the community of practice.

Knowledge systems were briefly introduced in step 4 of the implementation of the PREDICT methodology. They are a web-based electronic repository customized to the technical communities that brings together their data and knowledge. The repository is constructed in quantitative form to provide the methods and tools that the experts need to solve problems and make decisions. Constructing a knowledge system relies heavily on formal expert elicitation to structure the system and to "populate" it. The process of constructing this repository also provides a valuable learning opportunity by breaking down the complex system into manageable parts. Other advantages for using a knowledge system include: the stored knowledge is available at *customized* levels of detail for different users such as new project personnel, managers, and decisionmakers; updates and decisions are traceable (i.e., understanding of why we did this when we did it and what we knew back when); and the knowledge is available for the next system to be studied.

Because most technical professionals today are accustomed to using the web, HTML GUIs are convenient foundations for knowledge systems. Other options include commercially available languages such as IDL® (Interactive Data Language) and software such as IBM Lotus Notes®.

5. Statistical Issues and Analysis

5.1 Uncertainty, Fuzzy Logic, and Probability

A major portion of the statistical analyses used in PREDICT focuses on characterizing, combining, and propagating uncertainties through the system structure by using distribution functions of one type or another. Uncertainties enter into the system study in a number of different ways.

- There are uncertainties involved in determining weighting factors for combining experts and for combining other sources of data/information. It is recommended (Meyer and Booker, 1991) that equal weights be used if there is no additional information to indicate otherwise. However, sensitivity studies should be made to determine the impact of that maximum entropy solution.
- All the sources of data and information have uncertainties. When estimates are elicited from experts, uncertainty values, usually in the form of ranges, are also elicited. Physical models or simulations models have uncertainties regarding input-output relationships, in the choice of models (so-called modeling uncertainty) and in model parameters.
- As noted above, different measures and units are often involved in specifying the performance of the system. To map these into common units conversion factors are often required. These conversions can also have uncertainties and require a distribution function. For example, at Delphi Automotive, a two-parameter Weibull model is used to project the reliability forward into key time points of the systems' lifetimes: at 12 months and at 36 months (for warranty periods), and at 100,000 miles (for life considerations). This conversion from miles to time has an associated uncertainty distribution.

Probability theory provides a coherent way for determining uncertainties. There are many different interpretations or meanings of probability that are consistent with its axioms, Good (1965) provides eleven. Some examples include Relative Frequency Theory and Personalistic or Subjective Theory (including) Bayes Theorem. Because of the flexibility of interpretation permitted by the personalistic or subjective theory (Bement, et. al, 2000a), it is the one chosen for PREDICT. For example, it is possible to know something before observations are made, and to utilize that information. The subjective interpretation also allows us to handle rare and one-of-a-kind events, and interpret such quantities as a 10^{-9} failure rate.

Because reliability is a common performance metric and is defined as a probability that the system performs to specifications, probability theory is necessary. However, not all experts or their community think in terms of probability. We have found it useful to use alternatives such as fuzzy logic (Zadeh, 1965) for quantification when experts think in terms of rules such as if-then rules, and for characterizing a certain type of ambiguity uncertainty. For example, experts may have knowledge about the system expressed in statements such as "If the temperature is too hot, this component will not work very well." While that statement contains no numbers for analysis or probability distributions, it does contain valuable information and membership functions (from fuzzy control systems theory) are a convenient way to capture and quantify that information (Smith et.al., 1997, Smith et.al, 1998, La Voilette, 1995). Moving this information back into a probabilistic framework requires a bridging mechanism from these membership functions. It can be shown (Bement, et. al, 2000b) that membership functions may be interpreted as likelihoods; therefore the bridging can be accomplished using Bayes Theorem. This bridging is illustrated in figure 2 that depicts the various methods used for formulating uncertainty distributions.

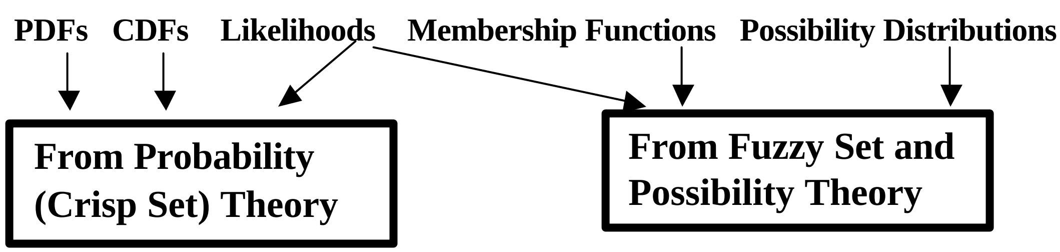

Figure 2. Theories for representing uncertainty distributions

5.2 Case Study: A Development System

As noted in the introduction, it is difficult to adequately test a newly designed system that is on a tight production schedule. This can cause a delay in production and/or result in insufficient time to correct problems. The latter contributing to faulty products possibly getting into customers' hands, which results in lack of customer confidence and customer dissatisfaction. Therefore, there is a clear need for understanding the performance of a newly designed system during its development program, even as early as the concept phase of development. Such a need can be met by estimating reliability using all available information at every lifetime phase, including when the system is an engineering concept. Gathering and combining all available information produces an estimate for the performance of the system. The following is an example of how the PREDICT methodology produces such an estimate.

5.2.1 Defining Performance and Structuring the System

Following the flowchart in figure 1, assume the performance metric is an uncertainty distribution for the reliability of a system. This metric is defined at various specified time periods, say 1 year for warranty purposes. The random variable for the reliability is $R(t)$, where t, is the time in years, and the uncertainty distribution function is $f(R;t,\theta)$, where θ is a set of parameters. For simplicity consider three specific sources of information for estimating $R(t)$ and $f(R;t,\theta)$: expert judgment, test data, and data arising from similar systems.

The next step is to structure the system. Consider a simple in-series system consisting of four levels as illustrated in figure 3:

- System level
- Design or Process level
- Subsystem (combination of processes or components)
- Individual components or processes

In reality, failure modes and mechanisms are identified below the individual component level, but these extensions will not be considered here. Reliability estimates for the higher levels may come from two sources: information from that level itself and also from the integrated estimate arising from the level below . The structure can be modified to accommodate this combination as shown in figure 4.

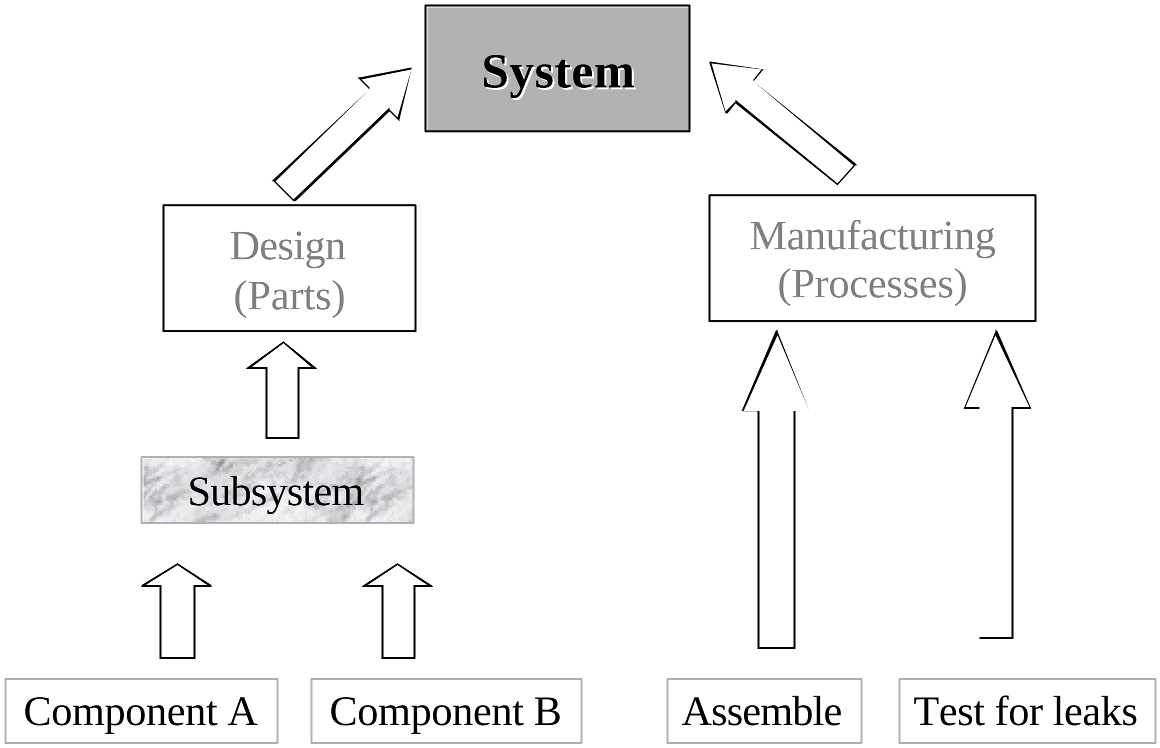

Figure 3. The system structure

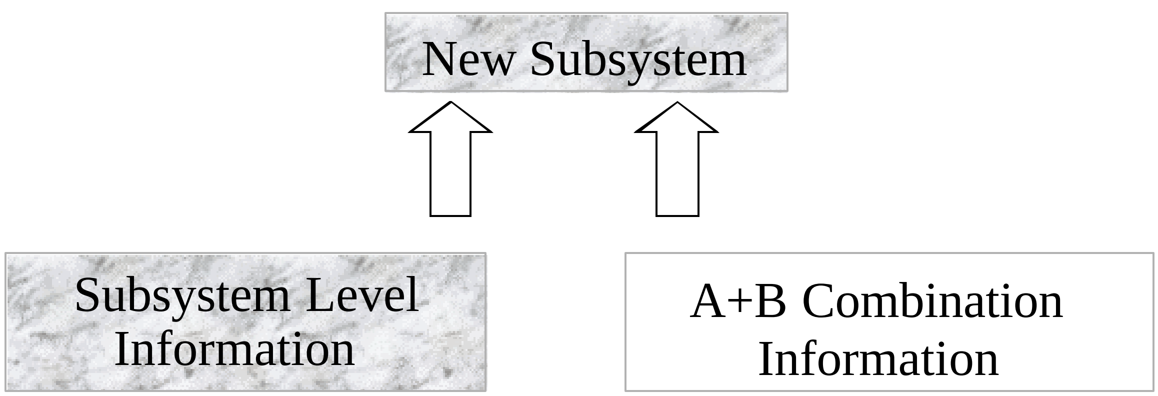

Figure 4. Higher Level Reliability Combinations

The reliability for each level of this in-series system is defined as the product of the reliabilities within that level and the system level reliability is the product of all the reliabilities of the parts (R_d) and processes (R_p):

$$R(t, \theta) = \prod_{j=1}^{n_d} R_d(t, \theta_j) \cdot \prod_{k=1}^{n_p} R_p(t, \theta_k^*)$$

for n_d parts and n_p processes, where $R(t, \theta_j)$ and $R(t, \theta_k^*)$ are a specific reliability model chosen by the experts, such as the two-parameter Weibull reliability function:

$$R_d(t, \lambda_j, \beta_j) = \exp(-(\lambda_j t)^{\beta j}) \text{ and}$$
$$R_p(t, \lambda_k, \beta_k) = \exp(-(\lambda^*_k t)^{\beta^* k}).$$

The reliability model must be physically appropriate and mathematically correct for the system. Of equal importance, the model and its usage must be culturally acceptable to the organization using it. The Weibull fits the infant mortality and useful lifetime (Kerscher, 1989) aspects of the system, provides a time dependent function, and, in this case, suits the implicit understanding of the design and manufacturing (processes) communities through their awareness of the corresponding hazard curve's "bathtub" shape. It should be noted that estimates are required for both parameters, for β (the slope) and for λ (the failure rate) for each component and process.

For a concept system, test data from prototypes or actual parts will be absent. Information sources at this point in the system's development reside mainly within the collective knowledge of the experts. Other information sources might include data from previous studies, similar parts, processes, and perhaps some physical model or simulation code outputs.

5.2.2 Analysis of Expert Judgment

A formal elicitation is necessary (following section 4.2) to understand what expertise exists and how it can be related to the reliability estimation, i.e., how to estimate the Weibull parameters. For this example, it is assumed that the experts are accustomed to working in teams, and reaching a team consensus is their usual way of working. It is not uncommon to learn from the elicitation preparation steps that not all teams think about performance using the same terms. Performance could be defined in terms of *incidences per thousand vehicles* (IPTV) which convert to failure rates for the product design, but in terms of *parts per million* (PPM) failures manufactured which translate to reliabilities for processes. Best estimates of IPTV and PPM quantities are elicited from the experts along with ranges of values. In this case, these three estimates are interpreted as the most likely (i.e., the median), maximum (worst), and minimum (best) estimates.

The job of the statistician is to work with the experts to convert these estimates to the parameters of the Weibull for both the design and manufacturing (or process) sides. The top portion of figure 5 illustrates how these estimates fit into the reliability calculations on both sides.

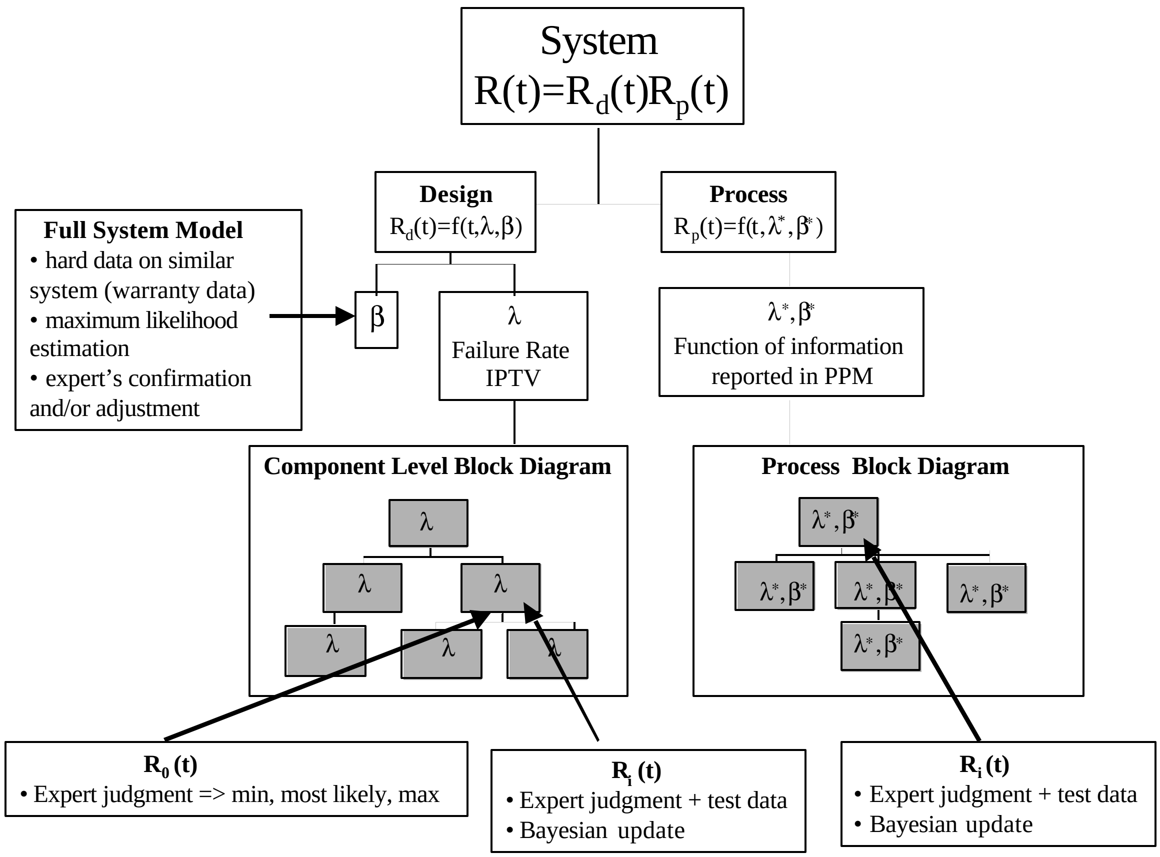

Figure 5. Dynamic system structure and model

Because IPTV at 1 year is a failure rate, a distribution for λ can be determined. Failure rates are often asymmetric distributions such as the lognormal or gamma. Because of the positive values, variety of possible shapes, and occasional interpretability of the parameters (the first parameter corresponds to a pseudo number of failures and the second parameter to pseudo total time on test), the experts chose the gamma. The best and worst cases were defined to represent the maximum and minimum possible values. However, accounting for the well-documented tendency of experts to underestimate uncertainty (Meyer and Booker, 1991), these values were equated with small tail percentiles. Sensitivity studies are recommended to demonstrate to the experts the effects of such a decision, ensuring that their initial estimates are not misrepresented.

Another difficulty arises when fitting three expert estimates to a two-parameter distribution. One of the three estimates will not match, and the experts may insist that the distribution exactly fit through all three estimates. A two-piece distribution (not a mixture of distributions), joined at one of the expert estimates can accommodate this request. Figure 6 illustrates the result of this implementation using a gamma.

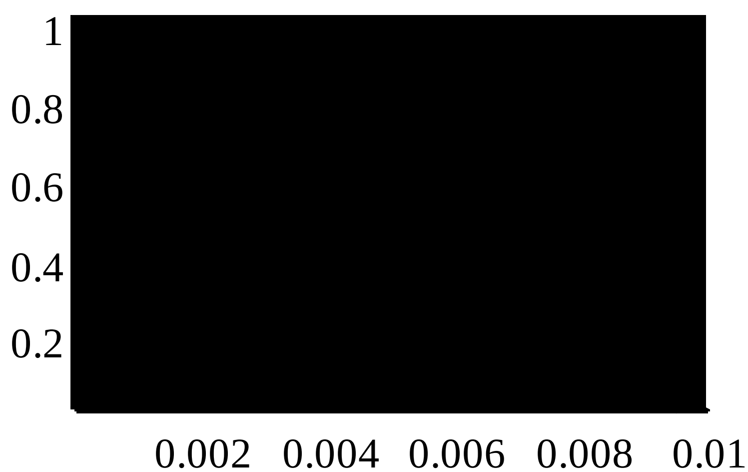

Figure 6. CDF of gamma formed from experts' estimates of {.05, 1.0, 15.0} IPTV

The experts agreed that the β parameter for the components and subsystems of the new system should correspond to that of previous, similar systems, for which warranty was available. Maximum likelihood estimates for β from Weibull fits of this warranty data provides a starting estimate that the experts were free to adjust or confirm for the new system. Warranty data is usually only available at the system or certain subsystem levels, making it necessary for the experts to make the final decisions about β values for all parts and processes at lower levels.

As part of the elicitation, experts were also asked to specify all known or potential failure mechanisms, or failure modes, for each part and process. Failure modes are failures in the components themselves, such as a valve wearing out, mistakes being made during the manufacture of components, or improper assembly of multiple components into a subsystem. For updating and documenting purposes, the percent or proportion contribution of each failure mode was also specified by the experts.

Processes are compilations of complex steps and issues, which must be considered to convert the experts' PPM estimates to Weibull parameters. Some of these issues relate to how quality control and inspections integrate with the process. For example, the reliability of the process depends upon the percent or proportion of items that slip through the quality control procedures (called *spills*). Quantities such as frequency and duration of these spills affect reliability, and these are elicited along with the functions required to specify their relationships to the PPM values provided for the processes themselves. Other issues are involved with failure modes. Through a series of transformations designed to account for these issues, the PPM estimates from the experts were converted to Weibull parameters λ^* and β^* for each process as depicted in figure 5. As on the design side, experts' estimates of best, most likely, and worst case values were used to fit an uncertainty distribution for the process reliabilities. The experts chose to use a beta distribution for the reliabilities translated from their three PPM estimates. The reasons include the beta's appropriate (0 to 1) range, its wide variety of possible shapes, and its occasional interpretability of parameters (the first parameter as pseudo number of failures and the second parameter as pseudo number of trials).

5.2.3. Initial reliability calculation

Once the parameters and uncertainty distributions were specified for the design parts and manufacturing processes, the initial reliability, $R_0(t, \lambda, \beta)$ was calculated, using Monte Carlo simulation. Because this model is time dependent, predictions at specified times are possible. Most of the data and expert estimates are given in terms of 1 year. For applications such as automobiles, three years is important for warranty reasons, and 100,000 miles is also important as a lifetime indicator. The change from time in years to time in mileage is one example of the need for a conversion factor. Such factors usually have uncertainties attached, so the conversion also requires an uncertainty distribution. This distribution was fit using maximum likelihood techniques applied to historical times-to-mileage data. This uncertainty distribution becomes part of the Monte Carlo simulation. The initial reliability calculation is concluded with system, subsystem, component, and process distributions calculated at these various time periods. Figure 7 shows the reliability for the total processes, R_p (t, λ^*, β^*), at t=1 year, 3 years and 100,000 miles.

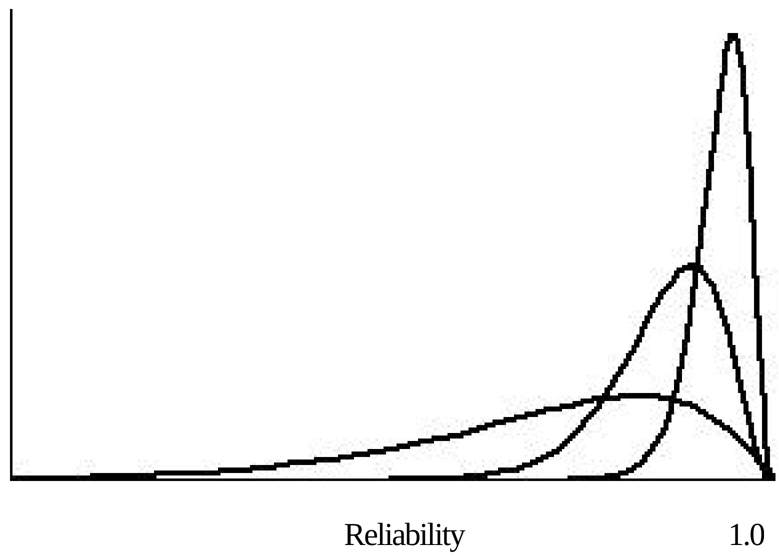

Figure 7. System Reliability Uncertainty distributions for 1 year (right), 3 years (middle), and 100,000 miles (left).

5.2.4 Tracking and updating the dynamic system

The initial reliabilities are for the conceptual system and may be quite poor with large uncertainties. Upon review, the experts can decide which parts or processes to change, where to plan for tests, what prototypes to build, what vendors to use, or ask *what-if* questions in order to improve reliability and reduce uncertainty. Before any usually expensive actions are taken (e.g. building prototypes), *what-if* cases are calculated to predict the effects on estimated reliability of such proposed changes or tests. These cases can involve changes in the structure, structural model, experts' estimates, and the terms of the reliability model as well as effects of proposed test data results. Further breakdown of components into the failure modes may be required to properly map these changes and proposed test data into the reliability model.

Because the system is under development or undergoing change, new information becomes available at various stages of its lifetime. Examples include design changes such as adding, replacing, and deleting parts and processes, supplier changes, prototype test data, production data, new engineering judgment, etc. Incorporating these changes and new information into the existing reliability estimates is referred to as the updating process.

New information and data from different sources or of different types (e.g. tests, process capability studies, engineering judgment) are analytically merged by combining uncertainty distribution functions of the old and new sources. This merging usually takes the form of a weighting scheme:

$$w_1 \cdot f_1 + w_2 \cdot f_2$$

where w_i are weights and f_i are functions of parameters, random variables, models, probability distributions, uncertainty distributions or reliabilities, etc. Experts often provide the weights, and sensitivity analyses are performed to demonstrate the effects of their choices. The $R_i(t, \lambda, \beta)$ boxes in figure 5 illustrate the general updating process.

Alternatively, Bayes Theorem can be used as a particular weighting scheme, providing weights for the prior and the likelihood through application of the theorem. Bayesian combination is often referred to as Bayesian updating. If the prior and likelihood distributions overlap (reinforce each other), then Bayesian combination will produce a posterior whose variance is smaller than if the two were combined via other methods, such as a linear combination of random variables or a mixture. This is one advantage of using Bayes Theorem.

Because test data at the early stages of system development are lacking, the initial reliability, $R_0(t, \lambda, \beta)$, is developed from expert judgment and forms the prior distribution for the system (figure 2). As the system develops, data and information may become available for only certain parts or processes (e.g. system, subsystem, component) and this would be used to form likelihood distributions for Bayesian updating. All of the distribution information in the items at the various levels must be combined upward through the system levels, to produce a final estimate of the reliability and its uncertainty at various levels along the way, until reaching the top, or system, level. Three different combination methods are used to form the next (updated) reliability, $R_1(t, \lambda, \beta)$:

- For each prior distribution that must be combined with a data-based or likelihood distribution, Bayes Theorem is used and a posterior distribution results.

- Posterior distributions within a given level are combined according to the structural model (e.g., multiplication of reliabilities for parts / processes in series) to form the prior distribution of the next higher level (figure 2).
- Prior distributions at a given level are combined within the same part / process to form the combined prior (for that item) which is then merged with the data (for that part or process). This approach is continued up the levels until a system level posterior distribution is developed.

As more data and information become available and are incorporated into the reliability calculation through Bayesian updating, they will tend to dominate the effects of the experts' estimates developed through expert judgment. In other words, $R_i(t, \lambda, \beta)$ formulated from many test results will look less and less like $R_0(t, \lambda, \beta)$ derived from expert estimates.

For general updating, test data and other new information can be added to the existing reliability calculation at any level and / or for any part or process. This data / information may be applicable to only to a single failure mode. When new data or information becomes available at a higher level (e.g., subsystem) for a reliability calculation at step i, it is necessary to back propagate the effects of this new information to the lower levels (e.g., component). The reason is that because at some future step, i+j, updating may be required at that lower level and its effect propagated up the structure. The statistical issues involved with this back propagation are difficult (Martz and Almond, 1997). It is also possible to back propagate by apportioning either the reliability or its parameters to the lower level according to their contributions at the higher level. While it can be shown that for well-behaved functions, solutions are possible, they may not be unique. Therefore, constraints may be placed on the types of solutions desired by the experts. For example, requiring that regardless of the apportioning mechanism used to propagate downward, the forward propagating maintains the original results at the higher level.

General updating is an extremely useful decision tool for asking *what-if* questions and for planning resources, such as tests, to determine if the reliability requirements can be met before actually beginning production. For example, the reliability uncertainty distributions calculated using simulation are empirical with no particular distribution form, but due to their asymmetric nature and because their range is from 0 to 1.0, they often appear to fit well to beta distributions. Suppose a beta distribution of the form:

$$\text{Beta}(x, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)} x^{(a-1)} (1-x)^{(b-1)}, \quad 0=x=1,\ a>0,\ b>0$$

is fit to a component reliability uncertainty distribution at some stage, $R_i(t, \lambda, \beta)$, resulting in parameters a = 81.9 and b= 1.01. The experts want to determine what would be gained by building 40 prototypes, testing them, and assuming all passed. Taking advantage of the beta as a conjugate prior for the binomial data, the new component reliability distribution, for $R_{i+1}(t, \lambda, \beta)$ would be a beta with parameters a = 81.0 + 40 = 121.9 and b = 1.01 + 0 = 1.01. The median improves slightly (from, 0.991 to 0.994) but, more importantly to the experts, the 5th percentile improves from 0.96 to 0.98, providing an incentive to invest in the prototypes.

The general updating cycle continues through the lifetime of the system as indicated in figure 1. Figure 8 depicts the tracking of the reliability through the system development indicating three percentiles (5th, median, and 95th) of the reliability uncertainty distribution at various points in time. The individual data points begin with the initial reliability characterization $R_0(t, \lambda, \beta)$ for the system and continue with the events associated with the general updates, $R_i(t, \lambda, \beta)$, the *what-if* cases, and incorporation of test results (depicted on the figure with vertical lines). As previously noted, asking *what-if* questions and calculating the effects on reliability of those provided valuable information for designing and modifying prototype building and test planning, before costly decisions were made.

Graphs like figure 8 were constructed at all the levels of system to monitor the effects of updating for individual parts and processes. Graphs were made for these levels at the desired prediction time values (i.e., 1 year, 3 years and 100,000 miles) to determine if reliability requirements were met at those important time points in the life of the system.

Plots like figure 8 capture the results of the experts' efforts to improve reliability and reduce uncertainty. The power of the approach is that the roadmap developed leads to higher reliability and reduced uncertainty, and the ability to characterize all of the efforts made to achieve these improvements.

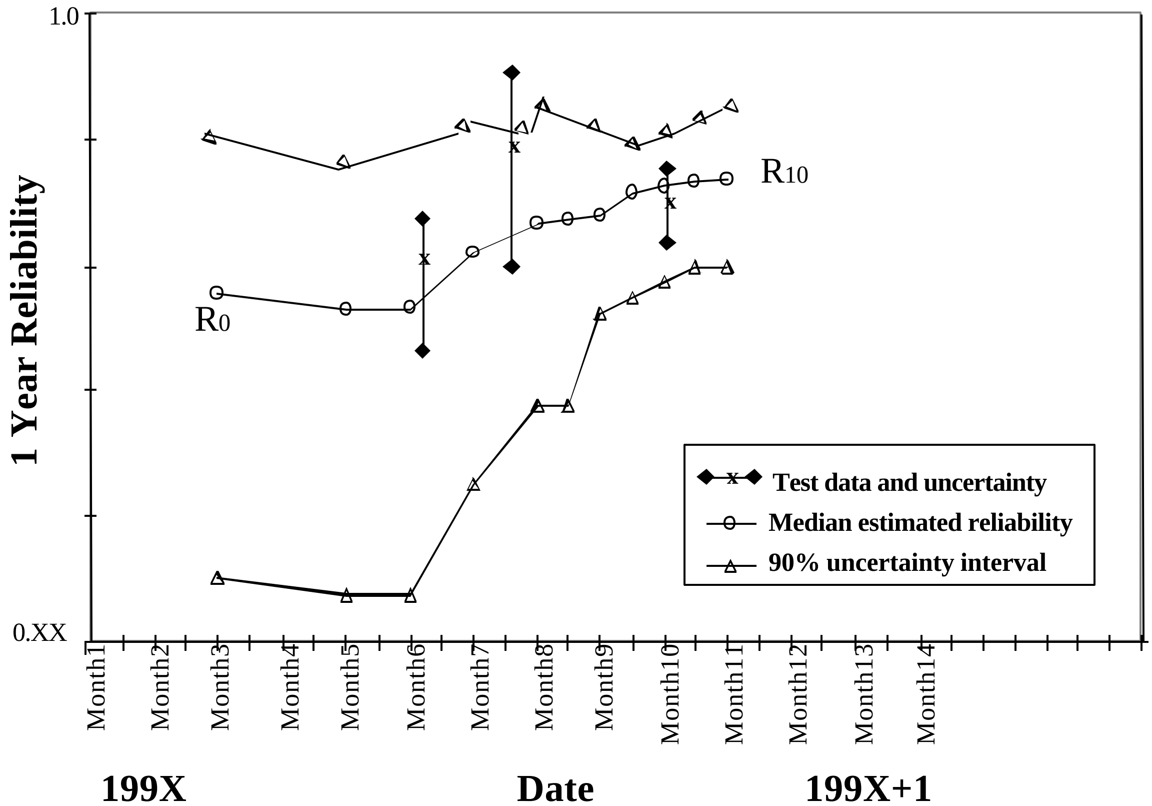

Figure 8. Tracking the system level reliability uncertainty distribution at 1 year.

6. Conclusions

While this application of Information Integration Technology has a proven track record of successful use, the PREDICT methodology does not claim to have solutions for all technical problems. There are many opportunities for research to expand the general Information Integration Technology base. A partial list of these opportunities is given below.

- The general areas of predictability and uncertainty analysis are not completely founded. More specifically in the uncertainty arena are challenges in understanding, specifying, quantifying, and propagating uncertainties. Are there, better methods of characterizing uncertainties than uncertainty distributions?
- Quantification of qualitative information has been a continuing research topic in expert judgment work. Our methods reflect those advances by quantifying rules

using fuzzy system control methods. Other quantification methods of qualitative information are needed. Do other disciplines hold the key for new methods?

- The PRA and decision analysis communities are branching into new methods of structuring the system with advances such as directed graphs, causal diagrams, networks, and process trees. Because the structure is so important for gathering and combining information sources relevant to the performance requirements, new methods are needed to accommodate ill-defined processes rather than just constructing a system as a sum of parts.
- More methods are needed to handle the back propagation problem, especially when dealing with empirical distributions and more complex structures.
- Dependencies between the various sources of information are another topic for more research. For decades, this issue has been tossed about in the literature regarding dependencies among experts, without substantial resolution. How should these be determined? How do they affect the process of combining the different sources?
- Methods of combining / integrating the various sources of information/data have relied on the traditional methods of combining distributions. Can other fields, like fuzzy logic, offer other solutions? Can metrics be developed to determine which method works best for which type of information integration problem? We have been investigating areas of information theory with entropy-based measures like Jeffreys' J, (1998), quantiles, or relative distributions (Handcock and Morris, 1999) for such purposes.
- Finally, research in the knowledge capture and representation fields is ongoing. These areas include self-elicitation and eliciting and analyzing tacit (implicit) knowledge. Although these involve different disciplines, the research has direct implications in methodologies, such as PREDICT, that deal with information integration.

This Information Integration Technology, PREDICT, has demonstrated its effectiveness for expertise capture, reliability, and performance estimation in the nuclear weapons program and for concept system development in the automotive industry. In the post cold war era, the basic philosophy of information integration is positively impacting the certification process of our nuclear systems. This same philosophy is providing the formal structure for taking advantage of a company's greatest asset—the knowledge and expertise of its engineers and designers. Our automotive and weapons customers agree that the greatest strength of Information Integration Technologies such as PREDICT is their ability to customize specific user needs, making them valuable methodologies for all design or engineering communities. This is because users establish a core of expertise that perpetuates through the Information Integration Technology resulting in a permanent shift in the way they currently think about reliability. We believe Information Integration Technologies such as PREDICT will revolutionize the way products are developed and analyzed.

Acknowledgements

We wish to thank the co-developers of PREDICT—the experts and multi-disciplinary team members at Los Alamos—Masha Keady, Yvonne Martinez, Dale Talbott, Stephen Becker, Fred Mortenson, Richard Yactor, Ronald Smith, Lyle Edwards, and William Parkinson. We also wish to thank Carl Miller of Delphi E (Energy and Chassis Systems), Delphi Automotive Systems, Luis Salazar of the Los Alamos Stockpile Systems, Joe Martz of the Los Alamos Enhanced Surveillance Program, and Kenneth Freese of the Los Alamos Technology Transfer Program. PREDICT would not have been possible without their help and support. We owe a great debt of gratitude to two statisticians at Los Alamos, Mark S. McNulty and Sallie Keller-McNulty (Group Leader of the Statistical Sciences Group), for providing us with excellent reviews and comments.

References

AMSTAT News, Issue 267, August/September, 1999.

Ascher, W., *Forecasting: An Appraisal for Policymakers and Planners*, John Hopkins University Press, Baltimore, Maryland, 1978.

Bement, T.R., J. M. Booker, K. F. Sellers and N. D. Singpurwalla, "Membership Functions and Probability Measures of Fuzzy Sets," Los Alamos National Laboratory report LA-UR-00-3660, 2000a.

Bement, T.R., J. M. Booker, S.Keller-McNulty, and N. D. Singpurwalla, "Testing the Untestable: Reliability in the 21st Century," Los Alamos National Laboratory report LA-UR-00-1766, submitted to *Science*, 2000b.

Booker, J.M. and Meyer, M.A., Sources and Effects of Interexpert Correlation: An Empirical Study, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 1, 135-142, 1988.

Booker, J.M. and M.A. Meyer, "Elicitation and Analysis of Expert Judgment" *Encyclopedia of Statistical Sciences* (Samuel Kotz, editor), 2000.

Booker, J.M., R.E. Smith, T.R. Bement, W.J. Parkinson, M.A. Meyer, "Example of Using Fuzzy Control System Methods in Statistics," LA-UR-99-1712, *SPES Newsletter*, (Section on Physical and Engineering Sciences of ASA), vol. 5, p11-12, Spring 1999.

Dempster, A. P., "A Generalization of Bayesian Inference (with discussion), " *Journal of the Royal Statistical Society, Series B,*30, 205-247, 1968.

Dubois, D. and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.

Good, I. J., The Estimation of Probabilities, MIT Press, Cambridge, MA, 1965.

Handcock, M.S. and M. Morris, *Relative Distribution Methods in Social Sciences*, Springer, 1999.

Jeffrey, R. C., *The Logic of Decision*, Second Edition, University of Chicago Press, Chicago, 1983.

Jeffreys, H., *Theory of Probability*, Third Edition, Oxford Classic 5 Series, Clarendon Press, Oxford, 1998.

Jensen, F.V., *An Introduction to Bayesian Networks*, Springer, New York, 1996.

Kerscher III, W.J., “A reliability model for total field incidents,” *Proceedings of the Annual Reliability & Maintainability Symposium*, 1989 Jan, pp 22-28.

Kerscher, W., Booker, J., Bement, T. and Meyer, M., “Characterizing Reliability in a Product/Process Design-Assurance Program,” *Proceedings of the International Symposium on Product Quality and Integrity*, Anaheim, California, and Los Alamos Laboratory report, LA-UR-97-36, 1998.

Laviolette M, Seaman J Jr, Barrett J, and Woodall W. “A probabilistic and statistical view of fuzzy methods, ” *Technometrics*, vol. 37 249-281, 1995.

Mann, N.R., R.E. Schafer and N.D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data* John Wiley & Sons, New York, 1974.

Martz, H.F. and Almond, R.G. “Using Higher-level Failure Data in Fault Tree Quantification”, *Reliability Engineering and System Safety*, vol. 56, no. 1, 29-42, 1997.

Meeker, W. Q. and L.A. Escobar’s new book, Statistical Methods for Reliability Data, John Wiley and Sons, New York, 1998.

Meyer, M.A. and Booker, J.M., *Eliciting and Analyzing Expert Judgement: A Practical Guide*, Academic Press, London, 1991, being reissued by ASA-SIAM, Philadelphi, PA, February, 2000.

Meyer, M.A., Booker, J.M., Bement, T.R., Kerscher III, W.J., “PREDICT—A New Approach to Product Development,” 1999 R&D 100 Award winner, *R&D Magazine*, Vol. 41, p 161, September, 1999. (Also, Los Alamos National Laboratory document, LALP-99-184, August, 1999.)

Ortiz, N.R, Wheeler, T.A., Breeding, R.J., Hora, S., Meyer, M.A., and Keeney, R.L., Use of Expert Judgment in NUREG-1150, *Nuclear Engineering and Design*, 126, 313-331, 1991.

Payne, S. *The Art of Asking Questions*, Princeton University Press, Princeton, New Jersey, 1951.

Smith, C. A. B., “Consistency in Statistical Inference and Decision (with discussion), ” *Journal of the Royal Statistical Society*, *B*23, 1-37, 1961.

Smith R, Bement T, Parkinson W, Mortensen F, Becker S and Meyer M., “The Use of fuzzy control systems techniques to develop uncertainty distributions,” *Proceedings of the Joint Statistical Meeting*s, Anaheim, CA, 1997.

Smith, R.E., J.M. Booker, T. R. Bement, M.A. Meyer, W.J. Parkinson, M. Jamshidi, “The Use of Fuzzy Control System Methods for Characterizing Expert Judgment Uncertainty Distributions,” *Proceedings PSAM 4 International Conference*, September, 497-502, 1998.

Zadeh, L.A., “Fuzzy Sets,” *Information and Control*, vol. 8, 338-53,1965.