Pacific Northwest National Laboratory

Operated by Battelle for the
U.S. Department of Energy

Uncertainty and Sampling Issues in Tank Characterization

A.M. Liebetrau
B.A. Pulsipher
D.M. Kashporenko

G. Chen
F. Gao
S.A. Hartley

RECEIVED
AUG 05 1997
OSTI

June 1997

MASTER

Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC06-76RLO 1830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information, P.O. Box 62, Oak Ridge, TN 37831;
prices available from (615) 576-8401.

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161

The document was printed on recycled paper.

Uncertainty and Sampling Issues in Tank Characterization

A.M. Liebetrau
B.A. Pulsipher
D.M. Kashporenko

G. Chen
F. Gao
S.A. Hartley

MASTER

June 1997

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

Executive Summary

A defensible characterization strategy must recognize that uncertainties are inherent in any measurement or estimate of interest and must employ statistical methods for quantifying and managing those uncertainties. Estimates of risk and therefore key decisions must incorporate knowledge about uncertainty. This report focuses statistical methods that should be employed to ensure confident decision making and appropriate management of uncertainty.

Sampling is a major source of uncertainty that deserves special consideration in the tank characterization strategy. The question of whether sampling will ever provide the reliable information needed to resolve safety issues is explored. The issue of sample representativeness must be resolved before sample information is reliable. Representativeness is a relative term but can be defined in terms of bias and precision. Currently, precision can be quantified and managed through an effective sampling and statistical analysis program. Quantifying bias is more difficult and is not being addressed under the current sampling strategies. Bias could be bounded by 1) employing new sampling methods that can obtain samples from other areas in the tanks, 2) putting in new risers on some worst case tanks and comparing the results from existing risers with new risers, or 3) sampling tanks through risers under which no disturbance or activity has previously occurred. With some bound on bias and estimates of precision, various sampling strategies could be determined and shown to be either cost-effective or infeasible.

The number of required samples depends on several factors. If waste properties are sufficiently far from decision thresholds relative to heterogeneity and measurement uncertainties, very few samples may suffice for confident decision making. If waste properties are near decision thresholds, too many samples may be required. Tank grouping may provide a way to improve estimates and decrease uncertainty, but grouping strategy won't resolve the bias issue. The tradeoffs between sample-based decision making, monitoring and control, or intervention will require uncertainty analysis and value-of-information studies.

The evaluation of the relative worth of sample information should stem from an evaluation of the effect of uncertainty on perceived risk. In this report, uncertainty analysis methods are shown to be an effective means to evaluate the effect of uncertainty. Uncertainty analysis uses the structure of the Structured Logic Diagrams (SLDs) to estimate and propagate uncertainty to the point of calculating risk. Uncertainty analysis is used with sensitivity analyses to determine which uncertainties have the greatest effect on how well we understand risk. Uncertainty analysis provides the methods for understanding the effect of uncertainty and evaluating the relative benefit of decreasing uncertainty (by either adding more information/samples or taking an entirely different approach to resolving the problem). Specific examples of how statistical methods of uncertainty analysis are employed in conjunction with an SLD are demonstrated.

Uncertainty analysis methods are shown to be an integral part of the proposed new characterization strategy. They provide the tools for quantifying, propagating, and managing uncertainty. Sample representativeness issues must be addressed before one can determine whether sampling will provide sufficient information for confident decision making. Although the new characterization strategy requires a more rigorous treatment of uncertainty, uncertainty analysis methods coupled with risk-based value of information techniques can be performed and the issues associated with sampling and analysis can be evaluated, resolved, or eliminated.

Table of Contents

Executive Summary	iii
1.0 Introduction	1.1
2.0 Uncertainty Analysis Issues	2.1
2.1 The SLD as a Technical Basis for Tank Safety Resolution.....	2.1
2.2 Uncertainty Analysis for Tank Safety Issues.....	2.4
2.3 The SLD as a Technical Basis for Uncertainty Analysis.....	2.6
2.3.1 The SLD as a Computational Model	2.6
2.3.2 Sensitivity Analysis: Comparison of Alternatives	2.11
2.3.3 Sources of Uncertainty	2.12
2.3.4 Representation of Uncertainty	2.13
2.3.5 Estimation of Uncertainty	2.14
2.4 Uncertainty Analysis Example	2.15
2.5 Uncertainty Analysis Summary.....	2.24
3.0 Sampling Issues: Potential for Resolution of Tank Issues Via Sampling.....	3.1
3.1 Sources of Uncertainty Introduced Through Sampling and Analysis	3.2
3.1.1 Sample Representativeness	3.2
3.1.2 Measurement Uncertainty	3.4
3.2 Overcoming Sampling and Analysis Roadblocks.....	3.4
3.2.1 New/Improved Sampling Technologies.....	3.4
3.2.2 Sample Representativeness Approaches: Bounding Bias and Tank Grouping Studies.....	3.4
3.2.3 Estimating and Controlling Measurement Error.....	3.6
3.2.4 Statistical Models to Estimate and Combine Uncertainties.....	3.6
3.3 Evaluating Relative Merit of Overcoming Sampling Issues.....	3.7
3.4 Sampling Issues Summary	3.8
4.0 Large-Scale Problem	4.1
4.1 Logical Completeness and Large-Scale Problem Analysis.....	4.1
4.2 Large-Scale, Composite Problem Resolution.....	4.4
5.0 Acknowledgments	5.1
6.0 Bibliography.....	6.1
List of Tables	
2.1 Summary of Calculations Illustrating the Effect of "Perfect" Information on Spark Frequency (f_s)	2.24
List of Figures	
2.1 Components of a Technically Defensible Tank Safety Analysis Resolution Strategy	2.4

2.2 Schematic SLD Showing how Uncertainties Enter and Propagate.....	2.7
2.3 The Role of the SLD in the Complimentary Processes of Determining Information Requirements and Evaluating Performance	2.9
2.4 SLD Showing Estimation of Frequency of Dome Failure due to Flammable Gas Burn	2.18
2.5 Estimation of the Distribution of f_c by Monte Carlo Simulation.....	2.19
2.6 Estimation of the Variance of f_c by Propagation-of-Error Methods.....	2.20
2.7 The Cut Set Shown in Bold Lines Contributes to the Variance of the Top Level Event Through Bracketed Terms in the Expression for $\sigma_c f_c$	2.21
2.8 SLD for the FG Model with the Estimated Variances for the Uncertainty and Risk Distributions Shown in Figure 2.5.....	2.22
2.9 SLD Showing how the Estimated Variance for the Risk Distribution in the FG Model Changes as the Variance for the Uncertainty Distribution for One Input, the Spark Frequency, Changes from 5.5×10^{-6} to 0	2.23
4.1 Integrated Technical Acceptance Process.....	4.4
4.2 Full Extended Problem Interaction Graph for the Ferrocyanide Sub-Problem Set.....	4.6
4.3 The Extended Problem Interaction Graph for Criticality	4.8

1.0 Introduction

Issues concerning uncertainty are at the core of each safety issue. Uncertainty must be estimated and managed to acceptable levels. One of the major sources of uncertainty in the tank characterization process is sampling. This report outlines methods for explicitly quantifying, propagating, and managing uncertainty within the context of a more comprehensive tank characterization strategy.

This report contains three major sections: Uncertainty Analysis Methods, Sampling Issues, and Large-Scale Problem Issue. Tank characterization strategies should be based on risk. Risk is estimated using information that is less than exact and subject to significant uncertainty. Any characterization strategy must recognize the inherent uncertainties and employ methods that account for those uncertainties. In the second section of this report, uncertainty analysis methods are demonstrated within the context of Structured Logic Diagrams.

Any characterization strategy also involves sampling. Sampling issues are an area of major concern for tank characterization. The issues associated with the number of core samples required and the representativeness of samples are discussed in section three. Some recommendations for overcoming sampling roadblocks are discussed.

2.0 Uncertainty Analysis Issues

Issues concerning uncertainty are at the core of each safety issue, so uncertainty analysis must be an essential component of any successful strategy to resolve these issues. It is crucial, therefore, to develop a sound and flexible uncertainty analysis strategy as part of the overall tank safety issues resolution strategy. However, the uncertainty analysis must be based on a competent understanding of tank content phenomenology and a methodology for dealing with the large-scale, composite problems posed by the tank safety issue. The uncertainty analysis must also involve techniques that can respond to and bring under control the many sources of uncertainty implicit in each tank safety issue. Specifically, the uncertainty analysis must

- reflect a thorough, systematic presentation and analysis of the underlying, causative phenomenology
- use a methodology sufficiently powerful to decompose, reformulate, and analyze the large-scale, composite problem posed by the five tank safety issues
- provide the capability to quickly evaluate frequencies, probabilities, and uncertainties at any point in the composite problem structure and propagate these uncertainties throughout the structure. (The composite problem structure is represented in the proposed safety resolution strategy by a structured logic diagram (SLD). An SLD is used to develop and communicate the technical basis for each safety issue.)

A strategy that addresses these three requirements will provide a sound technical basis for resolving uncertainty issues associated with resolution of tank waste safety issues. It will also provide a broad range of options for sensitivity and screening analyses to eliminate non-essential (or insignificant) uncertainty issues. Finally, the uncertainty analysis in itself will point the way to the development of acceptable levels of risk for each safety issue.

2.1 The Structured Logic Diagram as a Technical Basis for Tank Safety Issue Resolution

Phenomenology plays a crucial role in establishing the technical basis for resolving of each safety issue. As chemists, physicists, engineers, and statisticians, we must agree on basic phenomenology--how each kind of molecular species will react under normal and off-normal conditions in the tanks and what the reaction's immediate structural outcome will be--before we can resolve a particular safety issue. If we lack adequate information or understanding, we must also agree on how to conduct further information gathering activities such as sampling and analysis. To do this efficiently and unequivocally, **we must lay out all aspects of the phenomenology, as plainly and transparently as possible, but with sufficient technical detail to fully explore and characterize all safety issues associated with each tank.**

The issue of just how this technical basis is established is crucial. Above all, there must be a sense of reliability, a sense that the technical basis has a coverage, an adequacy, and a completeness that is technically convincing and substantively focused on realistic solutions. In other words, the problem definition for each safety issue must reveal its full scope through a comprehensive exposition of all alternatives and their interrelationships. The problem definition must ensure that each sub-problem--

each data collection and analysis problem, each model evaluation problem, each potential safety failure path, each off-normal causative event chain--is identified, documented, and evaluated. Techniques of reliability engineering are essential to ensure full comprehension of the larger composite tank safety problem. In summary, sufficient reliability must be engineered into the technical basis for each safety issue to translate objectives and decisions, made with respect to that basis, into realistic and effective actions.

The SLD has been selected as the vehicle to establish and communicate the technical basis for each tank safety issue. An SLD is the dual of a fault tree in the sense that the SLD is focused on a successful outcome, whereas a fault tree is focused on failure detection. Except for this distinction, both a fault tree and a SLD are used in the same manner. The SLD is an ideal tool for dealing with tank safety issues because it facilitates a systematic examination of all conceivable issues that must be resolved to bring a particular safety issue to closure. The SLD does this in a highly visible way by showing how all these issues, sub-problems, and alternatives are related to each in the context of the composite safety problem. Moreover, the SLD clearly shows that the analysis for a particular safety issue is explicitly tied to a clearly identified outcome (estimation of risk).

The purpose of this section is to show how SLDs can serve as the technical basis for uncertainty analysis, logical completeness issues, and the application of large-scale problem resolution methods. Structured logic diagrams for several of the most important tank safety issues are presented in detail in other sections of this report.

The SLD is a powerful tool for presenting and analyzing the underlying, causative phenomenology for each safety issue in a systematic and comprehensive way that reveals interrelationships among alternatives and sub-problems that must be solved to resolve a given safety issue. The first result of defining a SLD for any safety issue is that one is immediately made aware of the fact that tank safety problem is large and complex. Numerous safety issues have been identified and each of these, in turn, contains upwards of 100 or more individual sub-problems of varying degrees of difficulty. Moreover, many of these sub-problems presently lack adequate specification for problem analysis. Methods of large-scale problem analysis are clearly required to keep track of all missing or required subproblem specifications and changes therein. Large-scale problem analysis and decomposition methods, as used here, provide a rational process by which a composite problem is solved through the ordered application of problem decomposition and reformulation, made possible in large part, by an appropriate uncertainty analysis.

In each SLD, the top event is the demonstration of acceptable risk for the safety issue it addresses. All basic events describing initiating factors, conditioning event scenarios, primary event scenarios, secondary event scenarios, release scenarios, possible source term configurations and amounts, pathway scenarios, and uptake and consequence scenarios are identified with groups of cut sets. (A cut set is a collection of basic events [individual sub-problems] that, when solved, guarantee the occurrence of the top event.) The power of the SLD is simple: It makes explicit and plainly visible all possible cut sets that contribute to the success of the top event (i.e., acceptable risk).. Confidence in the SLD comes from the fact that it must be prepared by experts of the phenomenology and that each cut set must reflect information based upon an accurate, complete, and expert domain analysis.

Because the SLD makes the completeness and adequacy of this domain analysis immediately visible to any expert or peer reviewer, omissions, incorrect assumptions, and incomplete logic pathways can be quickly discovered and corrected. If the SLD is made part of a stable technical basis, it matures with the rapid elimination of errors and unsubstantiated expertise. The proof then comes about through the

construction of cut sets that lead to acceptable risk as a result of problem solving steps in pathways based upon the expertly established domain for the phenomenology.

A wide range of analyses are required to resolve any safety issue. If the composite tank safety problem is to be solved competently and conclusively, all sub-problems for each safety issue must be identified and either, eliminated, solved, or reformulation and then solved. A large number of subproblems may be eliminated from further analysis by identifying dominance relations among the subproblems. Through the analysis of cut sets in the SLD, the more risky and costly alternatives can be identified and eliminated from further consideration. The dominance relationships among subproblems will be driven by the most significant uncertainties in the data inputs and models (the uncertainty drivers) in the SLD.

The SLD provides a completely adequate template for an analysis of tank safety issues as a large-scale composite problem. The SLD provides a methodology sufficiently powerful to decompose, reformulate, and analyze the large-scale composite tank safety problem is possible. Without the aid of the SLD, the large-scale problem analysis and decomposition methodology could not be accomplished; the intricate, composite nature of tank safety problem would remain at least partially hidden; and we would be without a rigorous means for dealing with the problem in a systematic fashion. Thus, the SLD, when used to define and communicate a large-scale problem analysis, ideally meets the requirements for a technical basis for safety issue resolution and also as a template for the uncertainty analysis that must accompany the safety analysis.

The SLD is a powerful tool for presenting and analyzing the underlying, causative phenomenology for each safety issue because it allows the systematic and comprehensive application of analytical tools for problem resolution. But the SLD is not only a means for communicating our understanding of a particular safety issue. Indeed, the SLD serves as the phenomenological model itself. Moreover, as a computational template, the SLD provides the basis for the other essential components of a technically defensible tank safety analysis resolution strategy. These components are listed below and are illustrated in Figure 1. The SLD as a phenomenological model has been discussed in this section. Each of the components is discussed in a subsequent section.

1. **A phenomenological model.** In this case the SLD itself is the phenomenological model. Each SLD in this report captures a complete phenomenology for one of the tank safety issues, including the physical and chemical conditions necessary, the presence or frequency of potential initiators, the sustainability of the phenomenology, the actual reaction physics, reaction products and volumes, the inherent weaknesses or strengths of the containment vessel or its components, the resulting immediate consequences, transport and fate models, and long term and secondary consequence possibilities.
2. **A model for uncertainty analysis.** The uncertainty analysis model reflects the "flow" of information through the SLD and propagates uncertainty in input parameters and component models throughout the SLD. Input information about tank contents and conditions, including possible off-normal events and their consequences, are propagated upward to assess performance, and risk criteria are driven downward to establish requirements. (Section II).

Figure 2.1. Components of a Technically Defensible Tank Safety Analysis Resolution Strategy

3. **A model for evaluating the logical completeness of the tank safety analysis.** This model is based on the methods of reliability engineering and the use of cut sets to evaluate potential alternative paths through the SLD for the management of risk to safety. (Section III).
4. **A model for systematic resolution of the large-scale composite tank safety problem.** Each SLD poses a large number of sub-problems to solve. By treating the SLD as a decomposable large-scale problem, it is possible to eliminate, simplify, or combine sub-problems for faster, easier, and more economical resolution of safety issues (Section IV).

In summary, the SLD provides a technical basis for the rigorous and defensible solution of tank safety issues. By systematically exploiting the structure of the SLD, it is possible to achieve great simplifications in the resolution of the tank safety issues without ignoring the complexity of the problem or imagining that uncertainties do not exist. These advantages of the SLD are also highlighted in Figure 2.1

2.2 Uncertainty Analysis For Tank Safety Issues

For each safety issue (and for other characterization questions as well), the SLD provide a technically defensible structure for computing estimates of risk. These estimates, in turn, can be compared with acceptability criteria such as applicable regulatory or safety limits. For these comparisons to be meaningful, the risk estimates must be accompanied by reliable estimates of uncertainty. At a minimum,

the estimate of uncertainty accompanying a risk estimate should capture the variability in model parameters and the level of understanding of the processes involved. In particular, the estimate risk uncertainty should include variability in those input values that satisfy "data needs," and it should also capture the uncertainty due to errors inherent in the key phenomenological models of the SLD, especially those that drive (i.e., have the greatest effect on) the risk calculation. Quantitative uncertainty analysis is indispensable to any method undertaken to provide a technically defensible estimate of risk. Indeed, the analysis of a safety issue is incomplete until such an estimate is provided.

In this report, the term "uncertainty analysis" refers to the steps taken to provide a quantitative estimate of variability for risk that reflects variability in the data and uncertainties or errors in the modeling process employed to estimate risk. This analysis is essential for both the understanding of each safety issue and its credible resolution. The SLD logic and decision structure serves as a phenomenologically complete template (i.e., model) for the rigorous inclusion and propagation of all data source and modeling errors associated with each safety issue. In this sense, the SLD is an analytical or computational template that can be reused to produce estimates of risk uncertainty for any given set of inputs. If different inputs or models are used with this template, different results will be obtained, but the decision structure (template) remains stable. The uncertainty analysis can be repeated as necessary to respond to new information that reflects changes in the waste state or tank conditions (performance assessment), or to evaluate the efficacy of various mitigation strategies.

It is important to understand where uncertainties enter and how they propagate through the computational template used to resolve a tank safety issue. To reach the top level event in an SLD (calculation of a risk estimate), one must begin at the bottom with data inputs. Each input to the SLD (a data need) or one of its component models is viewed as a "measurement" that is subject to error. Figure 2.2 illustrates that each data input to the SLD computing template has an associated error that is used by one or more models of the phenomenology. Figure 2.2 also shows how Model 1 uses data inputs in its domain (a subset of all inputs to the SLD called a cut set) to produce output information (the Model 1 co-domain) that eventually becomes the input (domain) of another model higher in the SLD diagram (e.g. Model 2). Any errors in the inputs to Model 1 are propagated through this calculation to (the domain of) Model 2, which in turn carries the effect of these errors to higher levels of the computational structure. Moreover, through inaccuracy or incompleteness, Model 1 may also contribute directly to errors in the inputs to Model 2. A comprehensive uncertainty analysis will incorporate estimates of the effect on risk of all significant sources of error in the risk calculation. Depending on the particular situation, there many possible sources of error, some of which are identified in the next section.

The process of using the SLD template for uncertainty analysis has a dynamic connotation. The SLD represents a "flow" of information that provides the structure for a corresponding propagation of uncertainty. The SLD structure directs the "flow" of information as it moves from input (data needs) to the top level event, output of a risk estimate to compare with acceptability criteria. The uncertainty model involves obtaining estimates of uncertainty for each data need and then, at each junction in the SLD, showing how variability in the inputs at one level is reflected in (or propagated via suitable models) the output(s) that serve, in turn, as inputs at the next stage. The SLD serves as a template for calculating a risk estimate, and it also **acts as the structure for a companion composite uncertainty analysis model**. The uncertainty model involves obtaining estimates of the variability for each data input and then, at each junction of the SLD, showing how input variability, together with model uncertainty, is reflected in (or propagated to) the estimate of risk uncertainty. The SLD greatly facilitates this analysis by exposing the many sources of uncertainty associated with each phenomenology (safety issue) and by helping to focus on those sources that dominate the estimate of risk uncertainty.

Typically, an SLD contains many potential pathways than can lead to a successful outcome, not all of which are equally desirable. A sensitivity analysis can be conducted identify those alternatives or pathways that are most desirable (in the VOI sense) and also to eliminate undesirable alternatives. Sensitivity analysis refers to steps taken to estimate the effect on a model response (output) of variability in one or more input parameters. Through a systematic elimination of the alternatives (i.e., cut sets) that are dominated by others, the primary goal of sensitivity analysis is to achieve a parsimonious (minimal) representation of the "model" (the SLD and its key component models) by identifying non-influential variables and components that can be ignored in subsequent uncertainty analyses. Thus, sensitivity analysis is a quantitative tool to use in paring the SLD (i.e., the decision process) down to a minimal set of essential pathways. This paring or reduction process is especially important because of the potentially large number of sub-problems revealed by each SLD. The complexity of the analysis for a given alternative will depend on how close the estimated risk for that alternative is to established limits. Thus, the "essential minimum" components of any SLD may differ from tank to tank, depending on how close the estimated risk is to the acceptance criteria.

2.3 The SLD as a Technical Basis for Uncertainty Analysis

2.3.1 The SLD as a Computational Model

The technical basis for uncertainty analysis must reflect the essential logic of phenomenology for each safety issue. Moreover, the methods of analysis must provide the ability to rapidly evaluate distributions for data values and parameter estimates that reflect the tank condition or waste state so that the uncertainties can be quickly re-evaluated when changes (either actual or hypothesized) occur. This requires a strategy in which it is possible to rapidly propagate changes in probabilities at any point in the SLD to others so their effect can be assessed at other points in the problem structure. This key concept is illustrated in Figure 2.3. The technical basis for the uncertainty analysis is an SLD that reflects our current understanding of and strategy for resolving a particular safety issue. The SLD also provides the structure or "flow" for the corresponding uncertainty analysis. Starting at the lower left in Figure 2.3, it must be possible to propagate changes in input distributions upward through the SLD, with these changes being processed by each model enroute the top-level event (an estimate of risk). In this fashion, uncertainty in input distributions and process models will be reflected in estimates of risk uncertainty in such a way that components of risk uncertainty are specifically identified with particular contributing sources.

Figure 2.2. Schematic SLD Showing how Uncertainties Enter and Propagate

The risk estimate uncertainty obtained from this “bottom up” analysis can be used to evaluate our current state of knowledge (i.e., performance) relative to some established regulatory or performance criteria for risk (r_0). An acceptable risk is achieved when the probability (p) that the estimated risk (r) exceeds the limit r_0 is sufficiently small, that is, when $p = \Pr(r > r_0) < p_0$ for specified values of r_0 and p_0 . Moreover, a sensitivity analysis will reveal the main sources of uncertainty (uncertainty drivers) and points of greatest sensitivity for risk uncertainty.

Similarly, it is possible, beginning with some acceptable regulatory risk criteria (r_0), to traverse downward through the SLD tree. This “top down” analysis can be used to specify precise uncertainty bounds for models and data requirements. The required precision will depend on how close the estimated risk (r) is to the established criteria (r_0). In those cases where the risk is far from the limit, a very simplified analysis involving only one or two alternatives may suffice; in other cases, a more comprehensive analysis will be required.

The SLD provides a template for calculating risk for a given set of inputs, each of which satisfies a data need (i.e., each input is the value determined for one of the of SLD data needs parameters). The SLD template also provides the structure for a model to estimate the uncertainty that accompanies this risk estimate. That uncertainty model involves obtaining an estimate of variability for each data need (input) and then, at each junction in the SLD, showing how variability in the inputs is reflected in (or propagated to) model output(s) which, in turn, become inputs at the next stage. This process of propagation spreads throughout the entire SLD and terminates when an estimate of uncertainty for computed risk is finally

obtained. A schematic depiction of this process is shown in Figure 2.2. An example is given at the end of this section.

An initial step in establishing an uncertainty model is to estimate the uncertainty of each input variable. The appropriate method for estimating parameter uncertainty is determined by the method used to estimate the parameter. Data needs parameters (i.e., model inputs) may be 1) estimated from observed or experimental data, 2) derived from theoretical models, 3) obtained by simulation (which involves a model), or 4) determined by consensus (expert opinion).

Many statistical methods are available for estimating the variability in observed experimental data. The resulting variability estimates reflect measurement variability, confounded with variability attributable to sources that have not been accounted for in the modeling process. For example, if the analysis does not account for spatial differences in tank contents, then spatial differences become another (confounding) source of variability included in the resulting estimate of sampling variability. Analytical or differential methods are appropriate for estimating the variability associated with estimates derived from theoretical models of physical-chemical processes. Alternatively, uncertainty estimates for parameter estimates derived from models may be obtained by simulation, where input parameters to the simulation model have been assigned suitable distributions. Finally, methods exist for estimating the uncertainty in values that are established by expert opinion. Because the large-scale composite problem posed by any one of the tank safety issues may involve all types of uncertainty, any or all of these methods may be used to obtain estimates of parameter uncertainty. The SLD will enable the uncertainty analysis to employ the correct uncertainty estimator for the right sub-problem in the overall tank safety issue.

Figure 2.3 The Role of the SLD in the Complimentary Processes of Determining Information Requirements and Evaluating Performance

A second step in developing an uncertainty model is to estimate the contribution to risk uncertainty of each of the key models in the SLD. Uncertainty in input variables is propagated through various models to higher levels of the SLD as shown in Figure 2.2. Moreover, the models themselves may be sources of additional uncertainty in the risk estimate at the top level of the SLD. A model may contribute to risk uncertainty because it is an incomplete or inadequate representation of the phenomenon it is intended to model. A model may also contribute to risk uncertainty because the data used to calibrate or parameterize the model are incomplete or subject to error.

Methods for estimating model uncertainty exist, but are not so well known as those employed to estimate uncertainties for observed data. Possible methods of estimating model uncertainty include analytical or differential methods, simulation, and (where they exist) comparing the results of competing models.

The propagation of variability estimates and its analysis may be accomplished in various ways, depending on the complexity of the model and the availability of information. In those rare cases where the model is simple enough, the variability of its output(s) may be determined analytically. As a minimum every SLD will generate a tree structure of variances as seen in Figure 2.6. An example will further clarify just how these uncertainty structures can be used. Alternatives to analytical methods involve either some method of approximation or Monte Carlo simulation. Differential/adjoint methods of approximation involve calculating the derivative of the model output with respect to each input variable. These derivatives are then combined and evaluated at selected points to produce estimates of model output uncertainty. For models with a complex structure and a large number of inputs, it may be prohibitively expensive or impossible to compute the required derivatives. The remaining alternative is then to use Monte Carlo simulation. Results of the Monte Carlo simulation may be used either to estimate the derivative required to approximate the variance of the model output, or they may be used to estimate the output distribution directly.^a In either case, the resulting uncertainty estimates are carried along to the next stage of analysis required by the SLD.

For codes with long computing times, the Monte Carlo simulation may be too time consuming or costly. It is possible to minimize the number of runs required to achieve a stated precision by carefully designing the simulation or by developing approximations to the response surface of the underlying model over the range of interest. The Latin hypercube sampling (LHS) method, for example, is one that has proven very efficient for selecting the sets of input values for a Monte Carlo simulation. In many cases, it may be practical to develop response surface approximations that capture the "essential" features of the underlying code but which have much shorter run times. The key point is that methods for designing very efficient simulations studies are available. By means of a systematic sensitivity study, it is possible to reduce the total work required to perform an uncertainty analysis. The sensitivity analysis can be used to identify those models and parameters in the SLD whose variability does not contribute significantly to risk estimate variability. These noninfluential parameters/models can then be eliminated from consideration^b and the uncertainty analysis can be minimized by including only those variables that do significantly affect risk estimates in subsequent calculations. In other words, a sensitivity analysis can be used to ensure that the uncertainty analysis is done in a parsimonious (minimal) manner. The additional cost of a sensitivity analysis is modest because most of the computing required for the sensitivity analysis would be required for the uncertainty analysis even if the sensitivity analysis were not carried out.

---

^a A statistical method of estimating the derivatives employs the coefficients of the input terms in a first-order approximation, linear fit, to the model output.

^b One way to accomplish this is to set noninfluential parameters at their reference values in all subsequent calculations.

As shown in Figure 2.3 above, the "flow" in the SLD template can be in either direction. Regulatory requirements and constraints may be imposed to determine acceptable risk. Risk constraints in turn flow down through process models in the SLD until they are ultimately converted into uncertainty constraints for data needs. Conversely, current estimates of uncertainty for input parameters can be propagated upward through the SLD to estimate risk uncertainty. In particular, it is possible to make various statements about the confidence that regulatory limits are met. These statements can be used to determine whether additional information and/or mitigative action is needed. If the probability that estimated risk exceeds the acceptable limit is sufficiently small, then no further action is required and the issue is resolved. If not, then further information and/or mitigative action is required. In the latter case, the SLD identifies the acceptable alternatives and a decision-making tool, such as a value of information (VOI) calculation, can be used to help determine the least costly and most effective mitigative action. The SLD computational template provides an efficient mechanism for iterating this "up-down" process to achieve satisfactory resolution of the safety issue in question. It is possible to ensure, using the SLD template, that resolution is accomplished for each tank with the least costly solution from the class of alternatives with acceptable risk. This concept is discussed further in the next section.

It has been argued that the technical basis for uncertainty analysis must provide the necessary rigor and flexibility to handle a large number of comparative analyses (it is convenient to call them sub-problems) of alternatives associated with each safety issue. The technical basis must deal with data uncertainty, parameter estimation, and model uncertainty. And, in the context of a specific sub-problem, it must deal with independent and/or intersecting information domains (data or model inputs) that contribute (via the modeling process) to a common output (co)domain. Most importantly, for the rigor necessary to successfully resolve a particular safety issue, **the technical basis must be able to handle all probability distributions and sources of uncertainty that pertain to each safety issue.**

In summary, a large number of individual sub-problems identified for each of the major tank safety issues must be systematically resolved through problem decomposition and reformulation. These problems must be set up and solved repeatedly, with all critical probability and uncertainty relationships being considered and propagated to all levels of the problem structure (SLD), either to determine requirements (as we move down) or to evaluate performance (as we move up), until the issue is satisfactorily resolved. The task must be done systematically so that the most acceptable alternatives are quickly identified and the least acceptable alternatives are quickly eliminated. Finally, the task must be done as efficiently and economically as possible, using the most powerful off-the-shelf software, so that uncertainties can be quickly recalculated to evaluate new information or changing tank conditions. The SLD can provide a sound basis for an uncertainty analysis model that meets these requirements.

2.3.2 Sensitivity Analysis: Comparison of Alternatives

The SLD template acts not only as a computational tool for uncertainty estimation, but also provides the overall structure for systematically estimating the uncertainty for various alternative pathways to acceptable risk. By tying the structure of the uncertainty model closely to the SLD, it is possible to exploit the efficiencies that derive from the organizational structure imposed by the SLD to reduce the scope of the overall problem to its essential minimum. This minimization is accomplished by means of a systematic sensitivity analysis.

There are many alternative paths through a the typical SLD that can lead to resolution of the related safety issue. The "up-down" analysis illustrated in Figure 2.3 is a valuable tool for identifying and comparing alternatives. A sensitivity analysis can be used to compare the variability in risk from two possible alternatives (In the SLD context, these alternatives define subsets of data input variables and process models called cut sets) in order to select the best (in the VOI sense). Equally as important,

alternatives that are dominated by others can be identified and eliminated. Thus, the "up-down" uncertainty analysis can be used both to select the most desirable alternatives and to eliminate the least desirable ones. Used for this purpose, sensitivity analysis is a powerful tool for minimizing the scope of the safety analysis by "pruning" the SLD tree of nonessential branches. Of course, a systematic sensitivity analysis may lead to a different ranking of alternatives for each tank. A very simple analysis may resolve the issue for some tanks, while a full-scale analysis may be required for others. In any case, a particular safety issue will be resolved via the least costly alternative from the set of alternatives with acceptable risk.

Alternate paths through the SLD are characterized by subsets of the models and data input variables of the SLD called cut sets. These cut sets extend from input data at the bottom of the tree through model outputs further up the tree to the very top event, estimation of risk. At each level, the uncertainty analysis propagates the variability in these inputs through the model to the output(s). The model outputs(s), together with associated estimates of uncertainty become, in turn, inputs to other models at higher levels in the SLD. Large subsets of data and models can come together as inputs together at higher levels in the SLD structure. At any point in this process, the contribution to uncertainty from a particular subset of input variables or component process models, called a cut set, can be computed. This is done by estimating the uncertainty both with and without the contribution from the cut set of interest. The systematic application of this idea, called sensitivity analysis or cut set analysis, can be used to systematically identify the best alternative for resolving a given safety issue.

Sensitivity analysis of cut sets is illustrated in Figure 2.7, which shows the variance relationships among all pathways through the SLD to the top level event. The SLD structure is described in the example given at the end of this section. Note that the cut set on the left-hand side of the figure, indicated by the branch of the SLD with bold lines, contributes to the total variance of the top-level event through the terms shown in the box. Terms outside the box represent contributions from the other major cut set branch of the SLD tree. (In examples of realistic complexity, there will be many contributing branches.) Clearly, the total variance will be more sensitive to the branch or cut set that contributes most to the overall sum.

By systematically computing the total variance of the top-level event with and without the contribution of a particular alternative (cut set), it is possible to identify the best alternatives (in the VOI sense) and, further, to quantify the (maximum) value of reducing the variance contribution from a particular cut set. Equally as important, this process can also be used to identify the least desirable alternatives. By eliminating these alternatives from further consideration, it is possible in effect to "prune" the SLD structure to its essential minimum. This pruning process may result in a different outcome for each tank, but the safety issue in question will be resolved for each at the lowest cost alternative consistent with safety requirements. Moreover, the resolution process will take full account of uncertainty in the risk estimates on which a decision is based.

2.3.3 Sources of Uncertainty

With the assistance of Figure 2.2, one can begin to visualize the sources of uncertainty and how these uncertainties propagate through an SLD. Uncertainty in risk estimates can arise from a variety of sources. One of the most familiar sources is measurement variability due to sampling and instrument errors. There are many statistical methods for obtaining suitable estimates of measurement variability under a wide range of observational/experimental conditions. The matter of how to design data collection activities (experimental design) and analyze the resulting data is vitally important to the success of this (or any other) safety issues resolution strategy. This topic is discussed elsewhere in this

report. In this section, we assume that suitable input distributions for data needs have been established and are available for uncertainty calculations.

Uncertainty can arise for other reasons, such as a lack of information. The Historical Tank Contents Estimation (HTCE) model concentration estimates, for example, are uncertain because (among other reasons) the records used to derive them are incomplete. Incompleteness in the records implies insufficient or incomplete knowledge and understanding, and can be a major source of uncertainty. Observational information (data) can also be incomplete. Data from core samples, for example, is often incomplete because it lacks adequate spatial resolution.

Uncertainty also appears in the form of incomplete or inadequate phenomenological models (i.e., models that do not adequately predict the phenomenology they are designed to model). Each SLD contains many models that are potential sources of uncertainty. This has nothing to do with good or bad science. In fact, good science dictates that we recognize and rigorously deal with all sources of uncertainty. It is in this sense that each SLD also represents a tool-in-the-large for dealing rigorously with these composite uncertainty problems. The SLD template allows the problem to be solved by working through the larger problem, one subproblem and one source of uncertainty at a time, until the top level SLD event (acceptable risk) is achieved.

The top level event in the SLD for each safety issue is acceptance of the risk associated with possible harm or loss of safety due to off-normal events associated with that issue. The question of risk acceptability is resolved when each sub-problem in the SLD is resolved and the risk at each level is found to be acceptably low. Only then can it be asserted with confidence that there is no appreciable or unacceptable safety risk for the phenomenon in question. Each of the individual sub-problems, including the data collection, models, probability distributions, the uncertainty analyses itself, and all the other processes in the SLD are potential contributors of uncertainty. The SLD helps us deal in a systematic fashion with the inescapable sources of uncertainty and variability in characterization by displaying how that uncertainty will affect our results and where the significant sensitivities lie.

2.3.4 Representation of Uncertainty

Just how uncertainty is represented is extremely important in any structure with the logical rigor of the SLD. There are many ways to represent uncertainty. These include the variance associated with any data sample, model output, or parameter estimate. One can also consider the range determined by the highest and lowest possible values, the interquartile range, measures of confidence for the width of the intervals in which the true estimate lies and tolerance intervals. These are few of some of the more widely used representations of uncertainty. The interpretations of these measures may differ, depending on how that are derived. The variance, for example, is not appropriate in all cases. In particular, the variance is not appropriate for characterizing the variability of a parameter whose distribution is highly skewed or one that is bimodal. Moreover, not all measures of uncertainty may be used globally to describe the uncertainty for an entire set of variables that must be measured to resolve a particular safety issue. Global measures will have a different interpretation from individual data set measures.

The one representation, however, that always contains all the relevant uncertainty information for a variable is its distribution function. In fact, all other representations of uncertainty can be derived from the distribution function.^a Use of distribution functions, then, is a key to providing

---

^a Distribution functions may be empirical or theoretical, depending on whether they are determined from a set of observations or from theoretical considerations.

sound uncertainty analyses for the SLDs. Distribution functions (and the statistics derived from them) admit various interpretations depending on how they are derived. Distributions derived from observational or experimental data are measures of sampling variability (possibly confounded with other sources of variability or error that have not been adequately modeled such as, for example, operator or instrumentation error). Distributions derived from a Bayesian approach admit a "degree of belief" interpretation but lack the frequentist interpretation of those based on actual observations or on sampling data. Distributions may also be chosen to represent extreme or bounding/limiting cases. Inferences based on distributions obtained in this manner often have what is called in statistical terminology a minimax interpretation.

It is important to realize that regardless of how distribution functions are obtained, all are treated alike in any subsequent sensitivity or uncertainty analysis.

2.3.5 Estimation of Uncertainty

Just as uncertainties arise for a variety of sources and have a variety of interpretations, there are also a variety of ways to estimate uncertainty. The appropriate method in a given situation depends on the source of uncertainty, the complexity of the situation, how well the variable must be known, and the assumptions that one is willing to make. The SLD forces one to document any assumptions that are made. Methods for estimation of uncertainty generally fall into one of three categories: analytical, statistical, and simulation. The fundamental analytical method involves the mathematical derivation of the probability distributions of response (output) variables from the distributions of input variables appropriate to the model that relates the variables in question. For models that are complex enough to be of practical interest, deriving the mathematics for the analytical approach is generally an intractable problem. Consequently, some type of alternative is required. Alternatives include approximations, statistical methods, simulations, and the use of expert opinion. Approximations usually involve a Taylor-series type approximation to the true underlying distribution. In practice, multiple alternatives are usually used in combination. The class of differential/adjoint approximation methods, frequently referred to as propagation-of-error methods, are widely used for sensitivity/uncertainty analysis of models of physical-chemical processes.

Information required to implement approximation or propagation-of-error-methods may be obtained in a variety of ways. Two of the most widely used are statistical methods and expert opinion. Statistical methods may involve sampling data, obtained by direct observation of the variables of interest, or experimental data, which obtained from laboratory experiments and other indirect methods (e.g., from other tanks with similar properties). Statistical methods for sampling and analysis are discussed elsewhere in this report.

Methods exist for estimating uncertainty distributions by means of expert opinion when observational data are not available. An example of how this is done is also presented elsewhere in this report.

Key process models in the SLD may require simulation for estimating uncertainties and propagating them to the next level of the decision process. When simulation is employed, it is possible to obtain, not only some estimate of the uncertainty, but an estimate of the entire probability distribution of the response variable that is the output of the process model. Whatever the estimation problem, a variety of tools exist to provide estimates, to approximate complex models, and, in general, to propagate the real, underlying uncertainty to the top level event. In particular, methods (e.g., Latin hypercube sampling) exist for designing very efficient simulation studies that involve a minimum number of computer runs.

Because the large-scale composite problem posed by any one of the tank safety issues may involve all types of uncertainty, any or all of the methods identified here may be used to obtain estimates of parameter uncertainty at one point or another in the analysis. In any event, the data are analyzed to obtain the required estimates of (the distributions of) variability for the parameter(s) of interest. In subsequent analyses, all uncertainty estimates will be treated in the same fashion regardless of how they were obtained. The SLD will enable the uncertainty analysis to focus the right uncertainty estimator on the right sub-problem in the overall tank safety issue. The goal is realistic uncertainty estimates that are free of biases resulting from hidden or aliased variables.

2.4 Uncertainty Analysis Example

An example is presented to illustrate how an uncertainty analysis can be done using the structure of the underlying SLD as computational template. The example SLD in Figure 2.4 is used to show how parts (subproblems) of a composite problem might relate to one another and how the uncertainties in the parts propagate through the alternative branches (sub-problems). From the uncertainty modeling viewpoint, we will refer to the SLD template in Figure 2.4 as a model. In this example, the phenomenological model for an off-normal Flammable Gas (FG) release is used to estimate the frequency of dome failure due to FG ignition. The model in Figure 2.4 can be regarded as a part of the larger SLD that was developed to resolve the flammable gas issue. The model has been chosen for illustrative purposes and is not intended to be definitive.

The model depicted in Figure 2.4 has three levels. The output of the model--the top level--is an estimate of the frequency of dome failure due to a FG ignition (denoted by f_c). The estimated frequency f_c is modeled as the product

$$f_c = p_2 * f_i$$

where

$p_2 = \Pr(V_r > V^*)$ is the portion of the time that the volume of FG in the dome (V_r) is great enough to cause dome failure if ignited

and

f_i = estimated frequency of a FG ignition.

In this example, p_2 is calculated as the probability that the actual gas release volume exceeds the threshold V^* for dome failure.

At the second level, the threshold V^* is determined, from information about the structural integrity of the tank and the combustability of its contents. The data needs and the models used to determine V^* are not shown in Figure 2.4.

In the right branch at the second level, f_i is modeled as

$$f_i = p_1 * f_s$$

where f_s is the estimated spark frequency (a data need) and p_1 is the portion of the time the dome space is in a state where a spark would lead to ignition.

Finally, at the third level in this example, p_i is modeled as

$$p_i = f_r * \tau$$

where f_r is the rate of release of flammable gas from the waste into the dome space (a data need) and τ is related to the rate of ventilation (a data need). The data needs and the indicated models are summarized in Figure 2.4.

Notice that the top-level event and all intermediate results are built up quantitatively from variables and models at lower levels. This is true in general. Thus, it is possible to construct a chain of calculations that relate the top event (estimation of f_c in this example, estimation of risk in general) with data requirements at the lowest levels^a by means of a series of models (intermediate calculations). Beginning with measures of uncertainty in input variables, it is possible to estimate the uncertainty of the top event by propagating these uncertainties through the SLD computational structure. As indicated in a previous section, various measures of input uncertainty may be used.

Two alternatives for estimating risk uncertainty are illustrated, respectively, in Figures 2.5 and 2.6. (The structure of Figures 2.4 through 2.7 is identical to that of Figure 2.4, but the symbols have been suppressed to improve readability.) The first case involves determining the distribution of f_c by Monte Carlo simulation. Distributions were assigned to the input variables f_r and f_s (The distributions reflect uncertainty in f_r and f_s , respectively and may be determined using the methods described previously. A distribution was also assigned to p_i , the portion of the time that dome failure is possible. Realizations of f_r , f_s and p_i were drawn repeatedly from their respective distributions. The value of f_c was calculated for each realization and the resulting values were tabulated to produce the distribution for f_c shown in Figure 2.5. From this distribution, it is possible to estimate the probability that f_c exceeds some critical value. In general, the results are satisfactory when this estimated probability is small.

Although not shown in this example, the Monte Carlo simulation illustrated in Figure 2.5 can easily be extended to include estimates of the uncertainty in the models used in the calculations.

The second alternative for estimating the variability in f_c employs the same initial distributions for f_r , f_s and p_i as the previous case. However, instead of estimating the entire distribution, only the variance of each input variable and intermediate result is estimated. The variance of f_c is obtained, in turn by propagation. The formulae for propagation are summarized in Figure 2.6. Figure 2.6 shows how those variances propagate through the system, and how the estimated variance of the final output (risk) depends on variances of the initial inputs via the variances of the intermediate quantities.^b

---

^a Note that this process of assigning values to preceding inputs and outputs as values to following inputs is the first step in large scale problem decomposition, namely assigning values and computable specifications to the input data domains and the output co-domains for each subproblem. In the particular example for FG, we have a rather simple case. With the larger SLD, there are many domains and co-domains to specify, or eliminate, if possible. Doing this specification and keeping track of all domains and co-domains requires the machinery of large-scale problem decomposition methodology.

^b Simulation will be used to represent those subproblems of an SLD that require the running of an entire model. The model will be replaced with a representative transfer function, further eliminating the overall sub-problem complexity of the SLD. To obtain meaningful results from such a simulation, the input distributions must be estimated using the methods described in this section.

Clearly, the method selected to estimate uncertainty in a particular situation is (or ideally should be) closely tied to the way the uncertainty is represented. It is therefore highly desirable that uncertainty estimates for parameter values (inputs to the SLD or its component models) be derived at the time the parameters are estimated and that methods used to estimate uncertainty be compatible with those used to obtain the parameter estimates. In general, it is preferable to use distributions to describe uncertainty (as in Figure 2.5) rather than some measure derived from the distribution (such as the variance) because the latter can be misleading if the underlying assumptions for its use are not satisfied.

The analysis of cut sets is illustrated in Figure 2.7, which shows the variance relationships among all pathways through the SLD to the top level event. Note that the cut set on the left side of the figure and indicated by the branch of the SLD with bold lines, contributes to the total variance of the top-level event through the terms shown in brackets. Terms outside the brackets represent contributions from the other major cut set branch of the SLD tree. (In examples of realistic complexity, there will be many contributing branches.) A systematic analysis of the contributions to the variability of the top-level event will help to decide which alternatives (cuts sets) to pursue and which to abandon. Indeed, by setting the contribution from a particular cut set to zero, it is possible to see the reduction in uncertainty that would result from "perfect" information. The value of this reduction could be computed to determine whether it was worthwhile or whether some other alternative should be pursued. The effect of setting the variance of f_s to zero, for example, is illustrated in Figure 2.8. In Figure 2.8, the variance of f_c is computed from the input variances for the distributions used in Figure 2.5 and the variance formulae shown in Figure 2.6. Figure 2.9 is similar, except that the variance of f_s has been set to zero. The calculations in Figure 2.8 are summarized in Table 2.1. The estimated variance in the first case is

$$\sigma_{fc}^2 = 2.7 * 10^{-8}$$

If we suppose that $\sigma_{fs}^2 = 0$ (i.e., that the spark frequency is perfectly known), then

$$\sigma_{fc}^2 = 4.3 * 10^{-9}$$

Thus in this example, $\sigma_{fc}^2 = 4.3 * 10^{-9}$ is the limiting variance that can be achieved by reducing an estimate of uncertainty in f_s , the spark frequency, to 0. This information can be used to compute the cost of reducing the variability in f_s relative to that of other alternatives presented by the resolution strategy (i.e., the SLD).

While this section presents an overview of alternatives available for estimating uncertainties, it is impossible in such limited space to present a comprehensive discussion of all of the strategies available for estimating uncertainties in risk estimates for a particular safety issue. No one alternative is best in all cases.

Having the stable template of the SLD will allow precision in pinpointing and selecting the best method for the type of data and type of uncertainty in each situation.

Thus, an important part of an overall strategy for estimating uncertainty is the determination of a suitable method for estimating the uncertainty for each significant parameter or process model in the SLD. A comprehensive strategy will define the experimental design and data analysis to be performed, and will integrate results from laboratory experiments and other indirect observations, expert judgment, and supporting Monte Carlo and other simulation

Figure 2.4. SLD Showing Estimation of Frequency of Dome Failure due to Flammable Gas Burn

Figure 2.5. Estimation of the Distribution of f_c by Monte Carlo Simulation. The vertical axis for each probability distribution shows relative frequency.

Figure 2.6. Estimation of the Variance of f_c by Propagation-of-Error Methods

Figure 2.7. The Cut Set Shown in Bold Lines Contributes to the Variance of the Top Level Event Through Bracketed Terms in the Expression for σ^2_{fc}

Figure 2.8 SLD for the FG Model with the Estimated Variances for the Uncertainty and Risk Distributions Shown in Figure 2.5

Figure 2.9. SLD Showing how the Estimated Variance for the Risk Distribution in the FG Model Changes as the Variance for the Uncertainty Distribution for One Input, the Spark Frequency, Changes from 6.5×10^{-6} to 0

Table 2.1. Summary of Calculations Illustrating the Effect of "Perfect" Information on Spark Frequency (f_s)

σ^2	Case (a)	Case (b)
$\sigma_{f_r}^2$	4.0×10^{-4}	4.0×10^{-4}
$\sigma_{f_s}^2$	6.25×10^{-6}	0.0
$\sigma_{p_2}^2$	$0.52 \times 4.010^{-4} = 1.0 \times 10^{-4}$	$0.52 \times 4.010^{-4} = 1.0 \times 10^{-4}$
$\sigma_{p_2}^2$	7.3×10^{-4}	7.3×10^{-4}
$\sigma_{f_r}^2$	7.3×10^{-8}	1.0×10^{-8}
σ_{fc}^2	2.7×10^{-8}	4.3×10^{-9}

2.5 Uncertainty Analysis Summary

The principal elements of an overall strategy for uncertainty estimation are discussed in this report. It is impossible, in such a limited space, to present a comprehensive discussion of the strategies for estimating uncertainties in risk estimates, for problem decomposition, and for sub-problem reformulation for any particular safety issue. This is especially true since each tank safety issue is in itself a large-scale composite problem with a significant number of individual sub-problems. No one alternative for uncertainty estimation is best in all cases. An important part of an overall strategy for uncertainty estimation is the determination of a suitable alternative for estimating the uncertainty for each significant parameter or process model in each of the SLDs. A comprehensive strategy will employ statistical principles of experimental design and data analysis, results from laboratory experiments and other indirect observations, expert judgment and Monte Carlo simulation. This strategy may introduce new elements into an SLD, but will not substantially alter its logic or major cut sets. Indeed, once the computational template is constructed, the analysis can be quickly done (or redone) to assess the effect of new information or possible mitigation strategies. The comprehensive strategy will also employ sensitivity analysis to reduce the number of variables and alternatives that must be considered to its essential minimum. It is important to realize that this report does not address the very important question of how to design and implement statistical sampling plans to acquire the data to estimate the required input parameters and their distributions; this topic is discussed in another section of the report.

Issues concerning uncertainty are at the core of each tank safety issue and the methods used in their resolution. As each SLD is further developed and each sub-problem is further specified, additional development of a comprehensive and defensible uncertainty estimation strategy will become feasible. Uncertainty analysis is absolutely crucial to the satisfactory resolution of every tank safety issue.

3.0 Potential for Resolution of Tank Issues Via Sampling

Sampling of the tanks contents and subsequent analysis of the material is a key element of any characterization strategy. Under the proposed new characterization strategy, tank contents information is needed. Although there are many issues associated with sampling and analysis of tank material, sampling continues to be an important source of information and an important part of the new characterization strategy. The reliability of that information must be better understood and compared against other methods for gathering information or taking action.

This section focuses on sampling and analysis defined as the process of obtaining waste material from the tanks and obtaining a measurement through application of some procedure on the waste material. This section focuses only on waste material sampling and does not attempt to cover vapor sampling, expert elicitation, nor other forms of sampling. Uncertainty is always present whenever sampling because an attempt is made to describe some population without examining every member of that population. Attempts are made to control uncertainty by controlling the measurement methods, controlling the number of samples taken from that population, and ensuring that the samples represent the population.

Three questions that capture the primary concerns expressed when evaluating the effectiveness of the sampling program for "characterizing" the tanks are:

- **Question #1: How many core samples are required to provide adequate protection against incorrect decisions (confidence)?**
- **Question #2: Given the heterogeneity of the waste, will sufficient confidence ever be achieved from economically feasible sampling?**
- **Question #3: Should core sampling be halted until a better technical basis can be developed?**

Each of these questions will be considered in this section and the necessary information and assumptions for deriving the answers to these questions will be discussed. The statistical and analytical issues surrounding the sampling and analysis aspects of characterization will also be outlined.

Although one would desire a simple answer (yes/no or specific number of cores) to the above questions, the final answers are fraught with dependencies. In an attempt to summarize the response for each of these questions, the following general answers are provided.

- **Answers #1 & #2: The number of core samples required depends on the degree of heterogeneity and measurement uncertainty, required protection against making incorrect decisions or the required confidence in our estimates, distance from decision thresholds, and statistical parameters of interest. If waste properties are sufficiently far from decision thresholds relative to heterogeneity and measurement uncertainties, very few samples may be adequate to support confident decision making (i.e., perhaps with the FeCN issue). For this to be feasible, heterogeneity estimates must be bounded and assumptions must be deemed acceptable.**

If waste properties are near decision thresholds relative to heterogeneity and measurement uncertainties, too many samples may be required (e.g., organics issue). Tank grouping may provide one avenue around this by "pooling together" nominal/variability estimates to reduce uncertainties. The value of reducing uncertainty via sampling must be traded off against the cost of control or intervention. Thus, for some issues, sampling may lead to sufficiently confident decisions while for others, control and/or intervention should proceed without further sampling. Unfortunately, we can not provide a technically defensible answer to these questions until uncertainty and risk-based VOI analyses are performed.

- **Answer #3:** Targeted sampling should first resolve the sample representativeness issue (how representative are samples?). If current sampling is viewed as screening level information and is used to obtain preliminary estimates useful for estimating some aspects of heterogeneity and measurement uncertainty, continued minimal sampling is useful. However, confident decisions based on sample results will require that a defensible technical basis be developed (as outlined in the SLDs) and issues of sample representativeness be addressed. Moreover, tradeoffs between sample-based decision making, monitoring and control, or intervention must be evaluated using uncertainty analyses and a risk-based value of information approach before one can determine whether current sample-derived characterization is worthwhile and further define an optimal sampling scheme. Thus, a blanket statement that further sampling is not appropriate would be incorrect without first determining the representativeness of samples and the cost/benefit of sampling verses monitoring and intervention.

In terms of cost, the process of obtaining a core sample is significant. A ballpark cost of obtaining a core sample currently is approximately \$250K-300K depending on many factors. If a large suite of analyses is performed, the analysis cost can equal the sampling cost. If fewer analyses are required (screening or composite), the cost of analysis can be one-fourth to one-half the cost of sampling. Thus, the cost of reducing sampling uncertainty may be high but the payoff may also be significant.

In the segments that follow, sampling issues will be presented and potential approaches for overcoming some of the issues will be discussed. The role of sampling in development of the comprehensive characterization strategy will be discussed.

3.1 Sources of Uncertainty Introduced Through Sampling and Analysis

During sampling and analysis, uncertainty can be introduced from a variety of sources. Sampled material may not be entirely representative of the entire tank contents. This may be due to significant stratification of materials within the tank or may be due to the relative heterogeneity of the materials even in a "well-mixed" tank. The sampling method or device can also be a source of variability. Once the samples are obtained, sample handling and sub-sampling of the material can introduce uncertainties. Finally, the process of sample preparation and analysis introduces uncertainty. Each of these sources of uncertainty is discussed in greater detail below.

3.1.1 Sample Representativeness

Representativeness is a relative term. Although a black cat is different from a white cat, either one is representative of domestic cats. Yet the black cat may not represent the reflective properties of all cats in the dark. Their differences are relative. Similarly, some material within a tank may adequately represent the tank contents even though it is not exactly the same composition as some other material in the tank.

The term "adequate" is the key and can only be determined relative to some decision or estimation criteria. This relatively representative concept must be maintained throughout this section.

Several constraints and issues exist in current tank sampling approaches that affect the representativeness of sampled material. These include:

- riser availability and random access
- heterogeneity and spatial variability
- sampler performance and percent recovery of cored material
- homogenization and sub-sampling.

Riser availability currently limits access to the tank contents. It has been postulated that the volume beneath the existing risers may be the worst place to obtain a representative sample because of the disturbing effects of activity that has occurred under many of the risers. It may be that the material under the riser bears little semblance to the material in the rest of the tank or it may be relatively representative. Under any characterization strategy, the bias that may be introduced by sampling beneath existing risers must be examined and either resolved, bounded, or eliminated.

Material within the tank is not entirely homogeneous (another relative term). As the material in the tanks was introduced, acted upon, and extracted, natural layers and profiles developed. Both vertical and lateral spatial variability exists to some degree in the tanks. Some stratification exists within most tanks. If the lateral variability were minimal, just a few cores would be needed to characterize the vertical variability. If the vertical variability were minimal, a few samples from the top of the tank contents would be sufficient. However, both lateral and vertical variability appear to be significant, thus making vertical profiles from more than one lateral location necessary. If severe lateral stratification exists, the number of cores required will be uneconomical.

Sampler performance also affects representativeness. Full recovery of a core has been an issue in the past. If the sampler does not recover nearly 100% of the material expected, some material or layer may not be adequately represented in the sample. This deficiency could bias estimates obtained from sampling and would introduce another uncertainty. Either the sampler performance must be improved (many efforts are proceeding towards this end) or the bias must be bounded to be able to rely on sample results. Another important factor is whether the sampling procedure/device affects waste properties of interest. For example, moisture properties may be affected by either injection of water or dryout during sampling. Waste property maintenance must be considered.

Homogenization and sub-sampling occurs once the sampled material becomes available. Thus, a sample of a sample is obtained. This homogenization step is an attempt to minimize the uncertainty introduced from sub-sampling. The process of homogenization appears to successfully minimize the uncertainty instead of adding another level of uncertainty [Heasler, 1994]. The effect of sub-sampling of material as a source of uncertainty is probably constituent dependent. Often core composite samples are created by taking material from each homogenized vertical segment of a particular core sample. Compositing minimizes the effect of vertical variations but also represents another source of sub-sampling variability.

For tank samples obtained to date, possible uncertainties or biases due to the riser availability and sampler performance issues have not been estimable. The spatial variabilities and homogenization/sub-sampling variations have been estimated for some tanks where replicate cores are available using statistical analysis of variance (ANOVA) techniques. The spatial variability often is the major contributor to uncertainty for many constituents. This is further illustrated in a later section of this report.

3.1.2 Measurement Uncertainty

The process of obtaining a measurement includes both sample preparation steps and application of an analytical instrument. Usually duplicate analyses are conducted and the average used so as to minimize the analytical uncertainty. During the sample preparation process, digestions and dilutions are common and can be a potential source of measurement error. Particle size is an important factor that also has the potential of introducing variability. Different personnel can also be a potential source of uncertainty due to slight differences in how they implement the procedures. Some analytical methods are more accurate and/or precise than other methods. How quality control indicators (spikes, duplicates, QC/QA standards, calibration standards) are used can affect the estimated measurement uncertainty.

In general, measurement uncertainty for tank samples has been minor compared with the spatial variations in the waste. However, for some constituents measurement uncertainty is a major contributor to the overall uncertainty. Measurement uncertainty is not rigorously examined in this report.

3.2 Overcoming Sampling and Analysis Roadblocks

Several potential paths and tools exist that may lead to a successful sampling and analysis scheme for the tanks. Some of these are outlined below. Note that the value of these sampling pathways should be balanced against other alternatives such as monitoring or intervention. For sampling to be a viable pathway to obtain information for confident decision-making, either the sample representativeness must be bounded or estimated, or new sampling technologies must be developed.

3.2.1 New/Improved Sampling Technologies

Obviously, if there were less-expensive methods for obtaining samples at random or at specified locations within the tanks the sample representativeness issue would be more easily resolved. Some new technologies are under development but appear to be somewhat limited. Access to areas in the tanks that are not directly underneath the risers is needed to determine how well the material beneath the risers represents the entire contents. Moreover, if stratification within the tank exists, sample(s) from each strata may be required to provide adequate information for confident decisions.

Sampler performance studies must also be an integral part of any sampling program. Currently available samplers sometimes recover much less than 100% of the material expected. It is difficult to determine what these partial cores represent. Recent improvements have increased the reliability of the samplers but little has been done to quantify or bound the bias that may be introduced from the sampler.

3.2.2 Sample Representativeness Approaches: Bounding Sampling Bias and Tank Grouping Studies

One criticism of continued sampling is that we do not know if the samples obtained are relatively representative of the entire tank contents. This issue must be resolved before sample results can be used for confident decision making. The two related issues to be addressed are

- how representative are samples obtained underneath existing risers
- what is the spatial variability of the material in the tanks. Two approaches could be explored to resolve this issue.

Sampling bias due to riser availability could be estimated or bounded. Access to the material that is not underneath potentially nonrepresentative risers is needed. A few tanks that are expected to be highly heterogeneous could be used to bound the potential bias which could be introduced from sampling underneath existing risers. This might be accomplished by either adding new risers, employing a new sampling technology, or sampling through existing risers where little activity has occurred. If the sample-to-sample differences observed between "new" and "existing" riser samples are not significantly greater than the sample differences observed between existing risers, then the existing risers could be considered representative of the tank contents. If the sample differences between new and existing are significant, an estimate of the bias can be obtained and used as a bounding limit in future sampling campaigns using existing risers. If that bias is too large, sampling beneath existing risers would not provide the information needed for confident decisions and sampling should be halted until new technologies are available for sampling or tank material is mixed.

If the spatial heterogeneity within a tank is too great, a tank grouping strategy might be appropriate. Sampling requirements may be too costly if each tank is considered as a separate entity (population). If tank groups could be defined such that each tank within a group has similar composition and spatial inhomogeneity, the group of tanks could be treated as a single population. Thus, fewer samples would be required per tank to achieve the same degree of confidence in estimates of tank contents. For purposes of illustration, if we allow for several simplifying assumptions (an extremely oversimplified example), a one-sided upper 95% confidence limit on the mean (y) for a given population can be estimated as:

$$y + t*s/\%n$$

where t is derived from Student's t distribution (dependent on n), s is the estimate of the sampling and analytical standard deviation for an individual observation, and n is the number of samples obtained.

If two samples were obtained from a tank with one analysis per sample, the formula would be:

$$y + 4.46*s$$

If the within-tank spatial variability for a group of ten tanks was shown to be similar and could be "pooled", then the formula for a 95% confidence upper limit for a single tank where s is now the pooled standard deviation estimate, would be:

$$y + 1.28*s$$

Taking this one step further, suppose that for the tank group the tank-to-tank compositional differences are no greater than the pooled-within-tank variations. Then, the overall mean from the 10 tanks would provide an estimate of the tank contents for each tank. If y is the tank group mean and s is the pooled spatial standard deviation, the confidence limit formula would then be:

$$y + 0.39*s$$

If the within-tank variabilities are all similar (thus permitting pooling), the amount of information available and therefore the confidence the variability estimate is increased. In effect, greater precision is achieved by increasing the sample size. Therefore, the multiplication factor applied to the spatial variability estimate decreases from 4.46 to 1.28. If the average compositions across all 10 tanks are not significantly different, then we can average across all tanks and our information and confidence

increases. Thus the multiplication factor decreases from 1.28 to 0.39. There are many variations on this theme and the example described above, although greatly oversimplified, demonstrates the point.

By employing a tank grouping strategy, we may be able to obtain adequate information to make confident decisions from a few samples per tank. Estimates may be sufficiently precise to confidently conclude that the true mean concentrations are less than (or greater than) some decision threshold or to provide an accurate estimate for some parameter of interest. This approach assumes that the parameter of interest is the mean composition. If the decision criteria is based on an extreme value or the composition of a particular layer in the tank, this tank grouping argument would have to be revisited and may not apply.

3.2.3 Estimating and Controlling Measurement Error

Although spatial variations are usually the dominant factor affecting the overall uncertainty, measurement variability does often add significant uncertainty to some constituent determinations. It is not uncommon to find significant errors in data obtained from the analytical laboratory. Many of these errors are detected through statistical analyses conducted in a holistic manner on all of the tank data. When statistical evaluations are conducted for the purposes of developing a tank characterization report, "data cleanup" is a significant part of the effort.

A rigorous analytical laboratory QA/QC program should be followed to ensure traceability, appropriateness of the procedures and methods used and identification of anomalies. A program to identify problems with a particular data value and to examine data in the context of the entire tank sampling event is important. Over the past few years, significant efforts have been dedicated to ensure that accuracy and good precision occur in the analytical laboratory. However, some anomalies continue to slip through. An expanded treatment exploring all the potential sources of measurement uncertainty is beyond the scope of this report. It is sufficient to say that continued efforts to control this source of uncertainty are necessary.

3.2.4 Statistical Models to Estimate & Combine Uncertainties

Statistical methods must be employed whenever sampling or measurement occurs to derive estimates and understand the uncertainty inherent in those estimates. Statistical methods are being employed in the TWRS characterization program to provide best estimates of tank contents and determine the uncertainties associated with those estimates. Statistical methods are only as good as the data and so far, the possible biases introduced by sampling under existing risers and poor sampler performance have not been estimable.

Statistical methods exist and are being employed on tank sample data to quantify the uncertainty contributed from some of the major sources discussed above. Particularly, the spatial and measurement variations are being quantified when the sampling and analytical schemes permit. The statistical models employed depend on the sampling and analytical schemes used to obtain the data. The sampling scheme determines the form of the statistical model. Most statistical models for core sample data results are variations of the form:

$$y_{ijkl} = u + L_i + V_{ij} + S_{ijk} + A_{ijkl}$$

where y_{ijkl} is the observed result, u is the true mean concentration of the particular analyte for the tank, L_i is the deviation from the mean attributed to the lateral location, V_{ij} is the deviation from the mean attributed to the vertical location, S_{ijk} is the deviation from the mean attributed to the sub-sample (often

two samples taken from a homogenized segment), and A_{ijkl} is the deviation from the mean attributed to analytical variations. Terms in these statistical models may be deleted, combined, or added depending on the sampling and analytical scheme that is employed.

Examples showing the results of employing these statistical methods are shown for 13 tanks in Table 1 and Figures 1 through 13. The data were obtained by taking replicate cores in each tank, creating a composite sample for each core through a homogenization and sampling process, and conducting duplicate analyses on each composite sample. Thus, the statistical model in this case was:

$$y_{ijk} = \mu + L_i + S_{ij} + A_{ijk}$$

where S_{ij} represents the replicate composite (sub-sample) effect and the other terms are the same as defined above. The vertical effect term was not used in this model because of the compositing process.

Table 1 shows the estimated tank means and relative standard deviations (RSDs) for each source of uncertainty. The combined RSD (for the mean) is also shown. In Figures 1 through 13, the contribution of spatial, compositing (sub-sampling), and analytical uncertainties to the total variability in the data are shown. In general, the spatial uncertainty is the dominant factor affecting the total uncertainty (most of the plots are red). Analytical uncertainty is dominant for some constituents in some tanks. Differences between duplicate composite samples appears to be a minor source of uncertainty with a few exceptions (tanks S-104 and T-104).

These examples illustrate how statistical methods can be used to isolate and quantify uncertainties from various sources. More thorough descriptions of statistical modeling are found in several tank characterization reports and in (Hartley 1995). Such estimates of uncertainty are required to determine the sampling requirements for obtaining data that will ensure confident decisions.

3.3 Evaluating Relative Merit of Overcoming Sampling Issues

Several approaches to overcoming sampling issues were discussed above. There is a cost associated with reducing or bounding uncertainty and with control and intervention. The SLDs depict data or information needs for determining risk and achieving an acceptable risk. Uncertainty/sensitivity analyses should be performed using the SLDs as a basis for uncertainty propagation. The uncertainty/sensitivity analysis would show which data inputs and uncertainties significantly affect both the magnitude of the risk estimate and the uncertainty associated with that estimate. Uncertainty analysis is discussed in greater detail in another section of this report.

The uncertainty/sensitivity analysis should be used to determine whether efforts to overcome sampling issues and obtain "better" estimates will significantly reduce our risk estimate or significantly improve our confidence in that estimate. Without this uncertainty analysis, it is nearly impossible to answer the question of how many samples are required to provide sufficient information to make confident decisions. The effect of uncertainty on the probability of making incorrect decisions must be determined.

Finally, value of information (VOI) methods should be employed to evaluate the tradeoffs between sampling, control, and intervention. In some cases, the cost of obtaining information from samples that is adequate for confident decision-making may be too expensive when compared with control or intervention strategies. In other cases, the cost and benefit may favor more sampling. Without some

value of information study, it is impossible to determine which approach is optimal. The risk-based, VOI approach is also discussed in greater detail elsewhere in this report.

For both the sensitivity and VOI analyses, some baseline estimates or model of spatial and analytical variations is needed. Some baseline reference for these uncertainties is needed in order to compare the relative value of decreasing uncertainty versus control or intervention.

3.4 Sampling Issues Summary

Before one can answer the question of the number of samples required to make confident decisions, several issues dealing with sample representativeness and tolerance for making incorrect decisions must be addressed. Sample representativeness issues that must be resolved include whether the material sampled underneath a riser is relatively representative and whether samplers are recovering material from all layers in the tank. Some options for overcoming these sampling obstacles and statistical methods for estimating uncertainties have been presented.

The value of reducing uncertainty in risk estimates through sampling must be contrasted against intervention or control options. If after such an analysis the conclusion is that confident decision making for safety issues based on sample results is not feasible, then sampling should be focused on satisfying other users data needs. Uncertainty analysis coupled with a risk-based VOI approach would provide the insight and tools required for conducting these tradeoffs.

Answers to the questions dealing with the number of samples required and whether sampling will provide useful information can not be answered through a simple calculation or approach. Although it appears that the new characterization approach is complicated, the uncertainty analysis and VOI studies can be performed and the issues associated with sampling and analysis can be evaluated, resolved, or eliminated.

4.0 Reliability Methods

Reliability methods are often employed to determine probabilities of the failure. The SLDs are similar to failure trees but instead are success trees. Reliability tools can be used in conjunction with SLDs and uncertainty analysis methods to allow for a better direct visualization of a failure eliminates while providing for a systematic examination of all conceivable ways in which failure has to be eliminated. This section shows how reliability analysis methods can be used in the large-scale tanks problem.

4.1 Reliability and Logical Completeness

The SLD And Logical Completeness

A major issue in reliability engineering is the establishment of standard methods to ensure technical basis design reviews which provide a systematic examination of all conceivable safety failures, both normal and off-normal. The SLD template is the standard tool used to ensure all conceivable causes of tank safety failure have been eliminated or their probability brought into an acceptable range. Specifically, the SLD is a success tree, and a success tree analysis will be used to guarantee that all potential safety failures have been eliminated.^a The fault tree analysis(FTA) and its dual, the success tree analysis, "in all cases ... allow for a better direct visibility of a failure cause"^b

given a failure effect^c than does a tabular FMEA(failure modes and effects analysis).^d Note that the FMEA is "basically a design review process that provides for a systematic examination of all conceivable failures (or of all conceivable ways to eliminate failures and achieve success).^e

We added the italics, bolding font, and underline for emphasis.

To repeat the above quote and to rephrase it plainly in its dual form for success tree analysis: **the success tree analysis, in all cases, allows for a better direct visibility of a failure elimination while providing for a systematic examination of all conceivable ways in which failure has to be eliminated.**

Note that "better direct visibility of" in the above quote means better than could be provided by a FMEA. Since problem visibility (i.e., understanding all of the sub-problems of the composite problem that must be solved to achieve a satisfactory level of tank safety issue resolution) is a major concern, it was clear to our team that the success tree analysis is the preferred reliability engineering method for analyzing tank safety issues.^f The success tree analysis used for the tank safety issue problem is referred to in this report as a structured logic diagram (SLD). An SLD

---

^a Normally developed in reliability engineering as a Fault Tree Analysis, FTA.

^b in the case of the STA, success or a safety failure elimination.

^c in the case of the STA, an eliminated failure effect.

^d Klion, Jerome, Practical Electronic Reliability Engineering, Van Nostrand Rheinhold, NY, 1992, p 417.

^e Ibid., p 405.

^f See also Vesely, W. E., Goldberg, F. F., Roberts, N. H., and Haasi, D. F., Fault Tree Handbook, US Nuclear Regulatory Commission, New York, 1981.

will be provided for each major phenomenology. The detailed description for each SLD will be in separate reports. Each tank safety issue SLD will also be referred to at times as a major sub-problem set.

The SLD not only provides greater visibility for the sub-problems of the large-scale composite problem, it also can act as a type of proof that coverage of the domain is complete.^a That proof is the result of the fact that each SLD is composed of a union of cut sets. A cut set is a collection of basic events (individual sub-problems) that, when solved, guarantee the occurrence of the top event.^b

In every case, the top event chosen for each SLD is acceptance of the risk of a safety issue. All basic events describing initiating factors, conditioning event scenarios, primary event scenarios, secondary event scenarios, release scenarios, possible source term configurations and amounts, pathway scenarios, and uptake and consequence scenarios are included in groups of cut sets. The secret of the power of the SLD, as it is for the FTA and the FMEA, is simple; it makes explicit and plainly visible all possible cut sets that contribute to the success of the top event. Confidence in the SLD comes from the fact that the SLD must be prepared by experts of the phenomenology and that each cut set must reflect information based upon **an accurate, complete, and expert domain analysis**.

Because the SLD makes the completeness and adequacy of this domain analysis immediately visible to any expert reviewer or peer, omissions, incorrect assumptions, and incomplete logic pathways can be quickly discovered and corrected. If the SLD is made part of a stable technical basis, it matures with rapid elimination of error and unsubstantiated expertise. The proof then comes about through the construction of cut sets leading to acceptance of risk due to problem solving steps in pathways based upon the expertly established domain for the phenomenology.

The SLD can also be used to determine logical problem completeness in the sense of its coverage of all possible problem molecules and their associated phenomenology. This can happen in several ways. A peer review of the tank contents can hypothesize dimerization of all possible combinations (i.e., power set of the tank contents at the molecular or chemically active configuration level). This is the Leon Stock model for discovery of surprise molecules. This model for discovery of surprise molecules was proposed by Leon Stock University of Chicago, retired, at a tank waste characterization session.

Any hypothesized molecule that appears to be a feasible candidate will require sampling, probably of the head space gases for trace indicator molecules which can only be present if the events producing the surprise molecule actually took place. If a surprise molecule is found and substantiated, either an SLD will be developed for its phenomenology or it will be incorporated into one of the existing SLDs.

The following chart shows how this could happen. The chart summarizes the major sub-problem sets (each set is a large-scale composite problem). Each major sub-problem set contains all of

---

^a See also Johnson, W. G., The Management Oversight and Risk Tree-MORT, US Atomic Energy Commission, Washington, D. C., 1973.

^b See also Henley, Ernest J., and Kumamoto, Hiromitsu, Probabilistic Risk Assessment, Reliability Engineering and Analysis, IEEE Press, New York, 1991.

the sub-problems required to be solved in order to fully characterize the indicated tank safety issue and to finally characterize that safety issue as having acceptable or zero risk.

Note that the major sub-problem set Acceptance of P(New) Risk Objective does not exist now. It will only come about through exercise of the Expert Round Table Review (Leon Stock model) or the discovery or hypothecation of a surprise molecule. In the event that a surprise (fuel or explosive molecule) is detected and it cannot be integrated into one of the other major five sub-problem sets, addition of the new SLD, Acceptance of P(New) Risk Objective will complete the tank safety analysis. It is clear that, taken together, the SLDs constitute a technical basis for tank safety issues. This technical basis becomes the Maintenance Process for Acceptance of Technical Basis Risk, indicated below. Changes to the technical basis must be either an extension in terms of a new SLD or must be integrated into an existing SLD.

Figure 4.1 Integrated Technical Acceptance Process

4.2 Large-Scale, Composite Problem Resolution

One of the major drawbacks of tank safety issue work to date is that solutions to problems have been proposed at the macro level without considering the number, interrelationship, and depth of the underlying sub-problems that must be addressed to fully solve the larger problem. Macro-level problems have been usually focused on a tank while an understanding of a solution for a specific tank had to depend upon understanding the generic safety issue across all tanks. That understanding requirement has led to the SLD, a full exposure and identification of all the subproblems which must be addressed to resolve safety issues associated with a specific tank.. The real power of the SLD is to, not only identify the depth and complexity of sub-problems, but, to establish a rigorous, systematic process to eliminate, reformulate, and solve those sub-problems that are essential for achieving a final solution of the larger-scale composite tank safety problem. Our use of the SLD is especially powerful for decomposing, reformulating, and analyzing the large-scale, composite problems posed by tank safety issues. The first result of defining an SLD for any one of the safety issues is to become immediately aware of the fact that:

Each tank safety issue is a composite problem with a large number of sub-problems. At the first in-process review for this effort, it was clear that each safety issue was composed of many

individual sub-problems of varying degrees of difficulty. The initial accomplishment of identifying each sub-problem was a significant step in understanding and coming to grips with the real nature of the tank safety issue task. Each SLD reveals many sub-problems (see table below) that must be reformulated, eliminated, or solved in order to achieve characterization. This fact bears highlighting:

Each SLD represents a major safety issue sub-problem set that contains a large number of individual sub-problems, typically 100-150. The reformulation and simplification of the large-scale composite tank safety problem must begin with the reformulation and solution or else the simplification and elimination of these individual sub-problems.

For each of these major safety related sub-problem sets (SLDs), the number of individual sub-problems to be worked, models to be developed or applied, and the number of sub-problem types and their interrelationships to one another must be determined. This is the first step in the large-scale problem solving methodology introduced at the Centro di Teoria dei Sistemi (CNR).^a The principal tool of the methodology is the extended problem interaction graph. The full extended problem interaction graph for the ferrocyanide sub-problem set is shown below. The sub-problems are shown in abbreviated, numbered form to try to get the whole extended problem interaction graph on a single page. Even then some of the sub-problems on the right are chopped off. This extended problem interaction graph has 154 individual sub-problems. Note how neatly the SLD lent itself to this view of the ferrocyanide tank safety issue. The extended problem interaction graph for ferrocyanide was achieved by simply turning the SLD on its side, with all of its inputs on the left and the single output on the right. Viewed in this way the problem is a multiple input, single output (MISO) system. The multiple inputs are the data requirements essential for problem resolution. The single output is the top event of the SLD, acceptability of the risk of the tank safety issue.

---

^a Guardabassi, G. Large-Scale Problem Analysis and Decomposition Theory, Journal of the Franklin Institute, Vol 306, No 1, July 1978, Pergamon Press, Ltd., Northern Ireland.

Figure 4.2 Full Extended Problem Interaction Graph for the Ferrocyanide Sub-Problem Set

With the SLD now turned on its side to illustrate each tank safety issue as a composite problem, one can now begin to see the full magnitude of the composite tank safety issue problem. One also sees the interrelationship between each of the individual sub-problems. Preparation of the extended problem interaction graph is the first step in large scale problem decomposition and follows automatically from the SLD.

The extended problem interaction graph exposes the large number of sub-problems to be resolved and shows their interaction. It is evident that the larger problem solution will come through individual sub-problem resolution or elimination. The large-scale problem reformulation must first come through systematic review and reformulation of problems at the individual sub-problem level. This should be an iterative effort structured in general to 1) eliminate sub-problems with adequate solutions already in the Hanford literature, 2) combine sub-problems across major subproblem sets, 3) combine where possible data collection and analysis efforts, 4) use ancillary methods outside the technical basis such as from value of information theory to help further eliminate data requirements, off-normal event paths, primary reaction events, and other problems in each major sub-problem set which can be shown to be incredible through additional outside analysis. Again, it bears repeating, that **the SLD acting as the primary reliability engineering tool, for both evaluating cut sets and for tracking sub-problem interaction, will ensure that the overall problem reformulation takes place in a systematic and rigorous way.**

Each pruning of problem out of the problem set not only eliminates the requirement to address that problem, but also ensures that any future proposal to solve the eliminated problem is also eliminated as potential requirement. Knowing which problems we do not want to solve in this large-scale composite problem is as important as knowing the ones that must be solved. The same is true for probability densities or distributions. Since the SLD lays out which specific probability densities are important, uncertainty is bounded in the SLD framework in the sense that we know exactly which probabilities to address and which ones to ignore. This bounding by elimination and focus on what counts, is true for the SLD framework or its equivalent representations, more so than for any other approach.

Figure 4.3 The Extended Problem Interaction Graph for Criticality

Because of this and because the SLD framework lays out a complete problem specification, a problem simulation that collapses a large number of the sub-problems and most of the phenomenological models into single simulations is possible. This means that a wider variety of phenomenological states can be evaluated than can be obtained by running all the model sub-problems one at a time.

Fortunately, even though each phenomenology has a large number of sub-problems and uses a large number of models (FeCn--21), we can use an off-the-shelf complex modeling tool such as MATLAB with Simulink to dramatically reduce the number of sub-problems (FeCn--154) and the complexity of the extended problem interaction graph. The extended problem interaction graph for criticality, shown above on the preceding page, has over 189 sub-problems. There are actually more sub-problems than the count from the extended problem interaction graph because of the extensive sharing of data domains by different parts of the criticality graph. While the convoluted structure of this graph, makes its appearance as an extended problem interaction graph very busy, it is probably a good candidate for rapid elimination of sub-problems and easy consolidation of the overall composite problem. Also note the display of both the ferrocyanide and criticality extended problem interaction graphs is merely to reveal the large number of sub-problems and their interaction. The graphs will be brought under control of a matrix processor such as MATLAB. Each sub-problem, each link between problems will be exactly and precisely addressable in a large matrix problem format.

For example, we examine a piece of the criticality extended problem interaction graph. Input data for determination of the chemical form of the waste is

INITCHEMCOND =

[DMASS, DVOL, DTHERMO, DKINETIC, DCHEMCOMP, DSOLAB, DOXREDOX, DTEMP, DSOLSPEC]

and input data for the Geometric Location of the waste is

INITPHYCOND =

[DTEMP, DPOROSITY, DMASS, DVOL, DFLUIDDYNDATA, DVISCOSITY, DTORTUOSITY, DPARTSIZE, DDENSITY, DMECHENERGYDATA, DENTHALPY, DDIFFCOEF, DOTHERCONDUCT].

These are actual symbolic Criticality data input vectors taken from MATLAB. These and all other data requirements vectors will be processed by MATLAB, maintaining the vector structure of the data and processing each model as a transfer function, or as a matrix transformation on the input data. Uncertainty information and error estimates can be added to each of the input data components, to all process steps and to each model output. In this way, all estimates are carried automatically by the MATLAB processor throughout the entire SLD structure and the combined effect of uncertainty is known quickly for every change of estimate or recomputation. One could have developed a spreadsheet model to do the same thing, but spreadsheets are 1980s technology and truly cumbersome compared to MATLAB. The entire problem graph can be entered into MATLAB very quickly and processing begun immediately once all the physical relationships and model structures are specified. The vector product of

CHCOMP = INITCHEMCOND*Model1

is shown next. CHCOMP is the result of transforming the input vector INITCHEMCOND with the transformation array from Model1(M1). The MODEL1 array

MODEL1 =

```
[M11,M12,M13,M14,M15,M16,M17,M18,M19]
[M21,M22,M23,M24,M25,M26,M27,M28,M29]
[M31,M32,M33,M34,M35,M36,M37,M38,M39]
[M41,M42,M43,M44,M45,M46,M47,M48,M49]
[M51,M52,M53,M54,M55,M56,M57,M58,M59]
[M61,M62,M63,M64,M65,M66,M67,M68,M69]
[M71,M72,M73,M74,M75,M76,M77,M78,M79]
[M81,M82,M83,M84,M85,M86,M87,M88,M89]
[M91,M92,M93,M94,M95,M96,M97,M98,M99]
```

and the actual MATLAB symbolic multiplication is

CHCOMP = SYMMUL(INITCHEMCOND,MODEL1)

ans =

```
[DMASS*M11+DVOL*M21+DHERMO*M31+DKINETIC*M41+DCHEMCOM
P*M51+DSOLAB*M61+DOXREDOX*M71+DTEMP*M81+DSOLSPEC*M91,
M12*DMASS+DVOL*M22+DHERMO*M32+DKINETIC*M42+DCHEMCOMP
*M52+DSOLAB*M62+DOXREDOX*M72+DTEMP*M82+DSOLSPEC*M92,
M13*DMASS+DVOL*M23+DHERMO*M33+DKINETIC*M43+DCHEMCOMP
*M53+DSOLAB*M63+DOXREDOX*M73+DTEMP*M83+DSOLSPEC*M93,
M14*DMASS+DVOL*M24+DHERMO*M34+DKINETIC*M44+DCHEMCOMP
*M54+DSOLAB*M64+DOXREDOX*M74+DTEMP*M84+DSOLSPEC*M94,
M15*DMASS+DVOL*M25+DHERMO*M35+DKINETIC*M45+DCHEMCOMP
*M55+DSOLAB*M65+DOXREDOX*M75+DTEMP*M85+DSOLSPEC*M95,
M16*DMASS+DVOL*M26+DHERMO*M36+DKINETIC*M46+DCHEMCOMP
*M56+DSOLAB*M66+DOXREDOX*M76+DTEMP*M86+DSOLSPEC*M96,
M17*DMASS+DVOL*M27+DHERMO*M37+DKINETIC*M47+DCHEMCOMP
*M57+DSOLAB*M67+DOXREDOX*M77+DTEMP*M87+DSOLSPEC*M97,
M18*DMASS+DVOL*M28+DHERMO*M38+DKINETIC*M48+DCHEMCOMP
*M58+DSOLAB*M68+DOXREDOX*M78+DTEMP*M88+DSOLSPEC*M98,
M19*DMASS+DVOL*M29+DHERMO*M39+DKINETIC*M49+DCHEMCOMP
*M59+DSOLAB*M69+DOXREDOX*M79+DTEMP*M89+DSOLSPEC*M99]
```

MODEL1 is postulated to have a 9x9 transformation array. The actual array to be used will have fewer components since most physical models do not use every input parameter for every calculation. We can propagate an error through the model simply by creating the error vector e, where ei is the estimated error in the ith row of MODEL1.

We have for MODEL1 estimated errors, using a simple vector error model:

e = [e1,e2,e3,e4,e5,e6,e7,e8,e9]

Propagating this to CHCOMP, we have CHCOMP = symmul(e,MODEL1):

ans =

```
[e1*M11+e2*M21+e3*M31+e4*M41+e5*M51+e6*M61+e7*M71+e8*M81+e9*M91,  
e1*M12+e2*M22+e3*M32+e4*M42+e5*M52+e6*M62+e7*M72+e8*M82+e9*M92,  
e1*M13+e2*M23+e3*M33+e4*M43+e5*M53+e6*M63+e7*M73+e8*M83+e9*M93,  
e1*M14+e2*M24+e3*M34+e4*M44+e5*M54+e6*M64+e7*M74+e8*M84+e9*M94,  
e1*M15+e2*M25+e3*M35+e4*M45+e5*M55+e6*M65+e7*M75+e8*M85+e9*M95,  
e1*M16+e2*M26+e3*M36+e4*M46+e5*M56+e6*M66+e7*M76+e8*M86+e9*M96,  
e1*M17+e2*M27+e3*M37+e4*M47+e5*M57+e6*M67+e7*M77+e8*M87+e9*M97,  
e1*M18+e2*M28+e3*M38+e4*M48+e5*M58+e6*M68+e7*M78+e8*M88+e9*M98,  
e1*M19+e2*M29+e3*M39+e4*M49+e5*M59+e6*M69+e7*M79+e8*M89+e9*M99]
```

Note that more sophisticated error models than the simplistic one shown above can be quickly generated and incorporated into the MATLAB solution.

The matrix and vector handling off-the-shelf machinery is clearly here and available today to process, efficiently and cost-effectively, the multiple subproblems of the composite problem posed by each SLD. Before any of this work can be done, the next steps in a composite problem solving methodology must be executed. Since the overall problem has in effect been fully defined by the SLDs, and their extended problem interaction graphs, the next steps are:

- **specify the domain and co-domain of each sub-problem**
- **specify the mapping function represented by each sub-problem**
- **evaluate the structure for problem dominance, redundancy, and regions of infeasibility.**

The first bullet above, specify the domain and codomain of each sub-problem, is really the very next crucial step. Notice in the example given previously for uncertainty propagation, that the inputs and outputs for all boxes in the SLD had to be defined. Also the transformation of that data from input to output, or the mapping represented by each box, had to be defined. Most of this domain specification is yet to be done for each of the SLDs.

5.0 Acknowledgments

Key contributors to this report include Al Liebetrau, Brent Pulsipher, Dan Kashparenko, Feng Gao, and Guang Chen and Stacey Hartley. Al Liebetrau provided the sensitivity and uncertainty analysis perspective and Dan Kashparenko provided the systems analysis viewpoint. Brent Pulsipher developed the sampling issues section. In addition to significant contributions to the technical discussions surrounding this report, Feng Gao, Guang Chen, and Stacey Hartley provided the figures, the examples, and much of the proofreading. Larry Morgan participated in key formative discussions on this topic and supported the development of the proposed approach.

6.0 Bibliography

The following bibliography is provided for several reasons. First, it provided to demonstrate that the approach proposed here is practical and achievable, and not just a theoretical "pie-in-the-sky" construct. The ideas developed here are based on well-known and sound analytical principles, and the computations are tractable and have been used in real applications. Second, the bibliography is intended to demonstrated that PNNL has expertise in the application of methods for sensitivity, uncertainty, and systems analysis. Finally, the bibliography is provided for those who want to pursue the literature surrounding the proposed approach in some detail.

Bowen, M.W. and C.A. Bennett. 1988. Statistical Methods for Nuclear Material Management, PNL-5849, Richland, Wa.

Caselton, W.F. and Zidek, J.V. [1984], "Optimal Monitoring Network Designs," *Statistics and Probability Letters*, 2, 223-227.

Caselton, W.F., Kan, L., and Zidek, J.V. [1992], "Quality Data Networks that Minimize Entropy," in *Statistics in the Environmental and Earth Sciences* (P. Guttorp and A. Walden, eds.), Griffen, London.

Chambers, J.M. and Hastie, T.J. [1993], *Statistical Models in S*, Chapman and Hall, New York.

Currin, C., Mitchell, T.J., Morris, M.D., and Ylvisaker, D. [1991], "Bayesian Prediction of Deterministic Functions With Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, 86, 953-963.

Farris, W.T., Napier, B.A., Eslinger, P.W., Ikenberry, T.A., Shieler, D.B., and Simpson, J.C., [1994], *Atmospheric Pathway Dosimetry Report, 1944-1992*, PNWD-228 HEDR, Battelle Pacific Northwest Laboratories, Richland, WA.

Guardabassi, G., [1978], "Large-Scale Problem Analysis and Decomposition Theory," *Journal of the Franklin Institute*, Vol. 306, No. 1, July 1978, Pergamon Press, Ltd., Northern Ireland.

Hanselman, Duane and Littlefield, B., *MATLAB, Users Guide*, Prentice Hall, Englewood Cliffs, NJ 07632.

Hartley, S.A., K.M. Remund, B.C. Simpson, R.D. Cromar, C.M. Anderson. 1995. "Statistical Supplement to the Tank Characterization Reference Guide", WHC-SD-WM-TI-648, Prepared for Westinghouse Hanford Company by Pacific Northwest Laboratory, Richland, Wa.

Heasler, P.G, C.M Anderson, D.B. Baird, R.J. Serne, P.D. Whitney. 1993. "Statistical Evaluation of Core Samples From Hanford Tank B110", PNL-8745, Pacific Northwest Laboratory, Richland, Wa.

- Henley, Ernest J., and Kumamoto, Hiromitsu, [1991], *Probabilistic Risk Assessment, Reliability Engineering and Analysis*, IEEE Press, New York.
- Iman, R.L. and Conover, W.J. [1980], "Small Sample Sensitivity Analysis Techniques for Computer Models, with an Application to risk Assessment," *Communications in Statistical Theory and Methods*, A9(17), 1749-1842.
- Jensen, L., R.D. Cromar, S. Wilmarth, P.G. Heasler. 1995. "Number of Core Samples: Mean Concentrations and Confidence Intervals", WHC-SD-WM-TI-674, Westinghouse Hanford Company, Richland, Wa.
- Johnson, W.G., [1973], *The Management Oversight and Risk Tree-MORT*, US Atomic Energy Commission, Washington, D.C.
- Klion, Jerome, [1992], *Practical Electronic Reliability Engineering*, Van Nostrand Rheinhold, NY.
- Liebetrau, A.M., Neerchal, N.K., and Gao, F. [1995], "On Optimal Design of Computer Experiments," presented at MODSIM 95. In proceedings of the International Congress on Modeling and Simulation, Newcastle, NSW, Australia, Nov.27-30, 1995.
- Liebetrau, A.M., Neerchal, N.K., Gao, F. [1995], "Uncertainty Analysis for Systems of Linked Computer Models," Presented at EnvironMetrics 95. Sixth International conference on Environmetrics, Kuala Lumpur, Malaysia, 6-9 Dec. 1995.
- McKay, M.D., Conover, W.J., and Beckman, R.J. [1979], "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21, 239-245.
- Morris, M.D., Mitchell, T.J., and Ylvisaker, D. [1993], "Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction," *Technometrics*, 35, 243-255,
- Remund, K.M., C.M. Anderson, B.C. Simpson. March 29, 1996. "Hanford Waste Tank Grouping Study", Letter report prepared for WHC by Pacific Northwest Laboratory, Richland, Wa.
- Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H.P. [1989a], "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409-423.
- Sacks, J., Schiller, S.B., and Welch, W.J. [1989b], "Designs for Computer Experiments," *Technometrics*, 31, 41-47.
- Simulink Dynamic Simulation Software Users Guide*, The MathWorks, Inc., Natick, Massachusetts 01760.
- Vesely, W.E., Goldberg, F.F., Roberts, N.H., and Haasi, D.F., [1981], *Fault Tree Handbook*, US Nuclear Regulatory Commission, New York.

Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., and Morris, M.D. [1992],
"Screening, Predicting, and Computer Experiments," *Technometrics*, 34, 15-25.

Distribution

No. Of Copies

26 Pacific Northwest Laboratory
P.O. Box 999
Richland, WA 99352

G. Chen, K5-12
S.D. Colson, K2-14
F. Gao, K5-12
R.E. Gephart, K9-76
S.A. Hartley, K5-12
V.L. Hunter, K7-97
D.M. Kashporenko, K7-10
A.M. Liebetrau, K5-12
L.G. Morgan, K6-48
B.A. Pulsipher, K5-12 (10)
Information Release, K1-06 (7)