Risk Analysis, Vol. 19, No. 1, 1999

A Survey of Approaches for Assessing and Managing the Risk of Extremes

Vicki M. Bier,[1] Yacov Y. Haimes,[2] James H. Lambert,[2] Nicholas C. Matalas,[3] and Rae Zimmerman[4]

In this paper, we review methods for assessing and managing the risk of extreme events, where "extreme events" are defined to be rare, severe, and outside the normal range of experience of the system in question. First, we discuss several systematic approaches for identifying possible extreme events. We then discuss some issues related to risk assessment of extreme events, including what type of output is needed (e.g., a single probability vs. a probability distribution), and alternatives to the probabilistic approach. Next, we present a number of probabilistic methods. These include: guidelines for eliciting informative probability distributions from experts; maximum entropy distributions; extreme value theory; other approaches for constructing prior distributions (such as reference or noninformative priors); the use of modeling and decomposition to estimate the probability (or distribution) of interest; and bounding methods. Finally, we briefly discuss several approaches for managing the risk of extreme events, and conclude with recommendations and directions for future research.

KEY WORDS: Extreme events; risk assessment; risk management; extreme value theory; judgmental distributions.

1. INTRODUCTION

With technological progress, as ever-increasing amounts of energy are harnessed, there is an increasing need for methods for assessing and managing the risks of extreme events, which may be different from those used for less extreme events. This paper reviews some approaches for assessing and managing the risks of extreme events when we don't know much. However, this review should not be taken as being definitive; the intent is to suggest possible approaches, rather than to provide answers. We believe that further research and data collection on this important issue is needed.

[1] Department of Industrial Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706.
[2] Center for Risk Management of Engineering Systems, University of Virginia, Charlottesville, Virginia 22903.
[3] Hydrologist, Sterling, Virginia 20165.
[4] Robert Wagner Graduate School of Public Service, New York University, New York, New York 10003.

2. WHAT IS AN EXTREME EVENT?

Several methods for defining extreme events have been proposed. One proposal is to define extreme events in terms of their low *frequency*. However, this seems inappropriate. For example, a room temperature of exactly 72.34571° F may occur extremely rarely, but would hardly seem to be extreme by most people's definitions. Therefore, it would seem more reasonable to define extreme events as being extreme also in terms of their *severity*; unusually severe events tend to occur infrequently. Accordingly, extreme events will be defined throughout the

0272-4332/99/0200-0083$16.00/1

remainder of this paper to be events that are both rare and severe.

Other characteristics may be relevant, either in *defining* what we mean by an extreme event, or in explaining how such extreme consequences might arise. For example, extreme events are not only severe, but also outside the normal range of experience of the system in question. An event that is considered extreme in one system may be normal or average in another system. Also, while extreme events are often catastrophic, this need not necessarily be the case; events that are outside the normal range of experience of a system may be considered extreme even if they are not catastrophic. Slovic *et al.*[1] have pointed out that events considered to be extreme often share other characteristics in addition to low frequency and severe or catastrophic consequences, such as high degrees of dread and uncertainty, and the involuntary nature of many extreme events.

Extreme events are often caused by non-linear phenomena, such that a relatively small change in some causal factor can lead to a large increase in the severity of the event. One example of such nonlinearity would be threshold effects such as core melt or dam failure, where a small increase in some relevant parameter (e.g., core temperature, flood height) can result in a sudden transition from a near-accident situation to a severe accident. Additive or synergistic effects can also affect the frequency of extreme events, especially when system behavior is nonlinear. For example, normal flooding can become escalated to extreme levels due to a pile-up of nonextreme events such as soil saturation, heavy rains, etc.[2] Similarly, in the area of industrial safety, extreme accidents can result from the additive effects of less extreme problems in system design, maintenance, operation, and oversight.[3]

3. HOW CAN WE IDENTIFY POTENTIAL EXTREME EVENTS?

In some cases, risk analysts may be asked to analyze the frequency and/or severity of one or more specific classes of extreme events that have already been identified by the client or sponsor of the analysis, such as floods, earthquakes, airplane crashes, or explosions. At other times, the task of the analyst may be to identify and analyze *all* classes of events that could potentially lead to severe consequences in a particular context. In this case, in which the potential extreme events have not yet been identified, and possibly not even imagined, a systematic approach to identifying potential extreme events can be extremely helpful.

Several such approaches could be considered. For example, hazard and operability studies (HAZOPS)[4] are essentially inductive in nature (as opposed to the deductive processes involved in failure modes and effects analysis). Therefore, one possible strategy for ensuring that possible extreme events have been identified might be some type of customized HAZOPS procedure. The taxonomy of "surprises" developed by Fiering and Kindler[5] (see Sect. 5) provides another approach for identifying possible categories of extreme events.

Similarly, in hierarchical holographic modeling,[6] multiple models are developed to represent the various aspects of the system being analyzed. For example, for a civil infrastructure project, the sources of risk can be decomposed into functional, spatial/geographical, temporal/planning, and user/stakeholder aspects. Each of these sources can then be further decomposed in a hierarchical manner (e.g., sources of risk stemming from the functional aspects of the infrastructure can be subdivided into those related to life-critical, mission-critical, and convenience functions). Since each of the resulting models is in a sense a complete view of the system from a particular perspective, they are called "holographic" submodels. This approach has been used to help ensure completeness of failure mode identification in areas as diverse as global sustainable development[7] and software project development.[8]

4. WHAT PROBABILITY (OR DISTRIBUTION) DO WE WANT?

The concern of this workshop is "when and how you can specify a probability *distribution* when you don't know much" (emphasis added). However, in some cases only the probability of a single discrete event is needed (rather than a probability distribution over the magnitude or severity of the event, for example). Often, the event of interest can be described in terms of a threshold with respect to one or more parameters, in which case its probability can be defined as a tail probability of the distribution for the parameter(s). For example, the probability of bankruptcy can be viewed as a tail probability in a distribution for financial losses; if we know that a loss is severe enough to result in bankruptcy, we may not care very much *how* severe it is. Similarly, the probability of core melt in a nuclear power plant can be viewed as a tail probability in the distribution for reactor core temperature; if the temperature is hot enough to melt the core, we may not care very much *how* hot it is. Finally, the probability for a particular mode of dam failure can be viewed as a tail probability in the distribution for flood height. Once we know that the water is high enough to cause failure of the dam, the consequences of the dam failure may not depend very much on exactly *how* high it is.

In other circumstances, where threshold effects are nonexistent or small, the entire distribution of the event magnitude or severity (e.g., financial loss, temperature, or flood height) may be of interest. The conditional distribution (given that the severity exceeds some threshold that is sufficient to result in damage) may also be of interest. While assessment of extremely small tail probabilities is already a difficult task, the task of assessing an entire (possibly conditional) distribution is even more demanding of the analyst. Greater detail in the results is needed in this case, and the analyst may be more constrained in how he or she can transform or decompose the problem. For example, even though core melt probability can be viewed as a tail probability in the distribution of reactor core temperature, this is not usually how it is modeled and assessed. By contrast, if we required the entire distribution of core temperature, we may have little choice but to model the temperature explicitly.

Similarly, Katz and Brown[9] distinguish between situations in which *any* event exceeding a threshold is of interest, and situations in which only the *maximum* event exceeding the threshold is relevant. For example, in analyses of wave heights, there may be situations in which the damage is well represented as a function of the *maximum* wave height, regardless of how many waves at or near that height may have occurred.

5. IS A PROBABILISTIC APPROACH EVEN PRACTICAL?

In some cases, it is arguable that the problem at hand may be too information-sparse to handle in an explicitly probabilistic manner. (Note, however, that this is a controversial view; many Bayesian statisticians believe that all problems can be addressed probabilistically through reliance on what little information is available and the concept of "degree of belief" or subjective judgment.) Therefore, it may be worth reviewing other potentially useful approaches for analyzing extreme events. A few such approaches are discussed briefly.

One set of approaches to be considered might be described as semiprobabilistic or quasiprobabilistic methods, such as fuzzy set theory,[10, 11] Dempster-Shafer theory,[12, 13] or the theory of interval-valued probabilities.[14] Ferson[15] and Ferson and Ginzburg[16] have argued that such methods relax some of the axioms of probability theory, and hence may be applicable when it is not clear that those axioms are reasonable for modeling the problem at hand. Thus, for example, fuzzy set theory can be used to represent concepts such as what we mean by a "tropical" climate, which are not obviously probabilistic.[17] Alternatively, methods such as fuzzy set theory and interval analysis can be viewed as giving bounds (although in some cases quite loose bounds) on the results of a full probabilistic analysis when precise estimates of numerical probabilities and dependencies are unreliable or unavailable. Such bounding methods may therefore be more appropriate than full probabilistic analyses that may give a false appearance of precision; see, for example, Moore and Ferson[18] for an application of bounding methods to modeling of environmental exposures to hexachlorobenzene.

Next, some nonprobabilistic but quantitative approaches, such as chaos and catastrophe theories, may provide insights about particular classes of extreme events. Chaos theory is concerned in part with understanding "the complex boundaries between orderly and chaotic behavior."[19] Since some extreme events may be characterized as transitions from orderly to chaotic behavior, or for that matter *vice versa* (e.g., overheating may in some cases be associated with transitions from turbulence to more orderly patterns of coolant flow), methods that allow us to explicitly characterize such phase transitions may be desirable. Similarly, catastrophe theory is concerned with "*sudden* changes caused by *smooth* alterations in the situation"[20] (emphasis in original). Although not specifically intended for the study of disasters (despite the connotations of its name), this approach may nonetheless be well suited to studying those disasters and other extreme events that are caused by nonlinear system behavior. Study of catastrophes and phase transitions is potentially useful in the analysis of global climate change, for example, since qualitative changes in climate patterns are of central concern.

Finally, there are qualitative approaches to the analysis of extreme events that may provide useful insights, either in place of quantitative methods or as a supplement to them. For example, some have attempted to develop a theory of "surprise." Even though some of this work was not concerned with extreme events *per se,* it might well be relevant to defining and assessing the probabilities of extreme events. For example, Shackle[21] distinguishes between "counterexpected events" and "unexpected events." Here, counterexpected events are events that had been imagined, but consciously rejected as impossible; by contrast, unexpected events are events that had never been imagined. Occurrences of either type of event would generally result in surprise, and both counterexpected and unexpected events will often tend to be extreme (either in the sense of severity of consequences, or in the sense of being outside the normal range of experience of the system in question). Similarly, Fiering and Kindler[5] developed a taxonomy of counterexpected surprises that can affect the operation of water resource systems. These include structural collapse of a system component, errors in system design, changes in the hydrologic regime, new laws or environmental standards, surprises resulting from lack of understanding of how a system will respond to particular conditions, and demographic changes. Such taxonomies, while not directly answering the question of how we can assess the probabilities of extreme events, can be helpful, for example, in identifying which types of extreme events we should consider and thereby hopefully minimizing the number of "unexpected" (as opposed to "counterexpected") events.

6. IF SO, HOW CAN WE SPECIFY THE DESIRED PROBABILITY (OR DISTRIBUTION)?

Typically, we will have little or no data directly relevant to the question at hand. Therefore, most approaches to specifying the probabilities of extreme events rely on Bayesian rather than classical statistical methods. Bayesian methods become both more desirable and more problematic than classical methods when data are sparse. On the one hand, unlike classical statistical procedures, Bayesian analysis can be used in situations of sparse data, since important subjective and other nonstatistical types of evidence can be used in the decision or inference process in the form of an informative judgmental distribution. However, with sparse data, the results of Bayesian analyses become more sensitive to the analyst's interpretation of the evidence (e.g., in the choice of a judgmental distribution to represent nonstastical evidence), and hence more subjective and less readily accepted. The subjective nature of judgmental distributions is particularly problematic when assessing small tail probabilities, since behavioral research shows that people are poor at assessing small probabilities,[22] so there may be little reason to expect that people's judgmental distributions will be very reliable or reproducible in the extreme tails.

For example, tail probabilities are extremely sensitive to small changes in higher moments such as skewness and kurtosis. However, such higher moments are notoriously difficult to estimate. Even means and variances can be difficult to estimate; thus, Mosteller and Tukey[23] point out that "a few straggling values scattered far from the bulk of the measurements can, for example, alter a sample mean drastically, and a sample s^2 catastrophically." For higher moments, Frey and Burmaster[24] note that the uncertainty in skewness and kurtosis can be extremely large given even reasonably good-sized datasets (e.g., 19 values, which is a larger dataset than will often be available in many problems involving extreme events).

Similarly, Hammitt and Shlyakhter[25] warn against "assigning zero probability to events that are not 'impossible,'" since many methods for combining distributions across experts have the "zero-preservation property." With these methods, if one or more experts assign probability zero to event *j,* then the probability of that event resulting from the expert-combination algorithm will also be zero. However, what appear to be negligibly small tail probabilities in a prior distribution can have large effects on the posterior distribution if the observed data are extreme or surprising. For example, Berger[26] points out that in Bayesian estimation, "If. . .the likelihood function is concentrated in the tail of the prior, there are real problems; the form of the tail is very hard to specify, and different choices can drastically change the conclusion."

A number of approaches for dealing with such problems are discussed here, roughly in decreasing order based on the amount of empirical knowledge needed to use them. For example, when enough is known about the problem to construct an appropriate informative distribution directly, then the traditional guidelines for how to elicit judgmental distributions become relevant. When the exact functional form of the distribution is not known, but the first two moments and/or the behavior of the tails can be specified reasonably accurately, then approaches such as entropy maximization and extreme value theory can be useful. Finally, when it is difficult to say very much about the functional form of the relevant distribution directly, then other approaches (such as the use of modeling and decomposition, or "robust"[26] Bayesian methods) may be the only approaches possible.

Note, however, that these approaches are not necessarily mutually exclusive, and in fact are likely to yield the best results when used in tandem. For example, even if we do not intend to explicitly assess an informative judgmental distribution, good elicitation techniques can help in obtaining reliable estimates of moments or tail probabilities, for use in methods such as entropy maximization and extreme value theory. Similarly, the order in which the various methods are presented should not be taken as a judgment on their relative desirability. Thus, for example, modeling and decomposition might be good techniques for constructing a judgmental distribution even when it is also possible to assess the desired informative distribution more directly.

6.1. Methods for Eliciting Informative Judgmental Distributions Directly

Expert opinion has been widely used in practice. For example, Cooke[27] surveys a number of applications where quantities of interest have been assessed based primarily on expert judgment, in fields ranging from aerospace to the military to probabilistic risk analysis (PRA). Within the area of PRA in particular, Mosleh *et al.*[28] discuss applications of expert opinion to the assessment of component failure rates, seismic hazards, human error rates, accident consequences, and the severity of near misses.

A number of guidelines have been suggested for structuring the process of assessing an informative judgmental distribution.[29] Of course, many of the behavioral procedures that are useful in general are likely to be especially valuable when dealing with rare and/or extreme events, where naively generated estimates may be quite unreliable. More specifically, several elicitation methods seem particularly well suited to the analysis of extreme events, as discussed below.

For example, Armstrong[30] has advised that in long-range forecasting one should "avoid hiring the most expensive expert. . ..Instead, spread the budget over a number of less expensive experts." The reason for this advice is essentially that above low levels of expertise, the improvements in accuracy as a function of level of expertise are small. The same limitations on the value of expertise are likely to apply in assessing the probabilities of rare and/or extreme events. Almost by definition, there will be little opportunity to observe and study rare events that are outside the normal range of experience of the system in question, so "experts" may have little if any more knowledge than anyone else. Thus, the practice of using *multiple* experts, rather than a single "best" expert, is likely to be beneficial.

When considering Bayesian updating of a judgmental prior distribution with observed data, pre-posterior analysis[31] is another approach for ensuring the reasonableness of the assessed prior. This method may be particularly valuable when there are concerns about possible "surprises" in the data. The idea here is to perform a number of Bayesian updates of the prior distribution with *hypothetical* datasets (including some datasets that would be considered extreme or surprising), before analyzing the actual data. This enables people to see the implications of their chosen prior distributions; if the observed behavior is found to be unreasonable, the prior distributions can be readily modified without violating the assumptions of Bayesian theory. (By contrast, modifying the prior distribution after analyzing the *actual* data is more problematic, since in this case the modified prior distribution will already reflect the evidence with which it will be combined.) Pre-posterior analysis using extreme hypothetical data is particularly useful for ensuring that the tails of an assessed prior distribution are reasonable, something that is difficult to do by other means.

Pre-posterior analysis is in line with the suggestion by Bordley[32] that people have beliefs not only about the inputs to an analysis, but also about the range of outputs that could reasonably be expected. If one's beliefs about the outputs are stronger than one's beliefs about the inputs, then substantial adjustments to the judgmental prior distribution based on pre-posterior analysis may be expected. However, pre-posterior analysis is at least somewhat controversial. For example, Hattis[33] argues that while people must specify what they know (i.e., their prior distributions), Bayes' theorem is better at expressing what effect particular datasets ought to have on the resulting posterior distribution. Therefore, he suggests that revising one's prior distribution based on knowledge of the result of a Bayesian analysis may lead to worse rather than better prior distributions in some cases.

6.2. Maximum Entropy Distributions

The general idea of maximum entropy distributions (based on information theory)[34–36] is to use whatever partial information is available about the uncertain quantity of interest (e.g., mean, median, or mean and variance), but as little additional information as possible beyond that, to avoid inadvertently assuming more than is actually known. Therefore, the maximum entropy distribution is defined to be the least informative distribution that matches the specifed constraints.[37, 38] The resulting distribution can then be used either as a prior distribution for Bayesian analysis (if additional data are available), or as a (partially) informative distribution with no updating.

In practice, it may be necessary to restrict consideration to a limited family of parametric distributions. For example, if the quantity of interest is unbounded and only its mean is known, the true maximum entropy distribution will be improper.[26] This clearly weakens the claim that one is not assuming more than is actually known, since one is unlikely to know for certain that the distribution of the quantity of interest actually falls within the specified class. However, entropy maximization (even from within a restricted class of distributions) is still philosophically appealing if the alternative is to assume a particular distributional form arbitrarily or for reasons of mathematical convenience.

The method of entropy maximization is intuitively appealing when dealing with extreme events, since the notion of information is closely related to "rarity";[35] in particular, the occurrence of a rare (and hence, in ordinary usage, "surprising") event conveys much more information than the occurrence of an event that had already been expected. However, Berger[26] identifies some weaknesses of the maximum entropy approach. In particular, he argues that "maximum entropy priors. . .may raise concerns about robustness," especially when the information on which they are based (e.g., moments) is assessed subjectively, and hence not known very accurately. In addition, Berger notes that when one of the specified moments is the prior variance, this tends to eliminate from consideration distributions with long tails, which tend to give more robust results on average.

These concerns may limit the usefulness of maximum entropy priors in assessing probabilities of extreme events directly, since accurate estimates of distribution moments are unlikely to be available and it is precisely the tail probabilities that may be of greatest interest. However, the method is still extremely useful (e.g., in selecting distributions for model inputs).[39]

6.3. Extreme Value Theory

Some problems in assessing the risks of extreme events can be formulated in terms of the *maximum* (or minimum) of a sequence of events (e.g., the most severe of a sequence of floods).[9] In this situation, extreme value theory[40–43] can provide a useful tool for assessing risk. In its simplest form, extreme value theory provides asymptotic distributions for the minimum or maximum of a series of identically distributed (and typically independent) variables as the number of such variables becomes arbitrarily large. When the number of variables is large, the distributions for the maximum and minimum depend only on the tails of the original distribution; moreover, there are only three nondegenerate limiting distributions (the Gumbel, Frechet, and Weibull extreme value distributions). Therefore, when using extreme value theory, it may not be necessary to assess an entire prior distribution.

In particular, when the necessary convergence conditions are satisfied, it becomes possible to rigorously identify a functional form for the distribution of the maximum or minimum value without specifying all aspects of the distribution for the original variables, only the class of extreme value distributions to which it converges. Note that for some initial distributions, there may be no limiting extreme value distribution; however, limiting distributions do exist for most commonly used continuous distributions.[41] For example, in the limit, the maximum values of sequences of normal, logistic, and exponential random variables will all have Gumbel distributions (although generally with different parameter values). This is because, as the number of variables grows, the distribution of their maximum depends only on the tail of the distribution of the original variables.

Because the distributions of extreme values such as maxima and minima depend only on the tail of the initial distribution, it may also be desirable to specify parent distributions that have tails with particular shapes, and in fact to specify the left and right tails independently of each other. Most methods of fitting distributions, such as the method of moments, the method of probability-weighted moments,[44] and maximum likelihood estimation, are concerned with parameter estimation. As a result, they are not very sensitive to the shapes of the distribution tails, and are not well suited to partitioning a distribution into left and right tail sections. Therefore, when it is desirable to construct a distribution with tails of particular shapes, either the Tukey distribution[45] or the Wakeby distribution,[46] which is a generalization of the Tukey distribution, might be considered. The advantage of the Tukey and Wakeby distributions is that they can be used to fit one tail almost independently of the other. This is useful when one is interested in estimated percentiles in the fitted tail(s).

Of course, before using extreme value theory in any particular application, one must verify that the model is valid in that application and that the necessary convergence conditions are (at least approximately) satisfied. Extreme value theory has been extended to account for some departures from the usual convergence conditions, including dependence in the sequence of random variables,[47] and more recently nonstationary sequences.[48] Nonstationary sequences may be important, for example, in modeling the effects of global climate change on the risk of extreme events, and can be accounted for if the trend is known or can be bounded.

In some instances, however, extreme value theory may be of little or no benefit. For example, consider the problem of estimating the probability of a 10,000-year flood (i.e., one that by definition occurs with a frequency of roughly 10^{-4} per year) occurring during the 40-year lifetime of a new nuclear power plant, based on data from the last 100 years. If the mechanisms leading to severe and more moderate floods are thought to be fundamentally the same, then one might be able to extrapolate from the past 100 years of flood data (possibly with some estimated trend in flood severity) to obtain the desired estimate. However, if the physical mechanisms that give rise to 10,000-year floods are qualitatively different from the mechanisms leading to 100-year floods, then methods such as extreme value theory that rely on extrapolation from recent experience may be inherently unreliable. In such cases, there may be no alternative other than explicitly modeling the physical processes leading to extreme floods. The distribution of flood sizes in this latter case is an example of a "contaminated" or mixture distribution. When dealing with such distributions, which are frequently bimodal, the rare or extreme process giving rise to the second mode (in the low-probability tail of the distribution) may not have been observed at all, making extreme value theory and similar extrapolation methods inapplicable.

An even more severe problem arises if one is interested not in the magnitude of a 10,000-year flood if it occurs in the next 40 years, but rather in the magnitude of the most severe flood in the next 10,000 years. Such questions might be of interest, for example, in analyzing the likelihood of water intrusion during the lifetime of a subterranean nuclear waste repository. In this case, the nonstationary nature of the flood distribution over a 10,000-year time horizon would almost certainly not be reflected in the 100 years of available data. One would need to take into account long-range phenomena (such as the possibility of a new ice age) that would not typically have been observed in the data.

The above discussion points up the need for clear definitions and terminology. For example, the "return period" (in years) of an extreme event can be defined alternatively either as the reciprocal of the probability of observing that event in any given year, or as the expected waiting time until the event occurs. These two definitions are equivalent in the stationary case, but may not be the same under nonstationary conditions, when the expected waiting time may be a better measure of return period.[48] Similarly, the floods discussed in the last two paragraphs may both be referred to as "10,000-year floods," but in the first case we are interested in a short-term prediction regarding floods with a (roughly constant) rate of 1 in 10,000 years, while in the other instance the idea that the arrival rate or probability of severe floods is even approximately constant is highly debatable and we must model an environment that is changing over the time scale of interest.

6.4. Modeling and Decomposition

One approach to improving the reliability with which probabilities for extreme events can be assessed is to reformulate the problem so that the probability of interest is no longer the tail probability of a distribution, but rather some measure of central tendency. For example, in PRAs of nuclear reactors, the probability of core melt is not computed as the tail of the distribution of peak fuel temperature; rather, core melt probability is modeled explicitly as a function of individual component failure probabilities. This approach has the potential to yield much more reliable results, but may not be applicable to all situations. For example, if we need the entire distribution of fuel temperature (rather than only the probability of exceeding some threshold value), then clearly assessing only a measure of central tendency will not be sufficient.

More generally, decomposition in some form has the potential to greatly improve the assessment of extreme event probabilities. The basic approach is to express the probability of interest in terms of quantities about which we know more. The use of decomposition is widely accepted as a useful technique not only in risk analysis, but also in a number of related fields such as decision analysis, industrial engineering, and long-range forecasting. For example, Raiffa[49] states that "The spirit of decision analysis is to divide and conquer: Decompose a problem into simpler problems, get one's thinking straight in these simpler problems, [and] paste these analyses together with logical glue." Similarly, Armstrong[30] notes:

> Decomposition has a number of advantages. It allows the forecaster to use information in a more efficient manner. It helps to spread the risk; errors in one part of the problem may be offset by errors in another part. It allows the researchers to split the problem among different members of a research team. It makes it possible for expert advice to be obtained on each part. Finally, it permits the use of different methods on different parts of the problem.

He concludes that, although there has been little empirical work documenting the benefits of decomposition in practice, "the limited evidence suggests that it is a technique with much potential. Research on how and when to use decomposition would be useful."

One noteworthy advantage of modeling or decomposition is that it can facilitate the use of partially relevant data, which is valuable because of the sparse nature of data on extreme events. The use of such partially relevant data requires some type of model for how the data relate to the quantity of interest. For example, in risk analysis of nuclear power plants, data on accident "precursors" are sometimes used to help quantify the frequency of actual core melt accidents (or to assess the reasonableness of results obtained through other methods), based on how "close" each observed precursor is to an actual accident. Similarly, toxicologists use data from nonhuman animal species to help quantify human health risks, based on scaling models for extrapolating the observed responses from animals to humans, and from relatively high experimental doses to lower naturally occurring exposures. One can likewise imagine estimating extreme event frequencies by explicitly modeling the causes of those events, which as indicated above may be different than the causes of less extreme events.

Decomposition does have some disadvantages, of course. In particular, if the model used for decomposition is itself incorrect, then the results of the process may actually be worse than would be obtained by assessing the quantity of interest directly (see Mosleh and Bier[50] for a discussion of modeling error in the context of simple risk analysis models). One common form of such modeling error is to assume that the elements of the decomposed model are independent; if they are in fact positively correlated with each other (as will frequently be the case), the assumption of independence can lead to overly narrow uncertainty distributions.

The use of modeling or decomposition also does not completely eliminate the difficulties of estimating extreme event probabilities. In particular, there will often be computational problems associated with the assessment of tail probabilities even if they are estimated as the output of a model. For example, assume that we have a deterministic model for the phenomenon of interest (e.g., peak fuel temperature in a passively safe reactor, peak stress in a structure due to an earthquake, etc.). Furthermore, assume that we wish to generate a probability distribution for the quantity of interest by simulating to account for uncertainties about the input parameters to that model (such as heat transfer coefficients, material properties, etc.). Depending on the width of the desired distribution, simulation can likely give accurate estimates of measures of central tendency (such as the mean or median) with only a few thousand or even a few hundred samples. However, accurately estimating a quantity such as the 10^{-4} fractile of the distribution will require well over 10,000 samples, which may not be computationally feasible if the original deterministic model (i.e., the physical model of the phenomenon of interest) is already computation-intensive.

Methods exist for dealing with these computational problems. For example, importance sampling[51] (e.g., sampling preferentially from the extreme tails of the input distributions, and then correcting for this in the analysis) can substantially improve the assessment of extreme tail probabilities. However, these methods are difficult to apply, and can result in much worse (instead of better) results if applied incorrectly. Therefore, care must be taken in using them, and the appropriate approach may vary considerably from one context to another (e.g., models of natural convection cooling vs. fracture mechanics).

6.5. Other Methods for Obtaining Prior Distributions

Clearly, it can be difficult to justify specifying a particular functional form and/or parameter value(s) for an informative prior distribution if not much is known about the problem at hand. Moreover, it may not be feasible to decompose the problem down to the level where the probability of an extreme event can be assessed using exclusively data on nonextreme events. In such cases, general guidelines for selecting suitable prior distributions based on theoretical considerations may be beneficial.

A number of methods have been suggested to guide the development of prior distributions in Bayesian analysis. These include: hierarchical Bayesian methods,[52] in which partially relevant data are used to help construct the prior distribution; empirical Bayesian methods, in which the actual data for the problem at hand are used to help construct the prior distribution; standardized or "reference" priors,[53, 54] which are thought to reflect "reasonable uncertainty" in a wide range of problems; so-called "ignorance" or "noninformative" priors, which contain as little information as possible about the quantity of interest; and maximum entropy priors (discussed previously). All of these methods are discussed at least briefly by Berger,[26] and all have been suggested at one time or another as being appropriate when little is known about the quantity of interest, and hence the construction of a true informative prior distribution is difficult.

Each of these methods has some advantages and disadvantages. For example, Berger notes that empirical Bayesian methods work well when the number of data points available is large. However, he states that when only small or moderate amounts of data are available, hierarchical Bayesian procedures are preferred (because they capture the uncertainty arising from the fact that the moments of the data, for example, will be only estimates of the true moments, if the true moments even exist). Thus, hierarchical Bayesian procedures may be preferable for assessing probabilities of extreme events.

In the hierarchical Bayesian approach, the quantity of interest is viewed as being a sample from a population of similar quantities; the variability of that distribution is described using so-called "hyper-parameters" (e.g., the mean and variance of the population). However, personal experience with hierarchical Bayesian methods on the part of one of the authors has shown that the results can be quite sensitive to seemingly small changes in the hyper-parameters. For example, a relatively large number of data points sampled from the population of interest (typically dozens) is generally needed to significantly narrow the prior distribution for the population variance. Therefore, even small probabilities assigned to large values of the population variability in the "hyper-prior" (or first-stage prior) can have a large effect on the resulting posterior distribution.

Noninformative priors seem at first glance to be a reasonable approach when one does not know very much, but unfortunately this approach is not necessarily as objective as it appears, since for many problems the choice of a noninformative prior is not unique. The use of an (informative) reference prior may be desirable in cases involving public decision-making, where different parties may disagree and wish to resort to a more "neutral" analysis that will be dominated by the likelihood function. However, the choice of a particular reference prior may be even harder to defend than the choice of a noninformative prior.

6.6. Bounding Methods

As a result of the limitations of the various approaches discussed above, much interest has recently focused on so-called "robust" Bayesian methods and other bounding approaches. Robust Bayesian methods generally involve updating an entire *class* of priors (rather than a single prior distribution) with observed data. Note that if the class is chosen carefully, the computational effort involved in considering all distributions in the class need not be substantially greater than for considering a single distribution. If all (or most) priors in a suitably broad class give similar results (e.g., a similar posterior distribution or optimal decision), this can lead to greatly improved confidence in the results of the analysis (although robustness does not necessarily imply correctness; the result might be stable, but still wrong).

In a similar spirit is probability bounds analysis,[55–58] which is a bounding approach used for propagating uncertainties in models (rather than for Bayesian updating). In this approach, the analyst specifies bounds on the cumulative distribution functions of the various parameters used as inputs to a model, rather than a single informative distribution for each input value. These input bounds are then propagated through the model. The uncertainty propagation process, which is computationally quite efficient, yields valid bounds on the cumulative distribution for the final result of the model (e.g., an estimated risk level).

Both of these bounding approaches can be useful when sufficient data are not available to justify the selection of a single distribution. However, Berger[26] points out that "There is no easy answer to the robustness problem." In particular, if plausible changes in a prior distribution (or plausible changes in the input distributions to a model) lead to large changes in results (e.g., significantly different optimal decisions), this simply indicates that there is not yet enough empirical information available to provide definitive support for any particular decision. In this situation, the only real solution is to obtain additional data, which of course may not be possible in the short term.

7. HOW CAN WE MAKE DECISIONS ABOUT EXTREME EVENTS?

The pitfalls of simple expected value decision-making are well known, but are likely to be even greater when there is the potential for extreme events (e.g., catastrophic accidents, bankruptcy, etc.). Essentially, expected value decision-making treats low-probability, high-consequence (i.e., extreme) events the same as more routine high-probability, low-consequence events, despite the fact that they are unlikely to be viewed in the same way by most decision-makers. For example, if two investments have the same expected profit, but one has a much larger chance of bankruptcy than the other, they are hardly likely to be viewed as equally desirable, due to attitudes such as risk aversion;[59] similar risk aversion applies to large multiple-fatality accidents.[60–64] Researchers have also documented that people tend to put greater weight on low-probability events in decision-making than expected value decision-making would indicate.[65–69]

Some other techniques have been suggested for lending visibility to low-probability, high-consequence events, so that they are not camouflaged by expected value calculations. For example, the use of *conditional* expected values (e.g., conditional on exceeding some threshold value)[70] can help focus the decision maker's attention on the tail of a distribution. Similarly, approaches such as multiple-criteria decision-making[70–74] can treat the risk of extreme events separate from the overall (i.e., unconditional) expected value. Such methods can provide decision-makers with less aggregated information about policy trade-offs and allow them to explicitly determine how much weight to give to extreme vs. nonextreme events based on their own value judgments. (Note, however, that such approaches run the risk of putting *too much* emphasis on the risk of extreme events relative to more routine outcomes.)

The ultimate extension of such approaches is expected utility decision-making, which is intended to resolve this problem by taking into account the differential impact of large losses (such as bankruptcy) vs. more moderate losses (for which utility functions may be approximately linear). Despite the theoretical niceties of expected utility decision making, however, some practical issues may need to be resolved, particularly when applying expected utility decision–making to the risks of extreme events. First, it may be difficult if not impossible to assign utilities to some extreme events. For example, a few years ago there were serious concerns that a nuclear war could set off a "nuclear winter," with the potential to destroy human civilization, and possibly even end all life on the planet. Few if any of us would probably be willing to put a finite utility on outcomes such as these. Similarly, consider the situation of the title character in *Sophie's Choice,*[75] who was forced to choose which of her two children to save.

Thus, Rescher[76] has argued that "*no matter what the balance of probabilities, the 'reasonable man' would not risk loss of life or limb to avert the prospect of some trivial inconvenience*" (emphasis in original), and that some hazards or losses are therefore fundamentally incommensurate with others. The point is perhaps debatable; people regularly risk life and limb crossing the street to avoid the inconvenience of missing a bus, for instance. However, at least the difficulty (if not the impossibility) of applying expected utility decision making principles to situations with the potential for truly extreme losses seems clear.

Furthermore, since we have seen that the probabilities of extreme events are likely to be highly unreliable, the expected utilities computed based on these probabilities will tend to be unreliable as well. If minor differences in a small probability can result in major changes to the expected utility, and hence the optimal decision, expected utility decision-making may be unable to yield definitive recommendations as to the best course of action. This begs the question, of course, about what other approach (if any) is capable of resolving this problem. We are led back to Berger's observation that "There is no easy answer to the robustness problem"[26]; if minor changes in either probabilities or utilities can give rise to significant changes in the optimal course of action, there is little to be done other than taking a gamble on what appears to be the best course of action, or delaying the decision and gathering more information.

8. SUMMARY AND DIRECTIONS FOR FURTHER WORK

As stated in Sect. 1, this paper was not intended to be definitive. Rather, we hope that we have stimulated thinking on some possible approaches for assessing and managing the risks from extreme events. Further research on this issue is clearly needed. Assessing the probabilities of extreme events is generally a difficult task, and is likely to remain so even given the results of research on better approaches for doing so. However, at least some suggestions and directions for further work can nonetheless be usefully put forth at this point.

For example, many agree that modeling and decomposition are one of the most promising approaches for assessing the risks of extreme events. This observation not only provides useful guidance to practitioners at present, but also suggests possibly fruitful avenues for further research. In particular, there may be some classes of problems for which useful decomposition or modeling approaches have not yet been identified, or for which severe computational problems exist. Further work in such areas may yield feasible and beneficial modeling approaches. More generally, as pointed out by Armstrong,[28] further research on which approaches to decomposition yield the best results would also be useful.

Similarly, the risk analysis community has paid little attention to recent developments in Bayesian statistics, such as the work on reference priors and robust Bayesian analysis.[26] Much work has been done in these areas, from which the risk analysis community might benefit. However, assessment of risks from extreme events is different in kind from many of the problems faced in other areas, due to the small amounts of data that are typically available on extreme events, so further work specifically focused on issues of concern in risk analysis might be desirable.

REFERENCES

1. P. Slovic, B. Fischhoff, and S. Lichtenstein, "Rating the Risks," *Environment* **21,** 14–39 (April 1979).
2. R. Zimmerman, "After the Deluge," *Sciences* **34,** 18–23 (1994).
3. R. Zimmerman, "Understanding Industrial Accidents Associated with New Technologies: A Human Resources Management Approach," *Ind. Crisis Quart.* **2,** 229–256 (1988).
4. H. Kumamoto and E. J. Henley, *Probabilistic Risk Assessment and Management for Engineers and Scientists* (IEEE Press, New York, 1996).
5. M. B. Fiering and J. Kindler, "Surprise in Water-Resource Design," *Int. J. Water Res. Devel.* **2,** 1–10 (1984).
6. Y. Y. Haimes, "Hierarchical Holographic Modeling," *IEEE Trans. Syst. Man Cybernet.* **11,** 606–617 (1981).
7. Y. Y. Haimes, "Sustainable Development: A Holistic Approach to Natural Resource Management," *IEEE Trans. Syst. Man Cybernet.* **22,** 413–417 (1992).
8. C. Chittister and Y. Y. Haimes, "Risk Associated with Software Development: A Holistic Framework for Assessment and Management," *IEEE Trans. Syst. Man Cybernet.* **23,** 710–723 (1993).
9. R. W. Katz and B. G. Brown, "Extreme Events in a Changing Climate: Variability Is More Important than Climate," *Climatic Change* **21,** 289–302 (1992).
10. L. A. Zadeh, "Fuzzy Sets," *Inform. Control* **8,** 338–353 (1965).
11. A. Kaufmann and M. M. Gupta, *Introduction to Fuzzy Arithmetic* (Van Nostrand Reinhold, New York, 1991).
12. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976).
13. W. Dong and F. S. Wong, "From Uncertainty to Approximate Reasoning: Part 1: Conceptual Models and Engineering Interpretations," *Civil Eng. Syst.* **3,** 143–150 (1986).
14. P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1991).
15. S. Ferson, "Using Fuzzy Arithmetic in Monte Carlo Simulation of Fishery Populations," in T. J. Quinn (ed.), *Management of Exploited Fish: Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations* (Alaska Sea Grant College Program, AK-SG-93-02, 1993).
16. S. Ferson and L. R. Ginzburg, "Different Methods Are Needed to Propagate Ignorance and Variability," *Reliability Eng. Syst. Safety* **54,** 133–144 (1996).
17. A. Bardossy and L. Duckstein, *Fuzzy Rule-Based Modeling with Applications to Geophysical, Biological, and Engineering Systems* (CRC Press, New York, 1995).
18. D. R. J. Moore and S. Ferson, "Probabilistic Risk Analysis: Possible Approaches when Information Is Limited (As It Always Is)," Draft manuscript (1997).
19. J. Gleick, *Chaos: Making a New Science* (Viking, New York, 1987).
20. T. Poston and I. Stewart, *Catastrophe Theory and Its Applications* (Pitman, London, 1978).
21. G. L. S. Shackle, *Imagination and the Nature of Choice* (Edinburgh University Press, Edinburgh, 1972).
22. S. Lichtenstein, P. Slovic, B. Fischhoff, M. Layman, and B. Combs, "Judged Frequency of Lethal Events," *J. Exp. Psychol.: Human Learn. Memory* **4,** 551–578 (1978).
23. F. Mosteller and J. W. Tukey, *Data Analysis and Regression: A Second Course in Statistics* (Addison-Wesley, Reading, MA, 1977).
24. H. C. Frey and D. E. Burmaster, "Methods for Characterizing Variability and Uncertainty: Comparison of Bootstrap Simulation and Likelihood-Based Approaches," NSF/UVA Workshop on When and How Can You Specify a Probability Distribution When You Don't Know Much, II, Center for Risk Management of Engineering Systems, University of Virginia, Charlottesville, VA (1997).
25. J. K. Hammitt and A. I. Shlyakhter, "The Possibility of Surprise and the Expected Value of Information," NSF/UVA Workshop on When and How Can You Specify a Probability Distribution When You Don't Know Much, II, Center for Risk Management of Engineering Systems, University of Virginia, Charlottesville, VA (1997).
26. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).
27. R. M. Cooke, *Experts in Uncertainty: Opinion and Subjective Probability in Science* (Oxford University Press, Oxford, England, 1991).
28. A. Mosleh, V. M. Bier, and G. Apostolakis, *Methods for the Elicitation and Use of Expert Opinion in Risk Assessment: Phase I—A Critical Evaluation and Directions for Future Re-*

search (U.S. Nuclear Regulatory Commission, Washington, D.C., NUREG/CR-4962, 1987).
29. D. von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research* (Cambridge University Press, Cambridge, 1986).
30. J. S. Armstrong, *Long-Range Forecasting: From Crystal Ball to Computer* (Wiley, New York, 1985).
31. H. F. Martz and R. A. Waller, *Bayesian Reliability Analysis* (Wiley, New York, 1982).
32. R. F. Bordley, "Lottery-Dependent Probabilities in Quantum Mechanics, Psychology and Bayesian Statistics," Paper presented at the 45th Meeting of the NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics (1992).
33. D. Hattis, personal communication (1997).
34. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* **27,** 623–656 (1948).
35. S. Goldman, *Information Theory* (Prentice-Hall, New York, 1953).
36. J. G. Greeno, "Evaluation of Statistical Hypotheses Using Information Transmitted," *Phil. Sci.* **37,** 279–294 (1970).
37. E. T. Jaynes, *Probability Theory: The Logic of Science* (http://bayes.wustl.edu/pub/Jaynes/book.probability.theory, 1995).
38. R. D. Levine and M. Tribus, *The Maximum Entropy Formalism* (MIT Press, Cambridge, 1978).
39. R. C. Lee and W. E. Wright, "Development of Human Exposure-Factor Distributions Using Maximum-Entropy Inference," *J. Exp. Anal. Environ. Epidemiol.* **4,** 329–341 (1994).
40. A. H. Ang and W. H. Tang, *Probability Concepts in Engineering Planning and Design,* Vols. 1 and 2 (Wiley, New York, 1984).
41. E. Castillo, *Extreme Value Theory in Engineering* (Academic, Boston, 1988).
42. E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
43. R. R. Kinnison, *Applied Extreme Value Statistics* (MacMillan, New York, 1985).
44. J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis, "Probability Weighted Moments: Definition and Their Relation to Parameters of Several Distributions Expressible in Inverse Form," *Water Res. Res.* **15,** 1049–1054 (1979).
45. J. W. Tukey, *The Practical Relationship between the Common Transformation of Percentages or Fractions and of Amounts* (Technical Report 36, Statistical Techniques Research Group, Princeton University, Princeton, NJ, 1960).
46. J. C. Houghton, "Birth of a Parent: The Wakeby Distribution for Modeling Flood Flows," *Water Res. Res.* **14,** 1105–1110 (1978).
47. M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes* (Springer-Verlag, New York, 1983).
48. R. Olsen, J. H. Lambert, and Y. Y. Haimes, "Risk of Extreme Events Under Non-Stationary Conditions," Draft manuscript (1996).
49. H. Raiffa, *Decision Analysis* (Addison-Wesley, Reading, MA, 1968).
50. A. Mosleh and V. M. Bier, "On Decomposition and Aggregation Error in Estimation: Some Basic Principles and Examples," *Risk Anal.* **12,** 203–214 (1992).
51. R. Y. Rubinstein, *Simulation and the Monte Carlo Method* (Wiley, New York, 1981).
52. S. Kaplan, "On a 'Two-Stage' Bayesian Procedure for Determining Failure Rates from Experiential Data," *IEEE Trans. Power Apparatus Syst.* **PAS-102,** 195–202 (1983).
53. H. Jeffreys, *Theory of Probability* (Oxford University Press, London, 1961).
54. A. Zellner and A. Siow, "Posterior Odds Ratios for Selected Regression Hypotheses," in J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (eds.), *Bayesian Statistics* (University Press, Valencia, Spain, 1980).
55. S. Ferson and T. F. Long, "Conservative Uncertainty Propagation in Environmental Risk Assessments," in J. S. Hughes, G. R. Biddinger, and E. Mones (eds.), *Environmental Toxicology and Risk Assessment,* Vol. 3 (American Society for Testing and Materials, ASTM STP 1218, Philadelphia, 1995).
56. S. Ferson, "Quality Assurance for Monte Carlo Risk Assessments," *The Third International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society* (IEEE Computer Society Press, Los Alamitos, California, 1995).
57. S. Ferson, "Probability Bounds Analysis Software," in A. Gertler (ed.), *Computing in Environmental Resource Management* (Conference Proceedings, Air and Waste Management Association and the U.S. Environmental Protection Agency, Pittsburgh, 1997).
58. S. Ferson, L. R. Ginzburg, and H. R. Akcakaya, "Whereof One Cannot Speak: When Input Distributions Are Unknown," *Risk Anal.* (in press).
59. R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives* (Wiley, New York, 1976).
60. J. W. Pratt, "Risk Aversion in the Small and in the Large," *Econometrica* **32,** 122–136 (1964).
61. C. Starr, R. Rudman, and C. Whipple, "Philosophical Basis for Risk Analysis," *Ann. Rev. Energy* **1,** 629–662 (1976).
62. P. Slovic, B. Fischhoff, and S. Lichtenstein, "Risk Assessment: Basic Issues," in R. W. Kates (ed.), *Managing Technological Hazard: Research Needs Opportunities* (Institute of Behavioral Science, University of Colorado, Boulder, 1977).
63. J. M. Griesmeyer and D. Okrent, "On the Development of Quantitative Risk Acceptance Criteria," *An Approach to Quantitative Safety Goals for Nuclear Power Plants* (Advisory Committee on Reactor Safeguards, U. S. Nuclear Regulatory Commission, Washington, D.C., NUREG-0739, 1980).
64. R. L. Keeney, "Equity and Public Risk," *Operat. Res.* **28,** 527–534 (1980).
65. M. G. Preston and P. Baratta, "An Experimental Study of the Auction-value of an Uncertain Outcome," *Am. J. Psychol.* **60,** 183–193 (1948).
66. R. M. Griffith, "Odds Adjustments by American Horse Race Bettors," *Am. J. Psychol.* **62,** 290–294 (1949).
67. R. C. Sprowls, "Psychological-mathematical Probability in Relationships of Lottery Gambles," *Am. J. Psychol.* **66,** 126–130 (1953).
68. P. Nogee and B. Lieberman, "The Auction Value of Certain Risky Situations," *J. Psychol.* **49,** 167–179 (1960).
69. M. M. Ali, "Probability and Utility Estimates for Racetrack Bettors," *J. Political Economy* **85,** 803–815 (1977).
70. E. Asbeck and Y. Y. Haimes, "The Partitioned Multiobjective Risk Method," *Large Scale Systems* **6,** 13–38 (1984).
71. V. Chankong and Y. Y. Haimes, *Multiobjective Decision Making: Theory and Methodology* (Elsevier, New York, 1983).
72. Y. Y. Haimes, K. Tarvainen, T. Shima, and J. Thadathil, *Hierarchical Multiobjective Analysis of Large-Scale Systems* (Hemisphere, New York, 1990).
73. Y. Y. Haimes, "Total Risk Management," *Risk Analysis* **11,** 169–171 (1991).
74. Y. Y. Haimes and D. Li, "A Hierarchical-Multiobjective Framework for Risk Management," *Automatica* **27,** 579–584 (1991).
75. W. Styron, *Sophie's Choice* (Random House, New York, 1979).
76. N. Rescher, *Risk* (University Press of America, Lanham, MD, 1983).