Assessment of Joint Intervenors' Multi-Attribute Approach

Douglas W. Hubbard and Sam L. Savage

October 21, 2016

On August 18, 2016, the CPUC issued Decision 16-08-018 adopting on an interim basis an Approach proposed by TURN and the Energy Users and Producers Coalition/Indicated Shippers (also known as the "Joint Intervenors") for risk analysis. The Utilities were given the opportunity to both vet and test drive the Joint Intervenor Approach (also known as "Multi-Attribute Approach" or "MAA")[1] and consequently engaged the authors of this report to assess the proposed Approach. Specifically, the authors were engaged to:

- Evaluate practical and theoretical limits and proper use of the method proposed by the Joint Intervenors;
- Describe how the use of probabilistic methods, particularly simulations, are currently being used.

---

[1] In response to comments, the draft decision was substantively modified. The last version stated that it (among other things):

> "Clarifies that the decision is 'interim' decision subject to the following before final adoption of any model:
> a. Vet the Joint Intervenor Approach 'foundational requirements' and how they operate in a real world setting;
> b. 'Test drive' the model, using a small set of detailed test problems (at least five), which are common across more than one utility;
> c. Review results of utility pilots that highlight equivalent or similar features of the Multi-Attribute Approach."

D.16-08-018, p. 177

Contents

Introduction

In the interests of a more quantitative approach to risk mitigation "consistent with the principle of just and reasonable cost based rates," the Commission's Decision 16-08-018 of August 18, 2016, expresses the objectives below (with page numbers):

1) properly quantifying risk reduction (pages 3, 6, 11, 35)
2) moving toward a standard (pages 5, 6, 7, 13)
3) transparency (pages 6, 49, 82, 89, 124)
4) moving away from subjective scores (pages 20, 76, 90, 105)
5) potential for short-term application (pages 12, 73, 81)

The Decision also "Directs Utilities to 'test drive' the Multi-Attribute Approach [MAA] using real world problems before full scale adoption of any methodology" [Decision p.2]. In preparation for such a test drive, the authors of this document have been asked to investigate MAA from a technical perspective and how it actually supports the above stated objectives of the Commission's Decision.

We have found that MAA, as described, does not support the stated objectives. MAA is not a probabilistic approach, nor is MAA likely to be an effective bridge to a probabilistic approach. Further, in our opinion, MAA does not adequately represent the risk in the Joint Intervenors' first example problem [Presentation p.15], and may even exaggerate the risk.

We find that widely used probabilistic methods based on simulation do support the stated objectives of the Commission. Also, in the last couple of years, proven probabilistic approaches have become even more practical and far more accessible through new technologies. Thus, work with probabilistic modelling can proceed expeditiously.

This document is divided into two main sections. The first addresses the problems we have identified with the MAA. The second addresses some steps the CPUC and the utilities can take toward more probabilistic approaches to risk modeling.

The Proposed Multi-Attribute Approach

While MAA may avoid some of the correctly identified problems of the methods used by the Utilities, it does not alleviate most of the stated problems and introduces others. It relies on an oversimplification that ignores uncertainty and risk as well as risk tolerance.

In the CPUC SED Preliminary Draft Report dated July 7, 2014, there is precisely one case where peer-reviewed, published research was cited to list the problems with the methods used by the Utilities. This was the paper titled "Problems with Scoring Methods and Ordinal Scales in Risk Assessment (Problems with Scoring Methods and Ordinal Scales in Risk Assessment, 2010)[2]" which was co-authored by Doug Hubbard (and Dylan Evans), who is also co-authoring this response. The Joint Intervenors later referenced this paper in the evidentiary hearing dated February 5, 2015 while interviewing a utility witness. That paper stated:

> "First and foremost, useful risk assessment methods must use explicit probabilities and magnitudes of losses expressed quantitatively instead of using surrogate verbal or ordinal scales."

SED and the Joint Intervenors were correct that the above-cited source finds many faults with the methods historically used by the Utilities. But the proposed MAA method does not actually address the concerns described in the paper. MAA is still basically a weighted scoring method of the type Hubbard and Evans argued against. Furthermore, MAA does not address another problem identified in the SED Report,

> "Scoring techniques often presume that the factors being scored are independent of each other – i.e. there are no correlations between factors. This assumption is rarely tested or justified in any way."

And there are even more significant problems related to the fact that MAA incorrectly equates expected value of a loss (that is, the probability-weighted average loss) to risk. On page 110 of Decision 16-08-018 it states that "probability distributions are used in the Joint Intervenor methodology; they are taken one step further, however into an expected value to allow calculation of risk reduction." Probability distributions (expressions of uncertainty) are often reduced to an *expected value*, a single number, also known as an *average*. The very definition of risk in MAA is an average single number defined as:

"Risk = Likelihood of Failure (LoF) X Consequence of Failure (CoF)"

[p. 5 1-22-16 Intervenor Presentation Slide Deck (Presentation)]

However, this is not "one step further" than a probability distribution, as described in the Decision. It is also not the same as using the "actual likelihood or probability" or the "complete probability

[2] Hubbard, Evans "Problems with Scoring Methods and Ordinal Scales in Risk Assessment" *IBM Journal of Research and Development* 2010 Vol. 54, Number 3, Paper 2, May/June 2010.

distribution" as described in pages 95-96 of the Decision. On the contrary, this is a classic error in risk management that often results in serious miscalculations[i,ii]. Following are some of the problems with using expected values in this way:

- MAA does not distinguish among widely differing risks
- MAA can miscalculate risks through at least two mechanisms
- MAA ignores the value of information
- MAA is not a Standard in Quantitative Risk Assessment (Monte Carlo is)

In addition to showing how MAA improperly addresses the mathematics of risk, we will show that other methods are much more widely adopted as a standard in quantitative risk assessment, and that these methods have recently become universally accessible through the common Excel spreadsheet.

MAA Does Not Distinguish Among Widely Differing Risks

In the example on p. 14-15 of Joint Intervenors' Presentation, MAA fails to adequately address risk. In their example, we are asked to consider a hypothetical risk that has a "50% chance of no injury, 25% chance of minor injury; 15% chance of major injury, and 10% chance of death. A total of 3 people will be exposed." Imagine the risk of an explosion in a substation. MAA reduces this risk to a single risk score of 51 as further described below.

But before considering the details, we must point out that the Joint Intervenor methodology would also give risk scores of 51 to the following very different risks.

1. A 51% chance of exposing 10 people to minor injury
2. A 1% chance of exposing 51 people to death
3. A 17% chance of exposing 3 people to death
4. The exposure of 3 people to a 2.9% chance of no injury, a 61% chance of minor injury, a 36% chance of major injury, and a 0.1% chance of death

Consider how this sort of risk measurement would be applied in comparing different risk mitigations to the risk of a certain fatality; that is, a 100% chance of killing one person. This would have a risk score of 100%*100 =100. Now suppose we are considering two different mitigations that require the same resources to implement.

- Mitigation 1 reduces the risk of a certain fatality to that of case 2 above, resulting in a 1% chance of killing 51 people.
- Mitigation 2 reduces the risk of a certain fatality to case 4 above, resulting in a 0.1% chance of killing 3 people but higher chances of various injuries.

Both mitigations result in a reduction in risk score of exactly 49. And according to the Joint Intervenor's white paper, p.5 "Because both risk mitigation actions reduce risk by the same amount, we should be indifferent between them if the resources required for implementation are the same." Yet these residual risks would be viewed very differently by most people, and would be clearly distinguished by a probabilistic approach that preserved the distributions of outcomes.

Because MAA considers all mitigations of the same expected value to be equally preferable, it fails to capture risk tolerance. Most decision makers exhibit some level of "risk aversion." That is, they would prefer a certain amount of $5 to a bet that gives a 50% chance of gaining $1,000,000 and a 50% chance of losing $999,990 (where the expected value is $5). Quantifying this risk aversion is a critical step in the risk/return analysis of decisions.

It is interesting to note that on this issue, MAA actually *deviates from the seminal work*[3] *in Multi-Attribute Utility Theory* (MAUT). The original theory of MAUT gives a detailed treatment of the real-world tendency toward risk aversion. Unlike MAA, MAUT clearly distinguishes expected value of a bet with an uncertain outcome from the certain value a risk averse decision maker would consider equivalent to it. Note that this issue cannot simply be addressed in all situations by making the utility functions nonlinear in a way which is independent of the probability of the event.

The problem of distinguishing different risks is compounded when MAA further collapses the risk score of 51 for safety with risk scores associated with financial and reliability risk categories. If the number 51 did not adequately represent the safety risk in this situation, then subsequent calculations based on that number can only become less meaningful as more categories of risk are added.[4]

It is also stated on p. 6 of the Decision that it is the "intent that the adoption of these additional procedures will result in additional *transparency*." However, imagine how differently the newspaper headlines would read for each of the "51"s above. By reducing each distribution to a single number, MAA would not increase transparency.

MAA Miscalculates Risk Through Three Mechanisms

Failing to differentiate risk properly leads to some serious miscalculations. Continuing with the example risk of MAA at page 15 of the Presentation, the safety score is computed as follows:

"Scaled value of (0.50 x 0) + (0.25 x 10) + (0.15 x 30) + (0.10 x 100) = 17 per person"

The "Scaled Value" reduces the *probabilistic* representation of the various outcomes and associated probabilities above to a single number, the weighted average. This step is a critical flaw in MAA, which lays the foundation for subsequent errors.

---

[3] R. Keeney, H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Cambridge University Press, 1993.

[4] There are further simple calculations on p. 16 of the Presentation that show two different mitigations that reduce the risk score of 19.125 by 3.825 and 6.375. Remember that buried in these numbers is the ambiguity of whether we have a 1% chance of killing 51 people or a 51% chance of 10 minor injuries. As we mentioned, earlier calculations with averages are simple to perform, but that does not indicate that they are meaningful.

After reducing this situation to the number 17, it is no longer possible to correctly address risk, and the trouble shows up immediately. To account for the 3 people exposed, MAA simply multiplies the 17 safety risk units per person above by 3 to get:

"51 units total" [Presentation p.15]

Just to see why this matters, let's look at two broad categories of common issues in risk analysis: diversification effects and nonlinearities.

Diversification Effects

As shown above, MAA suggests that we have *three* times the risk with *three* people than we had with *one* person. This indicates that instead of a 10% chance of one death, there is now a 10% chance of three deaths. But this indication may be highly inflated. For there to actually be a 10% chance of three deaths, we would need to know that *all* people exposed would necessarily suffer the *same* fate, for example, like what typically occurs in a plane crash. More likely, the workers would be in various locations around the failed substation, and would therefore suffer *independent* fates. But in this case the probability of three deaths in a single failure would be 10% X 10% X 10% X = 0.1% or one chance in 1,000.

Here we see in the first example presented by the Joint Intervenors, that MAA has the potential to exaggerate the risk of three deaths by as much as 100 fold. This is a systematic error inherent in MAA's reduction of probability distributions to expected values, which could lead to ratepayers funding potentially unnecessary mitigations. This effect is further magnified as the risk is rolled up across multiple substations, and ultimately across all assets in the system. For example, if we added up the risk of 10 substations with 3 people exposed, we would get a risk score of 510, which could be interpreted as suggesting one chance in 10 of killing 30 people. If the risks were independent, then the actual risk of killing 30 people would be 1 chance in 10,000,000,000.

In the independent case the risks are said to be *diversified*. When they are not independent, they are said to be *correlated*. MAA, which calculates with expected values instead of probability distributions, does not address the distinction between independent and dependent fates of those exposed. Expected values cannot express the concepts of *diversification*, *correlation*, and other important forms of statistical dependence that can only be modelled through simulation.

Nonlinearities

Another important error caused by Expected Values involves how it computes non-linear risks.[5]

[5] This is separate from the claim that non-linear value functions can accurately represent risk preferences. In this case, however, we are referring not to how preferences of risks are modeled, but how risks themselves are computed.

Suppose there are equal chances of a cool (75°), average (80°), or hot (85°) summer in a region whose power grid becomes stressed through excessive air-conditioning use. There are no reliability consequences of a cool summer. In an average summer there is expected to be one hour of outage. But a hot summer is expected to lead to 14 total hours of outage. Many planners still base their decisions on average conditions. So planning for the average summer, we might expect one hour of outage in a year. But the correct average of outage hours over the three possibilities is (0+1+14)/3 = 15/3 = 5 hours of outage.

In this example the use of averages underestimated the risk. The difference between the outcome of the average assumption (one outage hour) and the average outcome (5 outage hours) is known by mathematicians as Jensen's Inequality, and it guarantees that most calculations in risk management will be erroneous if based on expected values.

MAA Ignores the Value of Information

Another critical issue arises when probability distributions are reduced to expected values. In real world problems, expected values alone offer no way to compute the value of information. Some of the most cost-effective mitigation investments involve improving the accuracy of uncertain estimates. Without value of information calculations, these sorts of mitigations will not even be considered. When there are multiple uncertainties, such as the reliability of a new type of asset or the cost per mile of a mitigation investment, the value of improved estimates can only be found through simulation. Such simulations then pinpoint those estimates that are most important to improve. This occurred in a recent pilot with PG&E, when two competing mitigation projects were being compared.

> **Project 1:** Bury about 100 miles of line in the Santa Cruz mountains, where heavy vegetation damages lines.
>
> **Project 2:** Replace about 500 miles of line in the Central Valley, but keep it overhead.

This pilot decision included uncertainties such as those mentioned above and many more (91 in total) each with different interacting distributions of probable outcomes. If all the distributions were reduced to expected values, then the analysis favored burying lines in the mountains over replacing the overhead lines in the valley by over $12 million.

However, by keeping the uncertainties explicit and running a Monte Carlo simulation of the 91 uncertain variables it was discovered that a previously overlooked uncertainty was, in fact, important: the number of businesses served per mile in the Central Valley. PG&E improved their estimates of this and other uncertainties at minimal expense, then re-ran the analysis. The pilot decision was reversed, with the Central Valley project coming in $14 million ahead of the buried lines, bringing greater value to PG&E's customers.

MAA is not a Standard in Quantitative Risk Assessment (Monte Carlo is)

One of the objectives stated in the Decision is the standardization of risk communication methods[6]. We strongly agree with this goal and we argue that the most direct route to standardization is the adoption of existing probabilistic standards.

Recall that the previously mentioned paper by Hubbard and Evans (which was used to make points by both SED and Joint Intervenors) argues that risks must be "expressed quantitatively instead of using surrogate verbal or ordinal scales". The authors recognize this statement is not an innovation in risk assessment at all. It is merely the recognition of the prevalent standard.

SED has also acknowledged the value of a quantitative approach to assessing risk in a published white paper in 2015[iii], which is consistent with these standards of probabilistic modeling. That is, risk is not represented as a weighted score or an expected value, but a range of outcomes and associated likelihoods, known as a probability distribution.

We will argue, however, that the methods that MAA are based on are far from a standard in real-world quantitative risk analysis. We will begin to make this point by first describing standards for communicating risk and then the standard for how risk is actually computed by actuaries, statisticians, and the vast majority of professionals who perform quantitative risk analysis.

Standards for Communicating Probability Distributions

There are several standard ways to represent probability distributions. They all contain the same information, but some are more useful in specific contexts that others.

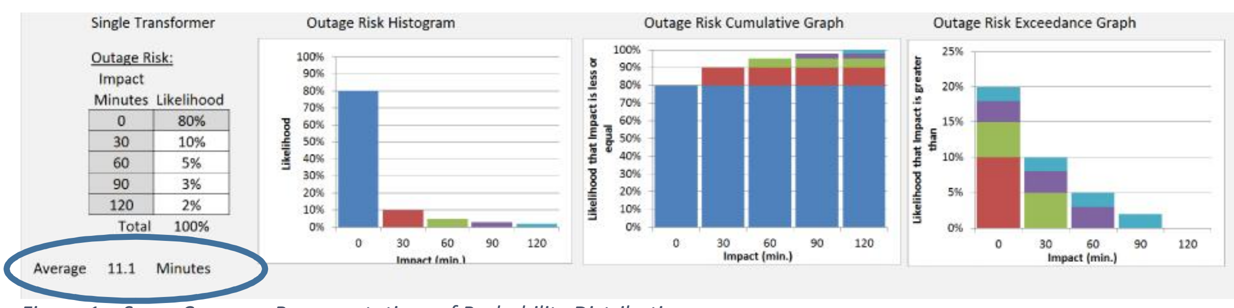


Figure 1 – Some Common Representations of Probability Distributions

We list five standard approaches below in terms of the risk of outage of a transformer.

1. A Table
   The Table, on the left in Figure 1, contains potential impacts and associated likelihoods that the impact will occur. In this case, the impact is the number of minutes of an outage, broken into 30 minute buckets. Note that the buckets may be defined as desired, for example, minutes, seconds, or even milliseconds, in which case the graphs would appear as smooth curves. The

[6] Decision 16-08-018 California Public Utilities Commission pg. 6

Average, or Expected Value is also displayed as 11.1 minutes, circled in blue.

2. A Histogram
   This is simply a bar graph of the table, in which the impacts appear on the X axis and the likelihood of that impact appears on the Y axis. Typically, the bars of a histogram are all the same color. We use different colors here, to better describe the additional graphs.

3. A Cumulative Graph
   A cumulative graph, like a histogram, has the impacts on the X axis, but the Y axis now displays the likelihood that the impact will be this amount or less. For example, there is a 90% chance that the outage will be 30 minute or less.

4. A Loss Exceedance Curve (LEC)
   A Loss Exceedance Curve is another particularly widely used method in quantitative risk analysis (right most graph in Figure 1). It shows the chance that a loss exceeds some amount during a given period. This method also provides a convenient way to compare risks and risk tolerance. In the example below (Figure 2), the loss is shown as a monetary amount in a given year. The risk tolerance curve explicitly states the acceptable level of a chance of a loss in a given year. In this case, it indicates that the organization can accept a 50% chance of a loss greater than $12 million in a given year but not a 10% chance of a loss greater than $20 million in the same period. The risk position shown indicates that the risks of larger losses are greater than what is acceptable to the organization.

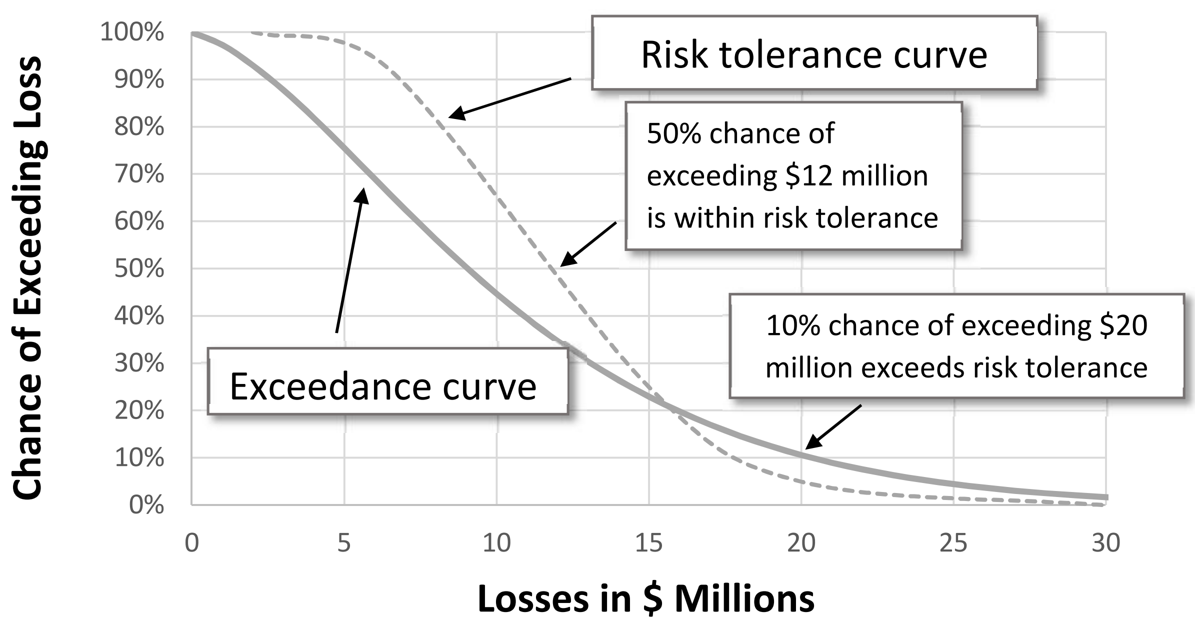


Figure 2 –A Loss Exceedance Curve Example

5. Auditable Data Sets
   The previous representations have been in use for over a century for communicating probability distributions. However, for communicating probability distributions between computer applications, arrays of data, called SIPs, may be used[iv]. These arrays offer the following advantages over the four representations above.

   a. They allow risks of individual assets to be aggregated across the system.
   b. They preserve the statistical dependence or correlations between risks.
   c. They provide an audit trail for combined risk calculations.
   d. They may be accessed by every-day software such as Microsoft Excel, allowing stakeholders to explore the chances that multiple categories of risk exceed specified limits.

For example, the US Geological Survey hosts a website that generates probability distributions of earthquake accelerations for any latitude and longitude in the United States[v]. These could be transmitted using this approach to probabilistic risk models of numerous utility assets. The resulting risks could then be further aggregated across the system.

The first four of these representations of probability distributions contain the same mathematical information expressed in different ways. The fifth, in addition, preserves the information of correlation and other forms of statistical dependence. MAA reduces them all to a single number; the Expected value (Average) of 11.1. That means that MAA would equate the outage risk above with what many would consider a greater risk, which nonetheless has the same Expected Value of 11.1; a 9.25% chance of a 120-minute outage and a 90.75% chance of no outage, as displayed in Figure 3.

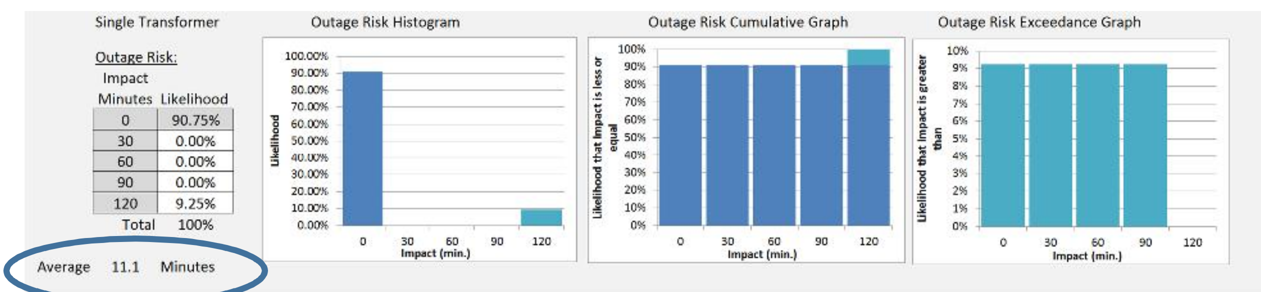


Figure 3 - A Different Risk With the Same Expected Value

A Standard for Calculating Probability Distributions: Monte Carlo Simulation

The Joint Intervenors cite MAUT as the basis of their proposed MAA (notwithstanding the point we made earlier in this document how MAA deviates from MAUT regarding risk aversion). While MAUT is not unheard of in quantitative risk analysis, by any measure we could think of, Monte Carlo simulations are much more widely adopted. Here are a few facts about the relative impact of these methods as standards. [vi]

- There are at least 30 different MC tools on the market, one of which claims to have 150,000 users and 93 of the Fortune 100 as clients as well as several government agencies and not-for-profits. (Furthermore, recent advances have now made this technique available to virtually all users of Microsoft Excel.)

- There are far more published articles in peer-reviewed research journals for Monte Carlo as a risk assessment tool than there are for MAUT[vii].
- Monte Carlo simulations are required training for the Society of Actuaries (MAUT is not a topic of any text on the required reading list).
- Monte Carlo simulation is a frequent topic at the American Statistical Association (well over a dozen sessions at JSM2016 in Chicago, where MAUT was not mentioned once).

There are reasons why MAUT is not the standard and Monte Carlo is. As we saw earlier, the use of Expected Values prevented the proper aggregation of risk across a group of only three individuals [Presentation p. 15]. In the insurance industry the aggregation of risk across groups is paramount, so they have long used a probabilistic approach.

The profession most associated with proper quantitative risk analysis, actuarial science, has known the value of using Monte Carlos for many decades. A 1962 article from the Society of Actuaries describes the problem of determining the probability distribution, *not* Expected Value "of the annual claim cost of a given group of lives for a given year[viii]." The article concludes that

> "The analytical solution of this problem would be extremely complex, and indeed any such solution which would be practical from a cost standpoint would necessitate making simplifying assumptions which would raise considerable doubts as to the validity of the conclusions. Therefore, it was decided to use the Monte Carlo technique, which is admirably suited to a problem of this nature."

These methods work by simulating uncertainty by generating random numbers on a computer, much like rolling dice. Hence the code name, Monte Carlo, bestowed when the technique was invented in 1947 as part of the Manhattan project. It is noteworthy that Monte Carlo was already chosen as the method of choice by the Society of Actuaries at a time when computers were millions of times less powerful and many thousands of times more expensive than they are today.

On page 102 of the Decision, MAA is compared to Monte Carlo as "an analytical (rather than simulation) approach that concentrates on the impacts of hazards on utility assets and potential consequences." As stated above, in 1962, aggregating risks of these sorts is far too complicated to be determined analytically and requires simulation[ix]. We would also argue that risks associated with utility assets are even more complicated than those in the life insurance industry.

So, the use of simulation is not just an opinion or choice, but often a mathematical requirement, at least in more realistic models of risk. As one of the original developers of both MAUT and Monte Carlo simulation stated:

> **"**This method (Monte Carlo) is designed to deal with problems of a more complicated nature than conventional methods…"[x] *-John von Neumann*[7]

[7] John von Neumann was the father of the modern computer and also one of the original developers of Utility Theory.

In other words, the forefather of MAUT itself would disagree with the characterizations of simulations by the Joint Intervenors.

Monte Carlo simulation is also explicitly recommended in the Hubbard, Evans paper referenced[8] by the Commission, which states that organizations:

> "…should use Monte Carlo simulations to explicitly model the relationships of components in a system and their correlations. In other words, instead of just adding or multiplying abstract factors, realistically descriptive functions are defined."

Since 1986, Monte Carlo add-in software has been available for electronic spreadsheets. The cost and ease of use of simulations has continued to improve dramatically since then. Now, thousands of scenarios per keystroke can be run in native Excel – quickly and with no specialized training or additional software.

Figure 4 displays an example of the results of a Monte Carlo simulation of 1,000 trials in native Excel, which can aggregate total hours of outage for between 1 and 10 independent transformers as modelled above. The results of 2, 5 and 10 transformers are shown.

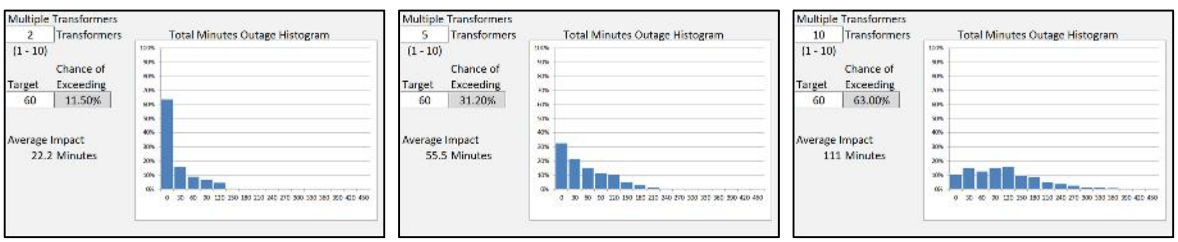


Figure 4 - Simulations of Two, Five, and Ten Transformers

Notice that the Averages (Expected Values) are 2 x 11.1 = 22.2, 5 x 11.1 = 55.5, and 10 x 11.1 = 111 minutes. This would be the only result returned from MAA. With simulation, however, the entire histogram is available.

What we discussed so far are just a few of the limitations of both MAUT and the version of it referred to as MAA including some important flaws discovered by other researchers. See the Appendix 1, "Further Comments on the Limitations of Multi-Attribute Utility Theory."

[8] Hubbard, Evans "Problems with Scoring Methods and Ordinal Scales in Risk Assessment" *IBM Journal of Research and Development* 2010 Vol. 54, Number 3, Paper 2, May/June 2010

A Path Toward More Probabilistic Modeling

The utilities have already been adopting probabilistic risk assessment methods that not only address the weaknesses of the previous methods used by the utilities but go beyond what is proposed by MAA. Both SDG&E and PG&E have already conducted – in limited cases – quantitative risk assessments using Monte Carlo simulation, empirical data, subject matter expert estimates, and explicit risk tolerances. For example, SDG&E is already using simulation models in the area of wildfire and aviation risk to drive multi-million-dollar mitigation investment decisions. PG&E's Electric Operations conducted three probabilistic pilots, performed by Hubbard Decision Research, and PG&E's Gas Operations created numerous probabilistic proof of concept models under a contract with AnalyCorp, Dr. Savage's consulting firm.

The path forward has been made easier with the development of new probabilistic approaches. In 2014, open standards were established for leveraging the new power of Excel as a Monte Carlo simulation platform. 2016 saw the introduction of free tools that allowed Excel users to create their own simulations. Thus the capability to perform powerful Monte Carlo is becoming ubiquitous.

SDG&E Wildfire and Other Risk Models

SDG&E has undertaken risk assessments using probabilistic methods, and is expanding its efforts on this front – for safety risks, as well as energy and project risks. This was done without direct involvement of Hubbard or Savage and we think this makes the case for how these methods can be used directly by the utilities themselves with minimal outside coaching.

As early as 2012, SDG&E began developing a highly quantitative model for assessing wildfire risk and the effectiveness of mitigation efforts. The resulting work product is currently the main tool to prioritize millions of capital dollars. Other efforts have been undertaken. Recently, a portion of SDG&E's aviation risk was analyzed to determine the effectiveness of a particular mitigation effort. Additionally, it is now common practice for SDG&E Enterprise Risk Management to review project or credit risk using probabilistic approaches that consider uncertainty of events and have a range of possible outcomes as outputs. Probabilistic approaches allow for the organizations to adjust the appropriate time horizon and risk tolerance for each risk, as well as provide a breadth of knowledge regarding the range of outcomes. SDG&E is currently undergoing pilot studies for additional safety risks and is determining the most suitable way to bring the risks together into a risk portfolio.

PG&E Pilot Conductor Decision

This example was the basis of the previously-described 91-variable conductor model. Recall that model was used to evaluate whether to replace cable with an underground system or an overhead system where the calculation of the value of information had a significant impact on the results. A highly simplified list of the key components of that model are listed below.

Elements of a Realistic (But Simplified) Description of the Conductor Investments Model:

- Investments to reduce wire-downs will have a duration and cost which is not perfectly predictable.

- Investments to reduce wire-downs will have some effect that cannot be known for certain.
- Wire-downs happen with an unknown frequency that varies from year to year.
- Wire-downs have varying costs of repair.
- When a wire-down occurs, it creates an outage for varying number of customers.
- The duration of that outage varies.
- The types of customers affected (the share of residential vs. business) varies.
- The non-revenue (societal cost) of the outage is uncertain.
- Some percentage of wire-downs produce fires, some percentage result in electrocution injury.
- When fires occur they have varying sizes, amounts of property damage, and amounts of injury.

All of these were expressed as ranges of continuous values, expressed as probability distributions, not binary yes/no outcomes or expected values. The specific distributions were based on multiple workshops with PG&E subject matter experts and, in some cases, historical data. All of these were used as inputs to compute a cash flow over a number of years which was in turn used to compute a net present value.

This allows us to answer questions like "What is the probability that losses due to wire-downs will exceed $5 million over the next 10 years?" or ""What is the probability the NPV will be negative?" Since the Monte Carlo simulation can produce an entire distribution (as in Figure 5 below) and Loss Exceedance Curve, any question like this can be answered.

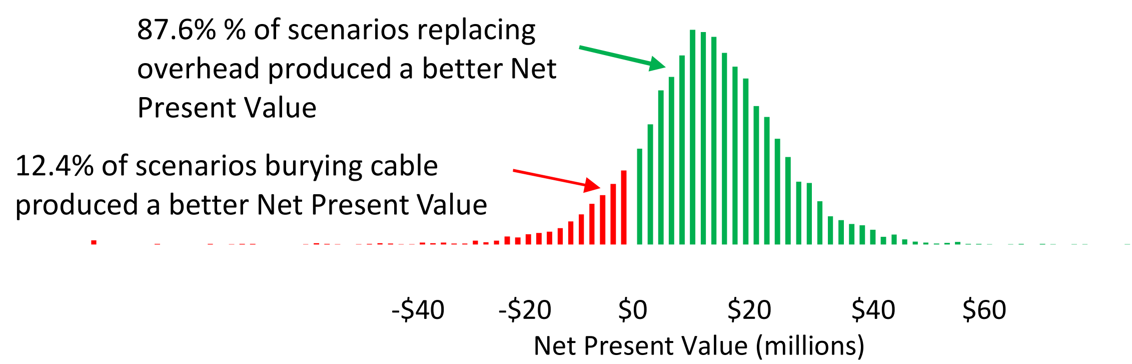


Figure 5 – An example of a Probability Distribution. The height of the bars indicates their relative likelihood.

It is precisely in these situations, however, where analytic solutions simply *do not exist.* The fact is that Monte Carlo simulations are so widely used because they are applied to problems where there simply is no analytic solution.

PG&E Gas Operations Pilot Studies

Over forty proof of concept models were created using open standard methodologies in native Microsoft Excel without additional software. Several of these models were developed entirely by PG&E personnel, with no prior simulation experience. One model, in particular demonstrates a multi-attribute probabilistic approach to optimal mitigation strategies, which we believe, if fully implemented would satisfy the goals of the Commission.

This interactive simulation model addresses both safety and reliability risk. It allows decision makers to experiment with budget levels and relative weights on safety and reliability, and determine the optimal portfolio of mitigations. This type of model provides a transparent means of conveying risk tradeoffs to stakeholders both inside and outside of the utility. (See Figure 6)

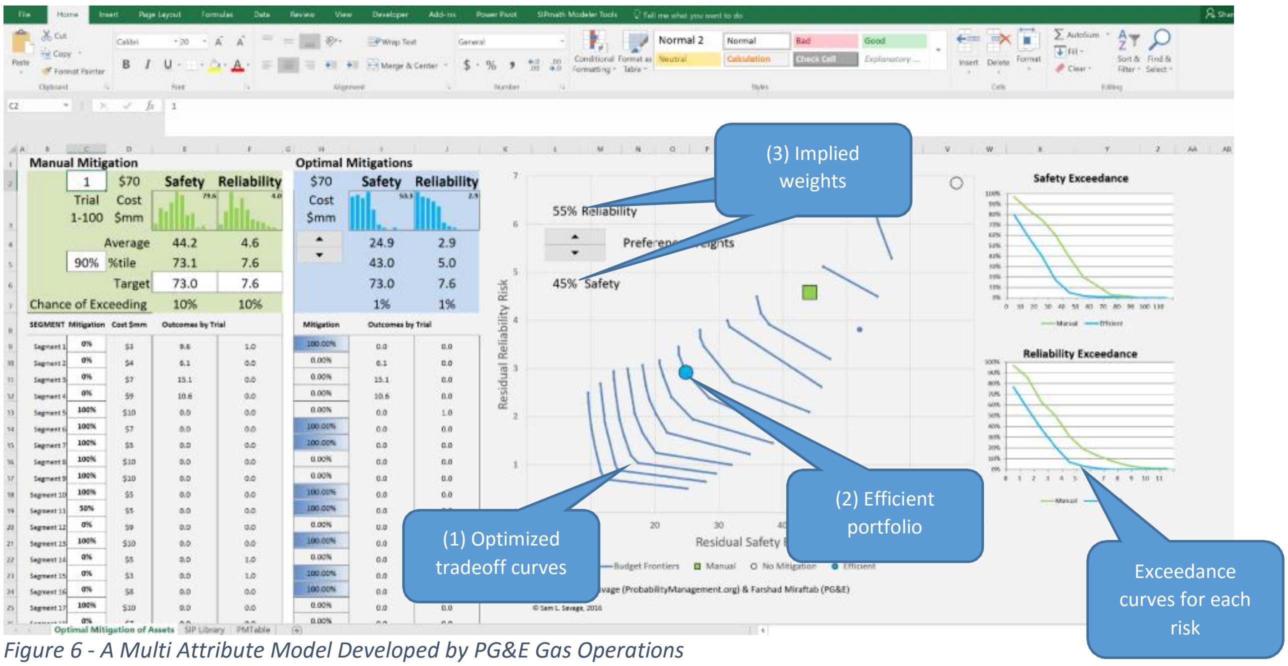


Figure 6 - A Multi Attribute Model Developed by PG&E Gas Operations

The pilot examples described above provide some basic insight into how Monte Carlo simulations are already being used at the utilities. For further explanation about how simulations are actually made see Appendix 2 “Further Comments on the Use of Simulations in Utilities”. This appendix also addresses some of the claims made about simulations by the Joint Intervenors and how simple “Top Down” models can be made fairly quickly while continuing to build more detailed models.

Finally, recent improvements in simulation methodology let simulations communicate with each other. This allows the probabilistic approach to be applied collaboratively at multiple levels within the enterprise. See Appendix 3 “Further discussion on recent important developments in Monte Carlo simulations.”

Summary

As these examples show, the Utilities have already begun using probabilistic methods. The methods shown above showcase situations where oversimplified analytic solutions would not suffice. They make the components of the model much more explicit while avoiding problems of abstract and arbitrary scoring and weighting.

Simulation is a vastly more accepted standard for risk analysis than MAA. A standard approach would be achieved faster thorough the use of widely available tools. In the past few years the full benefits of Monte Carlo simulation have been brought to the ubiquitous Excel spreadsheet without the need of any additional software. Furthermore, open cross platform standards have been developed that let Monte Carlo simulations communicate with each other, greatly facilitating the aggregation of risk across the enterprise.

We believe that a beneficial short-term activity would be to introduce all stakeholders to the benefits of the probabilistic approach, and the activities under way in this area at the Utilities. This would truly move the proceedings into a new dimension, and in so doing, open up new and constructive avenues of dialog between the stakeholders.

Biographies

Douglas W. Hubbard

He is the author of four books: *How to Measure Anything: Finding the Value of Intangibles in Business, The Failure of Risk Management: Why It's Broken and How to Fix It, Pulse,* and his latest book *How to Measure Anything in Cybersecurity Risk*. He has sold over 100,000 copies of his books in eight different languages. Two of his books are required reading for the Society of Actuaries exam prep. In addition to his books, Mr. Hubbard has been published in many periodicals including Nature, The IBM Journal of R&D, CIO, and Information Week.

Douglas Hubbard is the inventor of the Applied Information Economics (AIE) method and founder of Hubbard Decision Research (HDR). Mr. Hubbard's career has focused on the application of quantitative methods to solve current business issues facing today's corporations. Mr. Hubbard's consulting experience totals over 28 years and spans many industries including insurance, financial services, pharmaceutical, healthcare, utilities, energy, federal and state government, entertainment media, military logistics, and manufacturing. His AIE methodology, has received critical praise from The Gartner Group, The Giga Information Group, and Forrester Research. He is a popular speaker at IT metrics & economics conferences all over the world.

He is now serving on the American Statistical Association committee "Scientific Method for the 21st Century" to completely redesign statistical methods fundamental to scientific research. He is also the chair of Decisions and Measurement on the Probability Management board of advisors.

He has consulted to PG&E and is being compensated for this work by PG&E and SDG&E.

Dr. Sam L. Savage

Dr. Sam L. Savage is Executive Director of ProbabilityManagement.org, a 501(c)(3) nonprofit devoted to the improved communication and calculation of uncertainties. He is also author of *The Flaw of Averages – Why we Underestimate Risk in the Face of Uncertainty*, John Wiley & Sons, 2009. Dr. Savage is an Adjunct Professor at Stanford University and a Fellow of the Cambridge University Judge Business School. He is also the inventor of the Stochastic Information Packet (SIP), a standardized data array that lets simulations communicate with each other.

In August of 2015, Dr. Savage volunteered his time working with the Risk Lexicon Working Group Committee of these proceedings.

On October 16th 2015, Dr. Savage announced to the CPUC that he had been asked by PG&E Gas Operations to assist them in better understanding the discipline of probability management, but would not be involved in matters directly pertaining to SMAP. Subsequently he has continued to assist PG&E Gas Operations in exploring a probabilistic approach to risk management.

On October 22nd 2015, Steve Haine asked Dr. Savage for comments on his ALARP white paper which was nearly complete. Dr. Savage did this on a voluntary basis, and also created and presented a demonstration model at a workshop introducing ALARP at the CPUC on Dec. 4th 2015.

In July 2016, through a contract with Davies Consulting, Dr. Savage organized and delivered a 1.5 day workshop on probability management for SEMPRA in San Diego.

In August of 2016, Dr. Savage was asked to assist PG&E and SDG&E directly with these proceedings and is being compensated by them for this report.

Appendices

Appendix 1: Further Comments on the Limitations of Multi-Attribute Utility Theory

In addition to what was discussed in the main body of this document, there are many additional issues related to the use of MAUT and the Joint Intervenors' version of MAUT, which we refer to as the Multi-Attribute Approach (MAA). To understand these problems, it is important to understand where MAUT and MAA come from, how Nobel Prize winning research has contradicted some of the key assumptions, how MAA actually reduces transparency in an important way, and how all of the real contributions of MAA are already captured in simulations but without the use of arbitrary weights and scores.

A Brief Discussion of Background on Multi-Attribute Utility Theory

The formal explanation of Multi-Attribute Utility Theory (MAUT) originated with Ralph Keeney and Howard Raiffa in 1976.[9] They created MAUT as an extension of Expected Utility Theory (EUT), which was developed by Oskar Morgenstern and John von Neumann in 1944. Importantly, EUT, and therefore MAUT, is based on certain assumptions about the preferences of individuals, especially in regard to risks (more on that shortly). While MAA claims that it is based on MAUT, there are several differences including:

1) EUT and MAUT explicitly address risk aversion and shows how this is not modeled by expected values
2) EUT and MAUT describe several constraints on their use which, based only on the description we've seen so far, is not addressed in MAA.

What the Joint Intervenors propose is actually closer to something called "Simple Multi Attribute Rating Technique" (SMART) developed by Ward Edwards[10] in 1977. Edwards developed SMART specifically because he thought MAUT was too complex to be practical. Of course, this also meant giving up on the some of the appearances of rigor associated with MAUT. SMART went through multiple variations[11] and it was these variations, not EUT or MAUT, which used the point scales and weights in a manner more like what the Joint Intervenors propose.

---

[9] Keeney R., Raiffa H., *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Cambridge University Press, 1976

[10] Edwards W. "How to use multi-attribute utility measurement for social decision-making", *IEEE Trans Syst Man Cybern* 1977; 7: 326–340.

[11] Edwards W, Barron F, "SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement" *Organizational Behavior and Human Decision Processes*, 1994

Research Challenging the value of MAA

Ultimately, the stakeholder using SMART or MAUT presumes that decisions are improved in some way or, at a minimum, accurately describe preferences. Research challenges both of these assumptions. Research on effectiveness is important in making the case for any method since there is evidence that there is a type of "analysis placebo" where confidence in decisions can increase even if actual performance does not improve[12,13] (as the previously cited paper by Hubbard and Evans also pointed out). In other words, the perception of benefit is not the same as real benefit. This calls into question the validity of any study which merely asks whether users of a method are satisfied with the results. Individuals who use such methods are likely to perceive a benefit from them regardless of whether the benefit is real.

First, if an equation is presented as a good model of preferences, then we would expect that no matter what inputs we put into that model, the output agrees with the preferences of the stakeholder. This is the goal for MAUT or simpler scoring models like SMART. However, this does not always appear to be the case. In fact, <u>a Nobel Prize in economics was awarded for the development of Prospect Theory which proved that basic assumptions of expected utility theory were wrong</u> - especially where uncertainty is involved.[14] The researchers behind these findings, Amos Tversky and Daniel Kahneman, showed multiple situations where a model may capture preferences in one case but then fails to in another.

For example, Prospect Theory (and common sense) tells us that the same person would value a risk of losing $1 million differently if this were their entire savings vs. if they were already a billionaire. EUT and MAUT make no allowance for differences in starting points like this while the Prospect Theory research showed that people did treat such situations differently. EUT also implies that if a stakeholder was indifferent between two uncertain losses (say, a 50% chance of loss A and an 80% chance of loss B) then the stakeholder would still be indifferent if both probabilities were reduced by half. Again, empirical research shows individual preferences don't follow that model. In short, stakeholder preferences are more complicated than EUT or MAUT allow and certainly more complicated than SMART allows.

Another way to test a scoring model is to show that it actually shows measurable improvements in decisions in the long run over cases. That is, returns on portfolios of decisions are higher, risky events are less frequent and less impactful, or that other objectives are more likely to be achieved. For example, there is quite a lot of research about how estimates are improved by using statistical

---

[12] Heath and R. Gonzalez, "Interaction with others increases decision confidence but not decision quality: Evidence against
information collection views of interactive decision making" Org. Behavior Human Decis. Process., vol. 61, no. 3, pp. 305–326, Mar. 1995.

[13] C. Tsai, J. Klayman, and R. Hastie, Effects of amount of information on judgment accuracy and confidence" Org. Behavior.
Human Decis. Process., vol. 107, no. 2, pp. 97–105, Nov. 2008.

[14] Kahneman, Daniel; Tversky, Amos "Prospect Theory: An Analysis of Decision under Risk" *Econometrica* **47** (2), 1979.

forecasting methods[15] and we support the use of such methods. This is the sort of measurement of effectiveness proposed by one researcher in 2008[16] who proposed setting up long-term studies with multiple persons or organizations to measure real differences in performance. But no such research has been done to date for MAUT or related methods.

The Same Goals Are Achievable Without Resorting to Arbitrary Scores and Weights

We support the idea that we need to build probabilistic models based on evidence and statistically-sound methods. But without showing real, measurable benefits as described by the research shown above, the need for key steps of MAA needs to be questioned. With MAA, stakeholders are required to specify seemingly arbitrary scales and weights. Again, the same paper by Hubbard and Evans[17] cited as a source by both the CPUC and Joint Intervenors[18] points out potential problems with this.

> "Arbitrary features of the scoring scheme itself often have a larger impact on results than the users might be aware of."

The stakeholders are unlikely to fully see the impact of their choices of scoring methods and exact weights at first. In fact, this step would obscure the details of why they chose some scoring methods and weights instead of others. If they chose some of the weights they did because of what boils down to empirical beliefs about the true benefits of reliability, then this should be shown explicitly.

The utilities fully agree with the stated objective of transparency on the thought process of stakeholders. But the method proposed by the utilities is arguably much more transparent than MAA.

First, oversimplified risk models – where risks are binary outcomes with fixed impacts – lose far too much information. As the Monte Carlo simulation cases shown in the body of this document demonstrated, a large number of estimates and their relationships were explicitly modeled. Which assumptions were behind the highly simplified binary risk models of MAA could not be known.

Second, we should not avoid explicit valuation of reliability and safety. The intervenors originally claimed that they did not need to state the value of a statistical life (VSL). However, the application of simple algebra can compute VSL from a weighted score that contains both financial factors and death.

---

[15] William M. Grove et al., "Clinical versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment* 12, no. 1 (2000): 19– 30. 11.

[16] Clemen, R. "Improving and Measuring the Effectiveness of Decision Analysis: Linking Decision Analysis and Behavioral Decision Research" *Decision Modeling and Behavior in Complex and Uncertain Environments.* New York: Springer, 3-31. 2008.

[17] Hubbard, Evans IBM Journal of Research and Development 2010 Vol. 54, Number 3, Paper 2, May/June 2010.

[18] Gonzalez, Ana M. and Pun, Wendy M. and Shintaku, Michael J., "Reporter's Transcript", California Public Utilities Commission, February 5, 2015 pg 1323 to 1326.

In the example shown in the intervenor presentation dated January 25, 2016, a simple calculation shows the implied value of a life to be an exact number. In the example, death and a $500,000 loss are both given “100 points” but they are weighted at 6.25% and 75% respectively. So, according to these scores and weights, a death is equal to as many points as a loss of *exactly* $6 million ($500,000 x 75%/6.25%).

We agree that putting an exact value on something like a human life is difficult. But since anything like the example that MAA provided will produce an exact value, that problem is hardly avoided with MAA. On the other hand, we could put a range on a value like that instead of an exact number. This has already been done in the conductor model described earlier. We used a range that agrees with many of the Value of a Statistical Life (VSL) studies: $5 million to $15 million.

Note that the possibility of monetizing such values is acknowledged by the original creators of MAUT, Ralph Keeney and Howard Raiffa.[19] While Keeney and Raiffa stated that they did not think monetization was possible in all cases, the authors of this paper have yet to find any case in real-world where monetization wasn’t possible. This has been accomplished in models including ones built for measuring drought resilience in African villages, the value of food and water security in African villages, the value of saving endangered species, and reducing the exposure of children to pollutants.[20]

In each of these cases, policy makers were able to agree on a monetary range that represented these values. Ranges even help facilitate this entire process when there are many stakeholders: getting agreement on a range is far easier than getting agreement on a point.

In fact, since the conductor model used ranges for the monetary value of avoiding outages and deaths we are able to reverse-engineer weights for MAA if we wanted to do so. If we arbitrarily define the scoring methods as SafetyScore(Fatality)=100, ReliabilityScore(525 minute outage) = 100 and FinancialScore($500K) = 100, then we could algebraically deduce the weights based in a manner similar to how we deduced the implied value of a life of exactly $6 million in the Joint Intervener example. But since we use ranges, not exact points for the monetary value of safety and reliability, we get ranges for the weights as follows:

- Financial: 4.3% to 4.4%
- Reliability: 5.5% to 10.6%
- Safety: 85% to 90%

But the point of this exercise is to show that we don’t need to ask the stakeholders for arbitrary scores and weights at all. The same expressions of preferences are more directly captured by monetizing reliability and safety. And since we allow for the use of ranges we don’t need to require stakeholders to choose arbitrarily exact values. Most importantly, this makes the valuation explicit and transparent instead of attempting to hide it behind indirect scores and weights.

In fact, simulations can help stakeholders better understand these tradeoffs without ever resorting to scores and weights. Recent research indicates that interactive simulations can help managers make

[19] R. Keeney, H. Raiffa *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Wiley, 1976, pg 125.

[20] D. Hubbard *How to Measure Anything: Finding the Value of Intangibles in Business,* 3rd ed. Wiley 2014.

better decisions in the face of uncertainty.[21] Just as pilots practice in simulators to perfect their skills in unusual situations, decision makers can experiment with interactive models to develop a better understanding of the risk tradeoffs. These sorts of models help illuminate complex non-linear multi-attribute utility functions to facilitate a probabilistic approach to risk mitigation.

One of the authors (Savage) has applied interactive simulation and optimization at Royal Dutch Shell to help executives settle on a common portfolio of exploration projects.[22] Instead of safety, reliability and environmental risks, their portfolio decision hinged on such risks as inadequate short term profits, long-term reserves, percentage mix of oil and gas, and others. Although the stakeholders held diverse preferences across these attributes, the model helped them achieve a compromise by clearly displaying risk tradeoffs that had previously been masked by averages.

These interactive simulations offer another significant advantage as well: when probabilities are not known with precision, assumptions may be changed on the fly. Through this experimentation, the simulation may indicate which circumstances would cause one decision to dominate another. For example, at what probability of a Richter 7 earthquake would reinforcing a hydro power dam be a cost effective mitigation? Simulation may indicate the tipping point. Figure 6 in the main body of this document showed those tradeoffs explicitly.

Appendix 2: Further Comments on The Use of Simulations in the Utilities

In the body of this document we make the case that Monte Carlo simulations are not only a widely accepted, standard approach for quantitative risk analysis, but also a mathematical necessity for most real-world problems being modeled.

There were, however, points in the Decision where the Joint Intervenors were making claims that need further clarification. For example, consider the following points made in the decision document:

> "It is an analytical (rather than simulation) approach that concentrates on the impacts of hazards on utility assets and potential consequences" p. 102
>
> And
>
> "Whenever possible, analysis, real world experience and institutional knowledge is preferable to reliance on simulation. Simulations should only be relied on in those instances where a probability distribution cannot be developed analytically" p. 127

The implication here is that simulations don't consider the impacts of hazards and consequences and that simulations don't involve analysis, real world experience and institutional knowledge. These implications contradict the real-world practice of building simulations and needs to be challenged. First, we will give a simple explanation of what Monte Carlo means. Then we will address what the "analytic"

---

[21] Using Simulated Experience to Make Sense of Big Data, Robin M. Hogarth and Emre Soyer, MIT Sloan Management Review, December 16, 2014.

[22] Probability Management, Sam Savage, Stefan Scholtes and Daniel Zweidler, OR/MS Today, February 2006, Volume 33 Number 1.

vs. "simulation" distinction means, how simulations are actually built in practice, and how simple models can be produced relatively quickly in a phased-rollout.

A Brief Review of Monte Carlo Simulations

When the Joint Intervenors refer to a "simulation" they are presumably referring to the type of simulation using random sampling methods (any algorithm or program that is meant to represent some system in the real world could be referred to as a simulation, even if it is not based on random sampling methods). Generally, the type of simulation which uses random sampling methods is referred to as a Monte Carlo simulation. In short, a Monte Carlo simulation is <u>just a way to do the math when we don't have exact numbers</u>. In fact, it is the only practical way to do that math for all but the simplest problems.

For example, consider an infrastructure investment in a utility that could improve reliability while increasing capacity for growth (possibly reducing costs in the future). Even a simple cost/benefit analysis would likely have a number of variables including project cost, duration, impact on reliability, and even growth in future demand (which affects the benefits of reliability). If we had exact numbers for each of these the math would be simple – a net present value (NPV) calculation over some number of years. The result in this case would also be an exact number for some specific NPV. But almost all of these input variables listed above are uncertain. Since the input variables are uncertain, the NPV must also be uncertain.

What we need to do is compute a probability distribution for an NPV given the probability distributions of the inputs. By doing this we can answer key questions like the chance the investment will not break even. We can perform an actual risk/return analysis and produce the Loss Exceedance Curves mentioned in the body of this document. The Monte Carlo simulation method is a relatively simple approach to solve problems that would otherwise be mathematically complex or even impossible (a point also made by the quote from the Society of Actuaries journal in the body of this document).

In a Monte Carlo simulation, values are sampled from probability distributions – often thousands or tens of thousands of times. Each set of samples is referred to as a "trial" representing one possible outcome. After many thousands of such trials, a new probability distribution emerges for the computed value.

While the individual trials are random, the overall pattern after a very large number of trials is not. So it is not really accurate to say that the output of a Monte Carlo simulation is "just random." As more trials are run, the results converge on a consistently replicable result.

Analytic vs. Simulation

One distinction made in the Decision does not seem to have been adequately defined for the audience, specifically the "analytic" vs. "simulation" distinction. Mathematically speaking, the terms "analytic" or "analytical" simply refers to a procedure that can be executed in a finite number of steps. It is not by itself necessarily a benefit – in fact, it is entirely possible for a solution to be both analytic and completely incorrect.

The term "analytic" should not be confused with the more common meaning of the term "analysis." Analysis refers to an investigation by decomposition into parts – in other words, to look at each of the components of the thing being investigated. Monte Carlo models certainly do that and – we would argue – allow for much more detailed analysis than the MAA. As the 91-variable conductor example in the main body of this document demonstrates, Monte Carlo can allow for a large number of interdependent variables of different distributions.

It is true that a purely analytic solution is certainly possible in the simplest situations like those presented in the Joint Intervenor white paper[23]. In that white paper, all of the risks were presented as simple binary distributions – either something happened or it didn't. And if it did happen, the consequence was an exact amount. There was no allowance for a continuum of possible outcomes based on multiple uncertain variables each with their own continuum of outcomes. What experienced Monte Carlo modelers know is that in any significant, real-world problem, purely analytic solutions are very rare. This point was made clear in the body of the document.

How Simulations Are Actually Built

Having cleared up what the analytic vs. simulation distinction really means, let's address the implications about what simulations do or do not include. Again, the claim in question was that simulations are built without "analysis, real-world experience, and institutional knowledge" apparently unlike MAA which "concentrates on the impacts of hazards on utility assets and potential consequences." These points are stated as if such shortcomings were somehow a feature of the mathematical modeling method itself and not whether it was made by some analyst in isolation. Any method can be applied incorrectly but what the intervenors describe has nothing in common with how the models were actually built. So we need to make the following points.

1. **Simulations Do Use Expert Input**

   In reality, the Monte Carlo simulation examples described in the body of this document were built with direct involvement of several subject matter experts. This was also the case for the previous hundreds of models built for clients separately by the co-authors of this document. Often, this involvement takes the form of a series of working meetings where selected individuals who are knowledgeable of the system being modeled describe the problem, the key variables and their interactions in detail. The model-building is done in direct view of this team. Generally, the modeler is simply asking questions, proposing methods of modeling based on those answers, and then seeking input on that approach. It is more accurate to say that the

[23] Feinstein, Chuck and Lesser, Jonathan., "Intervenor Perspective Regarding an Improved Methodology to Promote Safety and Reliability of Electric and Natural Gas Service in California." On behalf of The Utility Reform Network/Indicated Shippers/Energy Producers and Users Coalition, 2016.

clients build the models and we are simply facilitators with knowledge of the modeling tools and methods.

Real-world experience and institutional knowledge are again directly used when estimates (expressed as probability distributions) are given for each of these variables which the subject matter experts wanted to include in the model. There are only two sources of data in each of the models – the calibrated estimates of experts and empirical data. Initially, the empirical data is limited to what we would call "arm's reach" data. Once the previously described information value calculation is done, we can decide where and how much further measurement is justified.

2. The Analysis of Simulations is More Realistic

All of the simulations described included a large number of interacting variables. Specific cause and effect relationships are modeled. The proposed MAA does not approach the level of analysis needed in real-world situations. The simulations already built at PG&E include specific cause and effect relationships. For example, corrosion increases the chance of leaks of pipes, of which some percentage lead to explosions and some of those may be in high impact areas. Likewise, wire-downs cause outages and the need for repair but some percentage of them cause fires and the damage they cause is a function of where the fires start and how big the fires grow. Breaker failures are more likely for certain types of breakers, and should they fail, different breakers will have greater or less impact. Again, most of these relationships are not simple binary relationships but a range of possible values.

3. The analysis used incorporates both empirical data and expert input

The subject matter experts themselves are actually trained to provide subjective probability estimates and their skill at this task is measured. This calibration method is critical in building models that use expert input even partially. Most experts are statistically "overconfident" in their assessments of probabilities. That is, the probabilities they provide represent less uncertainty than they actually have. Calibration training measurably improves this skill (one of the co-authors of this paper has calibrated over 1000 SMEs in the last 20 years). Without calibrating experts, the inputs they provide will have an unknown error – this will produce problems in decision analysis just as using an uncalibrated instrument in other areas of power production will cause problems.

In fact, we could find multiple examples of the use of Monte Carlo in power utilities – including a comprehensive survey conducted by the CPUC in 2009 which found a number of utilities using simulations for long-term procurement planning.[24] We also found examples of intervenors in other

---

[24] CPUC "Survey of Utility Resource Planning and Procurement Practices for Application to Long-Term Procurement Planning in California" April 2009.

locations using Monte Carlo to make their case against a utility[25] and in one where the utility was found to be in non-compliance with regulations for *not* using Monte Carlo simulations.[26]

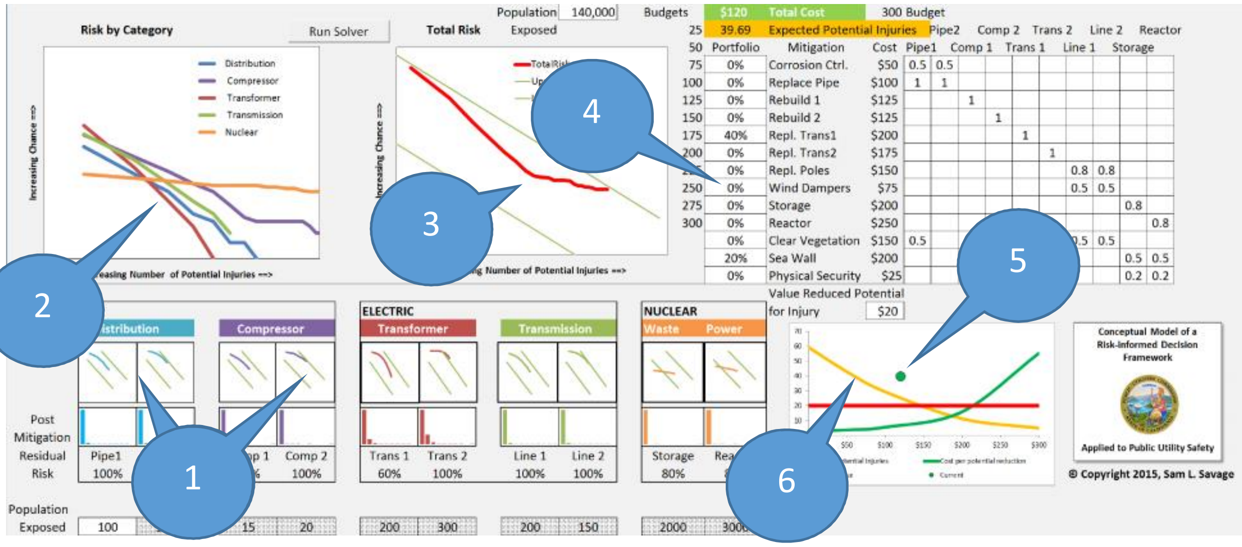
Using Top-Down Modeling: Simple Fast, More Realism Over Time

Note that the models that have already been built for the utilities were to solve specific problems and may be more detailed than what is initially needed. So it would be inaccurate to extrapolate from this experience and conclude that building proper probabilistic methods for the utilities would take years before any benefit from the effort could be seen. On the contrary, the best alternative to a complex probabilistic model is just a simpler probabilistic model.

The Utilities can take a "top down" approach that builds simple models across sets of different problems in the organization. Then those models can evolve over time to be more detailed. The best way to determine where more modeling detail (and more empirical measurements) should be targeted is by using the information value calculations as discussed in the body of this text. Remember, areas of a model with high information value are those areas where further uncertainty reduction through more detailed models or empirical measurements are more likely to improve decisions. The information values not only guide us to which parts of the model require further development, but also how to prioritize those areas. If we use the proper quantitative methods, we can always roll-up the risks of individual components of the model to portfolio-level risks no matter how detailed different parts of the overall risk model become. On December 4th of 2015, Dr. Savage presented a working conceptual model of such a top down approach along with an accompanying PowerPoint presentation at a workshop at the CPUC, as shown below.

[25] "Restoring Power: How Lawmakers Can Lower Your Electric Bill" by David G. Tuerck, Ph.D, Paul Bachman and Michael Head JANUARY 2015 Yankee Institute for Public Policy.

[26] State of New Hampshire Public Utilities Commission, Order Accepting Integrated Resource Plan, 2009.



The model features a number of exceedance curves for individual utility assets (1) which progressively roll up to asset families (2), and an eventual Total Risk exceedance curve (3). A portfolio of mitigation strategies can be chosen manually (4), resulting in a cost (X axis) and residual risk (Y axis) shown by the green dot (5). When optimization is applied, an optimal risk cost curve can be calculated in Yellow (6).

Appendix 3: Recent Advances in Monte Carlo

In 1986, Monte Carlo add-in software first became available for electronic spreadsheets, and this type of software probably now has over 100,000 users. However, it took special training to run. Furthermore, the results of these sorts of simulations could not be easily aggregated. That is, one could simulate the safety risk of a single gas pipeline segment, but to simulate the total risk across 100 segments would require a model 100 times larger that would simulate them all at once.

In 2010, significant Improvements in Microsoft Excel allowed extremely powerful Monte Carlo simulations to be performed without any additional software, raising the potential number of users of simulation to over 100 million.

In 2013, ProbabilityManagement.org was incorporated as a 501(c)(3) non-profit to promote the use of probabilistic modeling by leveraging the increased capabilities of Excel. Dr. Sam L. Savage, Executive Director, is joined on the board of directors by Nobel Laureate Harry Markowitz. The non-profit receives funding from Chevron, Lockheed Martin, General Electric, and Wells Fargo, among others.

The discipline of probability management describes the communication of Monte Carlo results between different simulations using data arrays called SIPs, or Stochastic Information Packets. A SIP is an array of Monte Carlo realizations, saved as a column of numbers. These columns can be used to calculate uncertainties using standard mathematical operations. Probability management is:

1. **Additive:** By passing SIPs between models while maintaining their order, simulations may now be aggregated.
2. **Actionable:** Mathematical operations can be performed on SIPs as if they were numbers.

3. **Auditable:** The SIP is an array of data, with provenance, which provides an audit trail. In this way, SIPs increase transparency.
4. **Agnostic:** The open SIPmath standard currently supports Excel, XML, CSV and JSON data formats. This means that probability management may be readily performed in most simulation environments. In particular, Microsoft Excel natively supports SIPmath, which has a user base over 100 million.

In the earlier example involving pipelines, probability management would allow each of the segments to be simulated separately, whereupon their SIPs could be summed up to get the total risk. This also means that statistical experts could produce SIPs for use by non-experts in their own simulations, just as experts in power generation produce electricity for non-experts with light bulbs.

In 2014, the open SIPmath standard was established by the non-profit to convey SIPs in XML, CSV, and XLSX format to create a fully cross platform environment for sharing results between virtually any Monte Carlo simulations. 2016 saw the introduction of free tools that allowed Excel users to generate their own random variables. Thus the capability to perform powerful Monte Carlo is becoming ubiquitous.

The field of probability management is new, and is recognized as a "transformational" technology by a world leading technology research firm, Gartner, Inc.[xi] Much of the probabilistic work that is being done at Gas operations within PG&E would likely not have been possible before advances made in 2016.

---

[i] Sam L. Savage, The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty", John Wiley & Sons, 2009.

[ii] Douglas W. Hubbard, The Failure of Risk Management: Why it's Broken and How to Fix it. John Wiley & Sons, 2009.

[iii] Preliminary SED Staff Whitepaper on As Low As Reasonably Practicable (ALARP) Risk-informed Decision Framework Applied to Public Utility Safety, Steven Haine, P.E. California Public Utilities Commission, Nov. 9, 2015.

[iv] Probability Management, Sam Savage, Stefan Scholtes and Daniel Zweidler, OR/MS Today, February 2006, Volume 33 Number 1.

[v] http://geohazards.usgs.gov/hazardtool/application.php

[vi] It should be noted that we don't believe that the particular approach proposed by the intervenors is even consistent with proper MAUT.

[vii] Searches of the JSTOR database in risk-related journals and articles produce 8,622 hits for Monte Carlo vs. 619 for MAUT over the same time period. (MAUT included searches on related terms, like Multiple Objective Decision Making, Multiple Objective Decision Analysis, etc.).

[viii] Russell m. Collins, Jr, "Actuarial application of the Monte Carlo technique", Transactions of Society of Actuaries 1962 vol. 14 pt. 1 no. 40.

[ix] Lucas, T.W, Kelton, W.D, Sánchez,P.J,Sanchez, S.M., Anderson, B.L. 2015. "Changing the Paradigm: Simulation, Now a Method of First Resort". Naval Research Logistics DOI 10.1002/nav.21628: pp 293-302.

[x] Von Neumann on Monte Carlo in seminal paper "Statistical Methods in Neutron Diffusion: With J. von Neumann and R. D. Richtmyer (LAMS-551, April 9, 1947).

[xi] https://www.gartner.com/doc/3388917