Corps Risk Analysis Gateway Training Module

Uncertainty

Series:
Corps Risk Analysis Online Training Modules

Prepared by:

Institute for Water Resources
US Army Corps of Engineers

Contents

Introduction

This module was originally developed as a web-based training on the Corps Risk Analysis Gateway. The content has been modified to fit this format. Additional modules are available for download on the IWR website.

The purpose of this training module is to acquaint you with the concepts of knowledge uncertainty and natural variability, which together constitute "uncertainty" and accompany many decision problems the US Corps of Engineers (Corps or USACE) faces. After completing this module you will be able to:

- Discuss the role of uncertainty in risk
- Identify two distinct levels of uncertainty
- Define the two primary causes of uncertainty
- List examples of quantitative values that can be uncertain
- List causes of uncertainty in empirical quantities
- Describe in general terms how probability distributions can be used to represent uncertainty

You are encouraged to read through all of the examples provided in this module, which look at specific concepts in more depth.

This training is approximately one hour.

This course includes a self-assessment; it's recommended that you be able to achieve 70% for successful course completion.

Chapter 1 - Uncertainty

1.0 Uncertainty

Risk analysis requires analysts to separate what they know from what they do not know. This methodology has a foundation in two principles:

1. Assessment of risks should be based on the best evidence and scientific methods available.
2. Risk analysis should focus appropriate attention to the unknown, which could affect decisions and decision making outcomes.

Thus, uncertainty occupies a central position in the risk analysis paradigm. One must make the most realistic assessment possible based both on what is certain and what is uncertain.

When analysts are not sure about any aspect of their work they are uncertain. In general, uncertainty derives from one of two sources:

- Knowledge Uncertainty: knowable facts that we may not know, but we could potentially research the fact to find out.
- Natural Variability: natural variability in the universe that may prevent us from knowing a value even when we have sufficient data and facts.

1.1 Uncertainty and Risk

Why are there risks? Because there is uncertainty.

We do not know how future events will turn out. Both the probability and consequence of most risky events are uncertain.

Example: Uncertainty and Risk

Risk is often referred to by the equation: Risk = Probability x Consequence, the two elements shown here:

Figure 1. Uncertainty in Risk

This equation is not a literal definition of how you calculate risk. It is more of a conceptual model that makes an important point. If either element of the risk is zero, there is no risk.

So, for example, if the probability of an event is very high but there are no consequences, there is no risk. Likewise if the consequences are large but there is no probability that they will occur, there is no risk. Risks exist because of uncertainty.

For example, there are things about the probability that we don't know and there are things about the consequence that we don't know. The probability is often uncertain because there may be facts, just knowable facts, about the probability situation that we simply don't know. These we call knowledge uncertainty.

And because we deal with complex natural systems, there is natural variability. Take a stream flow, for example. We might not know what is the mean high daily flow or the peak annual flow. Projecting exceedance frequencies for floods would be an example of knowledge uncertainty. But it is also an example of natural variability. Even if we did have 100 years of stream gage record, that would be of little use to us in trying to predict what the highest flow is going to be in the coming year.

In a similar way, consequences are affected by knowledge uncertainty and natural variability as well. We might not know how many houses or buildings are in the floodplain - that would be a knowledge uncertainty. But even when we know that, the amount of damage would be extremely variable.

A produce operation, for example a wholesaler, may bring in truckloads of produce (assembled deliveries for stores in the area). On one day there may be millions of dollars of potential damage and 24 hours later there may be virtually zero. Natural variability is also going to plague our estimates of the consequences of risks because they are simply factors that can always change.

Because there is knowledge uncertainty and natural variability in the world, and because these afflict both consequence estimates and probability estimates, our risks, our risk elements, our risk estimates, are all by nature uncertain.

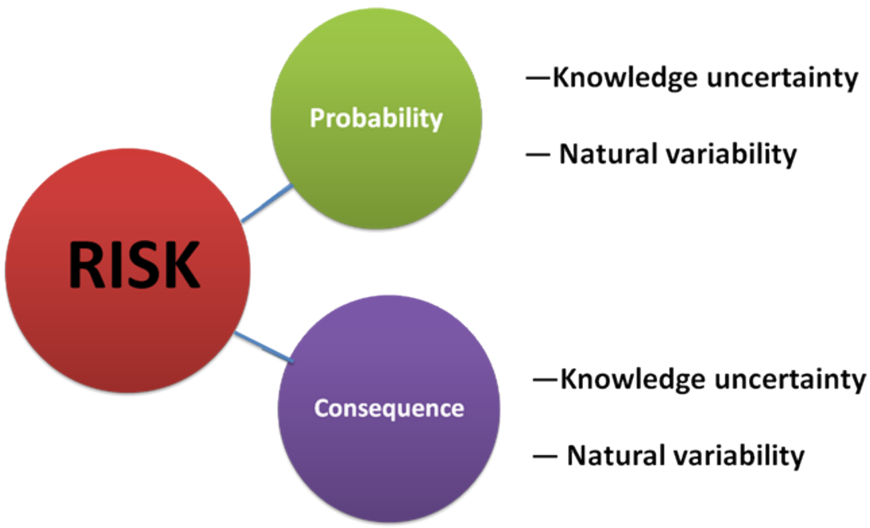


Figure 2. Elements of Risk

1.2 Two Levels of Uncertainty

Uncertainty occurs at two distinctly different levels of resolution: macro and micro. Macro-level uncertainty is typically the culmination or broad understanding of the micro-uncertainty and overall how does this uncertainty involve social values and policy level decisions. Micro-level uncertainty occurs at the level of the analyst's desktop and can be traced more directly to the source of uncertainty. These levels present distinctly different challenges to the Corps.

At the macro-level, the Corps operates in an uncertain and changing social environment. It is not always possible to know information such as:

- Whether the priorities of Congress will change

- What other Federal agencies will say or do
- How local residents feel about an issue
- How variability in climate will affect a project

At the micro-level, Corps analysts and decision makers every day are faced with many uncertainties such as:

- Lack of data
- Incomplete theory
- Imperfect models
- Unknown values
- Inherent variability of physical systems

Discussion: Two Levels of Uncertainty

One of the emerging constants in the modern world is uncertainty. Growing social complexity and an increasingly rapid pace of change are normal parts of the decision making landscape and they contribute a great deal to the uncertain environment in which we all operate. The world grows more complex.

Think of complexity as I use it here in a social sense. It refers to things like the size of a society, the number of parts, the distinctiveness of those parts, the variety of specialized roles that it incorporates, the number of distinct social personalities that are present, and the wide variety of mechanisms for organizing these into a coherent functioning whole.

We face an increasingly rapid pace of change in almost every arena. Scientific breakthroughs make commonplace things that were once impossible to conceive.

Much of this change is driven by rapid advances in technology. That technology changes social values and beliefs as well as the way we live and work. We see rapid increases in social economic and technologic connectivity taking place around the world. A computer virus spreads around the world in hours. A human virus spreads in weeks or months.

Relentless pressure on cost is now a fixture in all public decision making and patterns of competition are becoming unpredictable. As a result of these and other changes, we've entered a world where irreversible consequences unlimited in time and space are now possible. A new phenomenon of known unawareness has entered our lexicon. Everyone refers to Donald Rumsfeld's "unknown unknowns."

Despite the world's rapid advances in all kinds of sciences, we're increasingly dominated by public perception. Public perception is a palpable force. Risks in uncertain situations have a social context and possibility is often accorded the same significance as existence in the public's view. This view can find its way into public policy as well.

Responsibility in this more connected world has become less clear. Who has to prove what? And what constitutes proof under conditions of uncertainty? The future is fundamentally unknowable. We all live and operate in this uncertain reality. Decision making needs a culture of uncertainty and risk analysis provides just such a culture.

1.3 Two Sources of Uncertainty

Uncertainty = knowledge uncertainty + natural variability

Uncertain Knowledge

Knowledge uncertainty, also known as “epistemic uncertainty,” derives from a lack of knowledge on the part of the observer. Knowledge uncertainty exists when a true fact or a numerical constant exists, but you do not know it for any reason at all. It arises from situations such as:

- An incomplete understanding of a system
- An incomplete theory
- Modeling limitations
- Limited data

Knowledge uncertainty can be reduced, but it may be difficult or expensive to do so.

Natural Variability

Natural variability, also known as “variability,” or “aleatory uncertainty,” results from inherent variability in the physical world. It arises from factors such as:

- Random processes that produce variability of a quantity over time and/or space
- Random processes that produce variability among members of a population
- Natural, unpredictable variation in the performance of the system under study

Natural variability cannot be reduced or altered by obtaining more information, although collection of more information may improve our estimation of the natural variability that exists.

Example: Knowledge Uncertainty

1. The U.S. Army Engineer Research and Development Center (ERDC) conducted an extensive literature review focusing on the effects of woody vegetation on levees (http://operations.usace.army.mil/flood/pdfs/Vegetation-Levees-FactSheet.pdf). The study found a great deal of knowledge uncertainty concerning the effects of woody vegetation on levee stability:
    - Hydrologic data for small streams may not exist, so fundamental flow parameters are unknown.
    - There is a general lack of experimental data to characterize new engineering materials and processes.
    - Sometimes there is a poor understanding of the linkages between inputs and outputs in an ecosystem restoration project.
2. You may be required to collect information about a project site for your analysis. The values exist, but they may not have been measured or you may not know them. Such data might include:
    - You may be unfamiliar with a specific dam and not know how many tainter gates it has. If, in fact, it has five tainter gates, five is the true value and you have knowledge uncertainty.
    - A Corps lake has an average number of daily visitors in a year. You may not know that number. Even if the data have never been collected and the number has never been calculated there is still a true value for this statistic.

Example: Natural Variability

The Corps works with complex natural and manmade systems that are full of natural variability, such as the following:

- Amount of time it takes a tow boat to complete a lockage cycle
- Number of barges in a tow
- Draft of a vessel
- Peak annual flow on a stream
- Price of a cubic yard of concrete
- Daily number of visitors to a lake

There is also variability in any attribute of a population, such as:

- Strength of the rebar in a concrete dam
- Life of a light bulb in the rest room at a lake

Example: Sunbury

Introduction

As a quick review, let's revisit the two main types of uncertainty: knowledge uncertainty and natural variability. Knowledge uncertainty refers to uncertainty about facts that can be known, but are not currently known by the observer. Devoting resources to research the topic of which you are uncertain can lead to the reduction and, possibly, the elimination of knowledge uncertainty. In other words, spending time and money on the subject that is uncertain can lead the researcher to an exact and factual answer.

Natural variability, on the other hand, cannot be reduced or eliminated. This type of uncertainty stems from the fact that our world is variable and constantly changing. Because of this variation that is inherent in the physical world, there are issues that we won't be able to provide concrete answers for (regardless of how much time and money we devote to researching them). The best we can do in issues of natural variability is to improve our estimation.

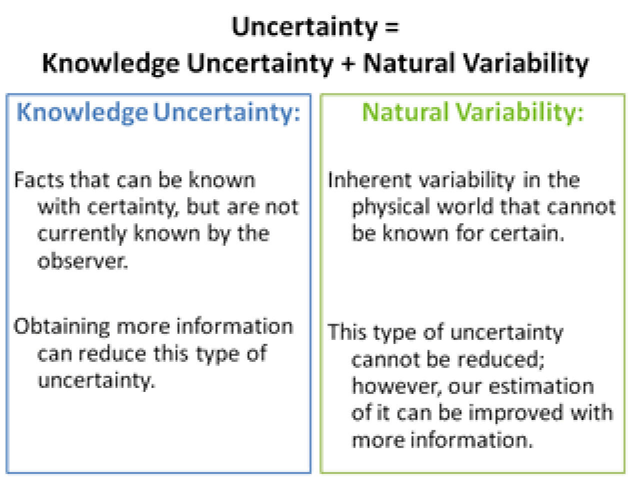


Figure 3. Knowledge Uncertainty and Natural Variability

Examples of knowledge uncertainty include uncertainty about things like mean temperature, median peak daily flow, and the findings of previous research on any given topic. These are all things that can be known exactly and with certainty, even if we don't know them at the outset of our project.

Examples of natural variability include uncertainty about things like tomorrow's temperature, the variation in mean temperatures from day to day, and next week's median peak flow. These are things that will be constantly changing as a result of our world's nature of variation. Although we can guess or estimate these values, we cannot provide an exact and certain answer for them.

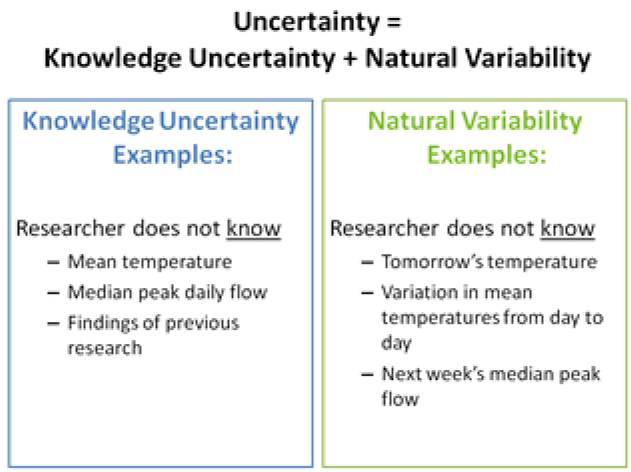


Figure 4. Given Examples of Knowledge Uncertainty and Natural Variability

To illustrate these topics regarding uncertainty in risk analysis, let's consider the town of Sunbury, Pennsylvania. The Susquehanna River runs through the town of Sunbury, putting it at risk for flooding. In case you're interested, here's a little trivia about the Susquehanna River: it is 464 miles long, making it the longest river on the American east coast. It's the 16th largest river in the United States and the longest river in the continental U.S. with no commercial traffic. Additionally, it drains 27,500 square miles into three states and discharges into the Chesapeake Bay.

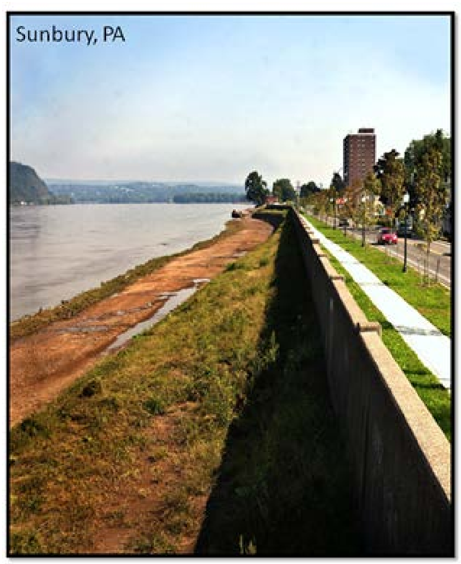


Figure 5. The Susquehanna River near Sunbury, Pennsylvania

In June of 1972, tropical storm Agnes caused the waters of the Susquehanna River in Sunbury, Pennsylvania to rise and tested the town's flood protection measures. Perhaps the most memorable of these measures is the Corps' floodwall. In the picture shown here, you can see the water lapping at the top of the floodwall.

Figure 6. Tropical Storm Agnes in Sunbury, Pennsylvania

The very next day — June 24, 1972 — the town of Sunbury showed its appreciation to the wall. As you can see here, the message speaks for itself.

Figure 7. Message Written on Wall Next Day

Sunbury - Peak Flow (Part 1)

Further considering, let's pose a question: What is the mean daily peak flow of the Susquehanna River at Sunbury, PA?

Now, you have to be sitting there thinking: I have no idea. This is a bit of information that goes into the set of things that we don't know. It's uncertain. But is it knowledge uncertainty or is it natural variability?

Well, the starting point would be: is there a truth out there – is there a true mean daily peak flow on the Susquehanna River? Clearly, the answer is yes. Therefore, there is knowledge uncertainty about that value.

What might be confusing when you try to think about this is the peak daily flow. Now, pay attention to the language. We are not focusing on the "mean" peak daily flow. The peak daily flow is highly variable. It varies from one day to the next. And that value we don't know because of natural variability.

The "mean" daily peak flow is a constant. There is no natural variability in that number. It is a constant; it is true. We just don't know what it is. So let's estimate it using @Risk.

Figure 8 on the next page illustrates what @Risk might look like if you were examining this question about daily peak flow.

You'll notice in this graph of a uniform distribution (in blue) that on the left you see the parameters. There's a minimum – we've estimated the minimum flow is 10,000 cubic feet per second (cfs) – and a maximum – we've estimated the maximum is 50,000 cfs. The rectangular block that you see in the graph is a representation of the uncertainty. The mean peak flow is somewhere between 10,000 cfs and 50,000 cfs. The fact that this is a rectangle with a uniform height is suggesting that any number in that range is as likely as any other. This is expressing the maximum uncertainty and suggesting we have no specific idea and are guessing.

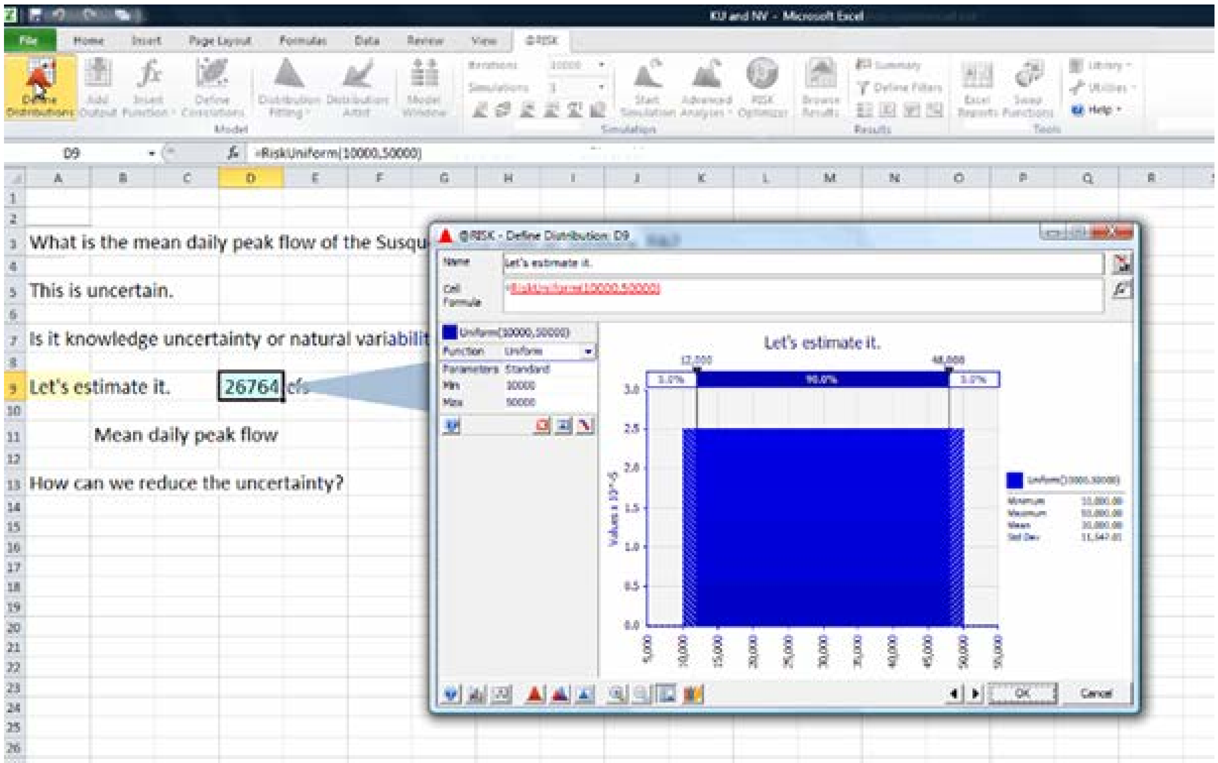


Figure 8. Calculating the Mean Daily Flow

We will learn to use probability as the language of uncertainty and variability as we move along in this risk analysis paradigm. In this example, the estimate of the mean daily peak flow is 43,120 cfs.

Sunbury - Peak Flow (Part 2)

As numbers are randomly selected from the distribution, a variety of different estimates are produced. This is because this is a value that’s in that set of things we don’t know. We have knowledge uncertainty. We don’t know the mean value. How can the uncertainty be reduced? One logical response would be to use the data.

We have data from the Susquehanna River gauge maintained by USGS. We have 26,834 daily observations from October 1937 through almost the end of November 2011. When we calculate the mean, we get a mean flow of 27,175 cfs with a standard deviation of almost 33,000 cfs. Our knowledge uncertainty is now gone. There is no longer any knowledge uncertainty surrounding the mean daily flow. We have the facts and we have reduced the uncertainty.

Table 1. Calculations Using Observations from Oct 1, 1937 through Nov 25, 2011

Count	26,834
Mean	27,175 cfs
Standard Deviation	32,953 cfs

But let's consider the other problem: What will be the daily peak flow tomorrow? We have all the facts now about the mean. The knowledge uncertainty is gone. We know the standard deviation to a high degree of certainty. But tomorrow's peak flow is one that still eludes us, and that's because of natural variability in the world: even though we know the mean and standard deviation, there's a great deal of natural variation.

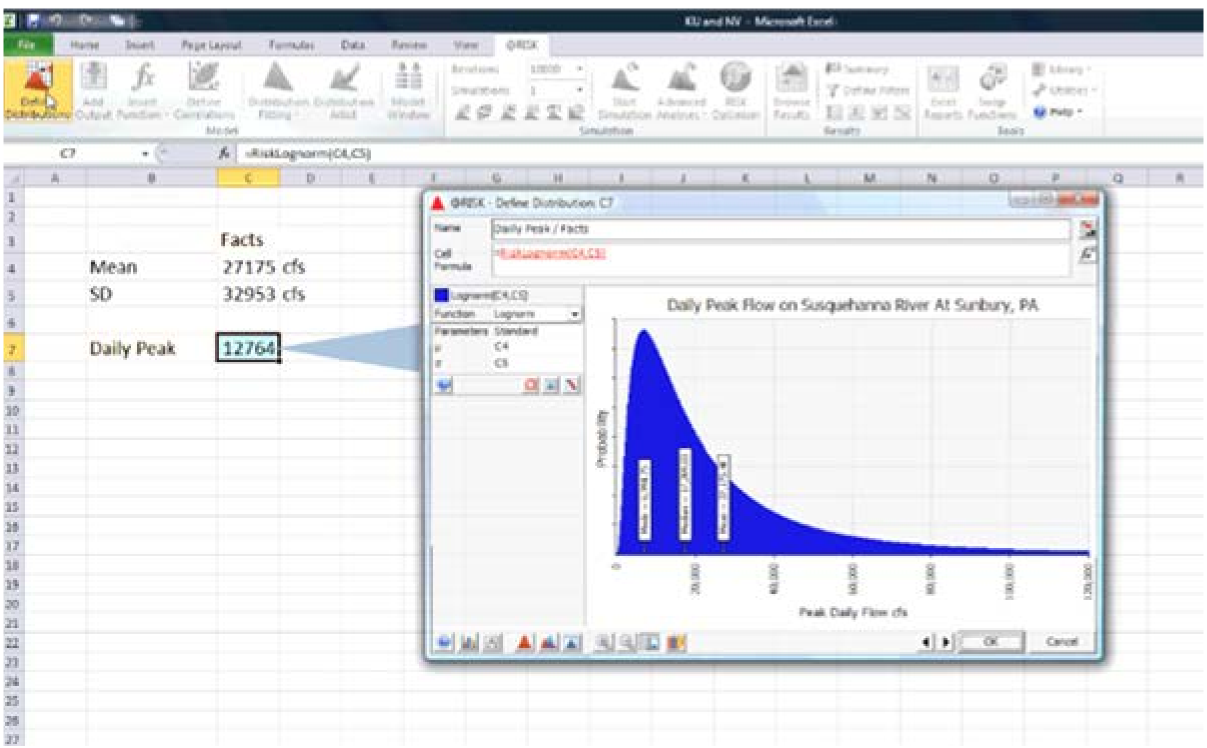


Figure 9. Daily Peak Flow Calculations

If we graph these data as is done in the illustration in Figure 9 above, if the mean is 27,175 and the standard deviation is 33,000, it yields the distribution illustrated in the figure. The mode is almost 7,000 cfs, the most common single daily flow. The median, the flow that's exceeded half of the time, is 17,300 cfs. And the mean, is about 27,200 cfs. On any given day the flow would be any one of those numbers from that distribution. The height of the distribution reflects the relative likelihood of observing that value. So that tail values well off to the right, they occur relatively infrequently. The values in the fat part, the high part of the curve, are the values that are most likely to be observed.

Table 2. Facts

Mean	27,175 cfs
Standard Deviaiton	32,953 cfs
Daily Peak	4,589

So, now you see the daily peak flow is estimated at 4,500 cubic feet per second. It could be 27,000. It could be 19,000. A probability distribution will be used to represent that natural variability. The Corps is quite familiar with this and if you're familiar with frequency curves you understand this very well.

So, in this simple illustration, we've talked about both issues here: knowledge uncertainty and natural variability. Knowledge uncertainty occurs when there is a fact that we don't know. Typically we will use our analysis to try to reduce some of that uncertainty by gathering data. In this case, we were able to gather enough data to feel like we had reduced the knowledge uncertainty entirely. But we were still left with natural variability. So, as you read through this module you will see stated: knowledge uncertainty can often be reduced. If there is a way to go and get the data and discover the truth, we can reduce knowledge uncertainty. But gathering more data is not going to make the variability go away. If we had another 26,000 data points, we would be no better equipped to predict the flow tomorrow than we are right now.

Example: Two Sources of Uncertainty in a Flood Risk Management Project

Suppose a community with 100 houses must be purchased as part of a nonstructural flood risk management project. We need an estimate of the costs. For simplicity, let's dispense with the complications of reality and assume the cost of the project is simply the cost of the 100 houses, each of which is different and does not change (see Table 3). We begin with no data and the cost of the plan is in the group of things we do not know as a bit of knowledge uncertainty. The true cost is an unknown fact and it is constant.

Thinking about the value of an individual house as a mind experiment, this is not a constant. Each house has a different value. So we begin this problem recognizing there is natural variability in the house values. Furthermore, because we have no data, there is knowledge uncertainty about that natural variability. Thus, the uncertainty at the outset of this analysis is due both to knowledge uncertainty and natural variability. That often happens and is one reason distinguishing the two is so difficult at times.

Table 3: Natural variability in residential structure values

House Values of 100 Houses									
$ 164,440	$ 142,096	$ 145,532	$ 177,127	$ 210,522	$ 60,093	$ 127,646	$ 128,727	$ 126,915	$ 106,376
$ 100,338	$ 103,326	$ 95,514	$ 72,080	$ 112,889	$ 171,822	$ 133,923	$ 116,687	$ 139,817	$ 124,073

$ 120,266	$ 155,576	$ 150,354	$ 142,339	$ 85,189	$ 172,617	$ 159,524	$ 115,638	$ 106,371	$ 203,075
$ 224,298	$ 161,243	$ 75,769	$ 174,455	$ 170,791	$ 97,741	$ 142,127	$ 141,764	$ 157,791	$ 173,773
$ 209,130	$ 105,239	$ 141,777	$ 114,693	$ 84,110	$ 188,728	$ 155,252	$ 233,973	$ 169,976	$ 130,085
$ 145,756	$ 152,926	$ 173,530	$ 222,267	$ 202,695	$ 160,931	$ 206,910	$ 209,136	$ 93,802	$ 104,709
$ 159,474	$ 80,379	$ 105,038	$ 110,281	$ 189,800	$ 203,081	$ 139,920	$ 96,868	$ 109,548	$ 142,553
$ 160,006	$ 216,130	$ 228,186	$ 134,182	$ 154,949	$ 181,276	$ 173,454	$ 177,881	$ 130,339	$ 204,240
$ 62,225	$ 172,778	$ 165,200	$ 194,453	$ 181,295	$ 71,542	$ 134,315	$ 181,183	$ 124,641	$ 253,245
$ 141,099	$ 207,225	$ 180,906	$ 143,332	$ 125,989	$ 231,911	$ 186,929	$ 169,933	$ 97,933	$ 164,969

Knowledge uncertainty is reducible while natural variability is not. Let's examine those ideas from the perspective of the omnipotent risk analyst who happens to know the true mean value of all 100 houses is $149,849 as shown in Table 4 below. Keep in mind the Corps analysts do not yet know this value. Because a true and constant value, i.e., a fact, exists data can be gathered to try to learn what it is. So, imagine that a random sample is taken and the cost of 35 houses is estimated. This evidence reduces the knowledge uncertainty. It suggests the true value is about $150,566[1], not a bad estimate at all.

Table 4: Summary Statistics for Residential Structure Values

Residential Structure	Mean	Standard Deviation
Sample	$ 150,566	$ 38,528
Population	$ 149,849	$ 42,975

The uncertainty has been reduced but not eliminated because we work with a sample value. Nonetheless we can now assume an average value of about $150,000 per house. The data collection has also produced some evidence about the natural variability in house prices. We'll return to this idea in a moment. For now, let's focus on the knowledge uncertainty.

After a sample, the Corps analysts have an estimate of the true mean but they will not know, as we do, how close it is to the true mean. In time, estimates will be prepared for each house. At that time they will learn the true value is $149,849. At that future point in time all knowledge uncertainty will have been eliminated and the true fact of the mean value of all houses will be known. The project costs 100 times the true mean house value.

The analysis began with a large uncertainty due primarily to knowledge uncertainty. In this instance it was possible to reduce knowledge uncertainty by conducting a sample and gathering some data. This effort greatly reduced the uncertainty. In this simplistic example it was possible to continue to reduce the uncertainty by gathering more data until the value of every house was estimated and the unknown value could be calculated with certainty.

It is not always possible to get to complete certainty, nor is it always necessary or desirable. In reconnaissance studies, for example, it is not unusual to work with large degrees of uncertainty. Cost estimates may be based on a 20 percent level of design detail or less, for instance. Moreover, it is not always going to be possible to gather data to reduce knowledge uncertainty at all.

The key idea for understanding when you are dealing with knowledge uncertainty is to ask yourself if a true value exists and if it is a constant. If the answers are yes, you are dealing with knowledge uncertainty. In this example there was clearly a true constant value, as the omnipotent risk analyst knew.

Knowledge uncertainty is not confined to numerical values. Much of knowledge is unknown and so the notion is extended to include all situations regarding factual matters that are in the group of things we do not know. For example, we may not know whether providing water in a specific quantity and quality in a given place at a particular time will restore either the functionality or the morphology of an ecosystem.

So far we have neglected the standard deviation; let's consider it now. The Corps analysts know the houses vary in value. That variability was clouded by knowledge uncertainty at the outset. Imagine standing in front of the homes and being asked the value of a specific house; you would have to say you don't know. Once the sample is completed, however, your knowledge uncertainty is reduced; we "know" the mean house value is $150,566 with a standard deviation of $38,528.[2]

Consider the table with all 100 house values again. Notice the variability in the house values. Now consider this simple mind experiment. Imagine the value of each house on a lottery game ping pong ball. What will the value of the next ball be? If we repeat that experiment, is that value a constant? Clearly it is not. Each selection will produce one of 100 different values.

Does knowing the sample standard deviation help you predict the value on the next ball? Again, the answer is no. Suppose we know the population standard deviation and the mean from the table of statistics above. Will this help us predict the value of the next ball? No, it will not.

The value of the next ball is uncertain. Analysts new to this distinction between knowledge uncertainty and natural variability are sometimes tempted to reason that the value is going to be something and we do not know it now so that is knowledge uncertainty. This is wrong. The value of the next ball is not a constant before it is chosen, it is a variable. It reflects natural variability. Spending more money to get the remaining 65 cost estimates after the initial sample provides a better estimate of the standard deviation. In fact it eliminates uncertainty about this measure of natural variability. But that fact does absolutely nothing to reduce the variability itself. The houses still have 100 different values because of the "system[3]" that produced housing values in this community. It is important to understand that collecting more data or doing research and analysis will not reduce the natural variability that characterizes housing prices in the community.

Interestingly, once a ball is chosen its value does become a constant. At this point, we may have knowledge uncertainty about the house value on the last ball selected because it is now a true value or fact that can, conceptually, be discovered. The only way to change the existing variability to a more desirable variability is to alter the system that has produced the original variability. Let us suppose the powers that be have decided that a more egalitarian mix of housing is desirable. They intend to enact this decision by making improvements to every house below the average value to bring it up to the average value. This yields the population shown in the table below.

Table 5: New Natural Variability in Residential Structure Values
Due to a Change in the System

House Values of 100 Houses									
$ 164,440	$ 149,849	$ 149,849	$ 177,127	$ 210,522	$ 149,849	$ 149,849	$ 149,849	$ 149,849	$ 149,849
$ 149,849	$ 149,849	$ 149,849	$ 149,849	$ 149,849	$ 171,822	$ 149,849	$ 149,849	$ 149,849	$ 149,849
$ 149,849	$ 155,576	$ 150,354	$ 149,849	$ 149,849	$ 172,617	$ 159,524	$ 149,849	$ 149,849	$ 203,075
$ 224,298	$ 161,243	$ 149,849	$ 174,455	$ 170,791	$ 149,849	$ 149,849	$ 149,849	$ 157,791	$ 173,773
$ 209,130	$ 149,849	$ 149,849	$ 149,849	$ 149,849	$ 188,728	$ 155,252	$ 233,973	$ 169,976	$ 149,849
$ 149,849	$ 152,926	$ 173,530	$ 222,267	$ 202,695	$ 160,931	$ 206,910	$ 209,136	$ 149,849	$ 149,849
$ 159,474	$ 149,849	$ 149,849	$ 149,849	$ 189,800	$ 203,081	$ 149,849	$ 149,849	$ 149,849	$ 149,849
$ 160,006	$ 216,130	$ 228,186	$ 149,849	$ 154,949	$ 181,276	$ 173,454	$ 177,881	$ 149,849	$ 204,240
$ 149,849	$ 172,778	$ 165,200	$ 194,453	$ 181,295	$ 149,849	$ 149,849	$ 181,183	$ 149,849	$ 253,245
$ 149,849	$ 207,225	$ 180,906	$ 149,849	$ 149,849	$ 231,911	$ 186,929	$ 169,933	$ 149,849	$ 164,969

The new mean is higher but of interest to us is there is now less natural variability in this community. The standard deviation has changed from $42,975 to $25,011. The natural variability has been reduced in this example[4] by changing the system. The new system still has natural variability, however, so it has not been eliminated.

---

[1] This result is based on an actual random sample taken from the 100 values. The reader should not expect every such sample to yield such a good estimate.

[2] These two facts coupled with the knowledge that the sampling distribution of the sample mean is normal due to the large sample size equips the savvy Corps analyst with a great deal of

useful information about the natural variability in housing prices. This is not the focus of this discussion, however.

[3] That so-called system includes location, school district, size of house, construction material, wear and tear, landscaping and many other systematic factors and influences on house value.

[4] Note that reducing natural variability is not always an improvement. There may be many instances in the Corps work where increasing natural variability is actually desired. Increased variability in stream flows is often considered a desirable feature of an urban ecosystem, for example.

1.4 Self Assessment

The following quantities are unknown and therefore uncertain. Are they uncertain because of knowledge uncertainty or natural variability?

What is the source of uncertainty for:

1. The value of an acre of land today in the downtown Springfield, Missouri floodplain
   - ○ knowledge uncertainty
   - ○ natural variability
2. The number of utilities crossing a channel to be enlarged to reduce flood damages
   - ○ knowledge uncertainty
   - ○ natural variability
3. The peak daily flow on the Susquehanna River on any random future day
   - ○ knowledge uncertainty
   - ○ natural variability
4. The amount of dissolved oxygen in the river below Tenkiller Dam
   - ○ knowledge uncertainty
   - ○ natural variability
5. The mean high daily temperature of water at a fixed access point below Tenkiller Dam
   - ○ knowledge uncertainty
   - ○ natural variability
6. The number of tainter gates in the entire Portland District
   - ○ knowledge uncertainty
   - ○ natural variability
7. The price of gasoline
   - ○ knowledge uncertainty
   - ○ natural variability

Self Assessment - Answers

The following quantities are unknown and therefore uncertain. Are they uncertain because of knowledge uncertainty or natural variability?

What is the source of uncertainty for:

1. The value of an acre of land today in the downtown Springfield, Missouri floodplain

   ◉ knowledge uncertainty **CORRECT**. *There is a true value that we just do not know.*

   ○ natural variability INCORRECT. *Over time this value may change but today it is a true value that we happen not to know.*

2. The number of utilities crossing a channel to be enlarged to reduce flood damages

   ◉ knowledge uncertainty **CORRECT**. *There is a true value that we just do not know.*

   ○ natural variability INCORRECT. *This value never changes, it is a true value that we happen not to know.*

3. The peak daily flow on the Susquehanna River on any random future day

   ◉ knowledge uncertainty INCORRECT. *The value of a future flow is not constant, it varies from day-to-day.*

   ○ natural variability **CORRECT**. *We could know the mean and standard deviation of the peak daily flow but this information does not help us know that value on any given day.*

4. The amount of dissolved oxygen in the river below Tenkiller Dam

   ◉ knowledge uncertainty INCORRECT. *This value varies from day-to-day, there is no true value to express here. Had we been asked the mean DO or any attribute of DO that is constant and unchanging the answer would be different.*

   ○ natural variability **CORRECT**. *We do not know this value because it changes with the temperature and other characteristics of the waterway.*

5. The mean high daily temperature of water at a fixed access point below Tenkiller Dam

   ◉ knowledge uncertainty **CORRECT**. *There is a true value that we just do not know. If we have a sample of flow data we can estimate it. As our sample size grows we are reducing uncertainty and our estimate may change but the truth does not change.*

   ○ natural variability INCORRECT. *This value never changes, it is a true value that we happen not to know.*

6. The number of tainter gates in the entire Portland District

◉ knowledge uncertainty **CORRECT**. *There is a true value that we just do not know.*

○ natural variability INCORRECT. *This value never changes, it is a true value that we happen not to know.*

7. The price of gasoline

◉ knowledge uncertainty INCORRECT. *The price of gas fluctuates with supply and demand conditions on a daily basis. We do not know the price of gas for the purposes of estimating gas costs because of this variation.*

○ natural variability **CORRECT**. *We could know the mean and standard deviation of gas, we could even have a good forecast but this information does not help us know that value on any given day.*

Chapter 2 - Why is it Important?

2.0 Why is it Important?

Why is it important to be able to tell the two sources of uncertainty apart?

- Knowledge uncertainty can often be reduced by research, collecting data, more analysis, taking a course, hiring an expert, and so on.
- Natural variability cannot be reduced by gathering more information.

Analysts who can identify the source of their uncertainty are much more likely to find an appropriate and effective way of addressing that uncertainty than are those who cannot.

The figure below shows how the sources of uncertainty can be deconstructed. The reason for being able to deconstruct the uncertainty in your analysis is that there are specific tools and techniques that are appropriate for the many different sources of uncertainty.

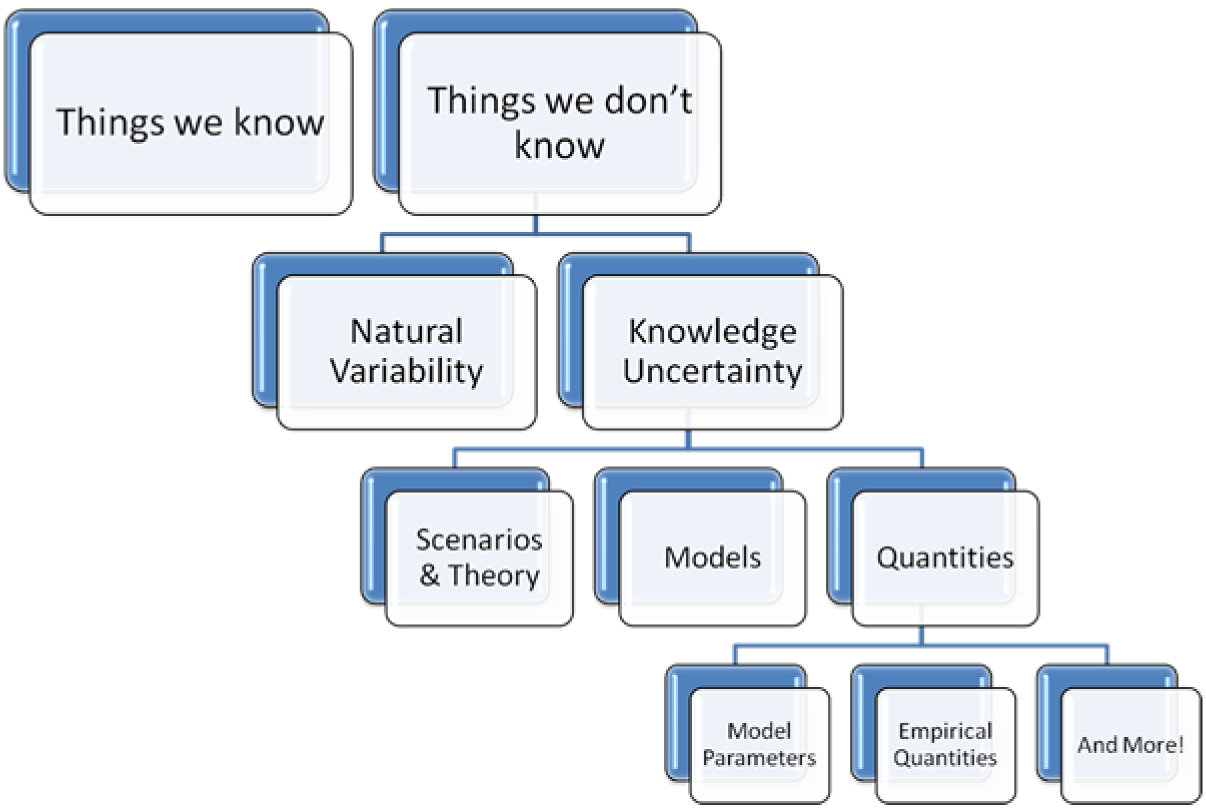


Figure 10: Two Sources of Uncertainty

EXAMPLE: SOURCES OF UNCERTAINTY

We'd like to take a moment to try to understand these two sources of uncertainty a little bit more clearly. So, with any work that you're doing, when we move into this risk analysis paradigm, this risk analysis way of dealing with decision making under uncertainty, we can always make a group of the things that we know. There are always facts that we know. There's information that we have. Likewise, with every project that we're working on, there's a group of things that we don't know.

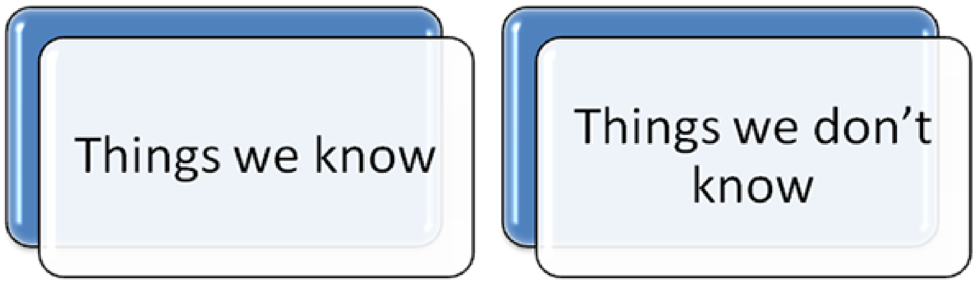


Figure 11: Things We Know and Things We Don't Know

These two sets of information are not always the same size. Sometimes, when a problem walks in the office at 4 o'clock on a Friday afternoon, the set of things that we don't know is very big – there's been an incident down at the lock and no one quite knows what it is, what happened, or has any details – and the set of things that we know is relatively small. But over time, as we get more information, the set of things that we don't know can often be reduced, and the set of things we do know can be increased. Now, in the risk analysis paradigm, the things we know would constitute our evidence, the science, the good analytical work that we're doing. What makes risk analysis different from other decision making frameworks is that it intentionally addresses that set of things that we don't know.

Imagine we've got a set of things we don't know on the floor right now. We can take that set and separate it into two smaller sets. The reason that we don't know things are because of knowledge uncertainty; that means that there are some true facts out there and, for whatever reasons, we just don't know them. So, knowledge uncertainty is intrinsic. It's a characteristic for the analyst. We might say that a number is uncertain, but that number isn't uncertain, it knows exactly what it is. It is we who are uncertain, we don't know what that number is.

Now, sometimes we have all the relevant facts, we know all the true parameter values, and we still don't know what a value's going to be. We might have great data on the average cost of a cubic yard of concrete but we don't know what it's going to be six months from now, because there's natural variability in that market. So, natural variability is another source of uncertainty.

So, imagine that we take that set of things that we don't know and divide it into two subsets. We find natural variability and knowledge uncertainty.

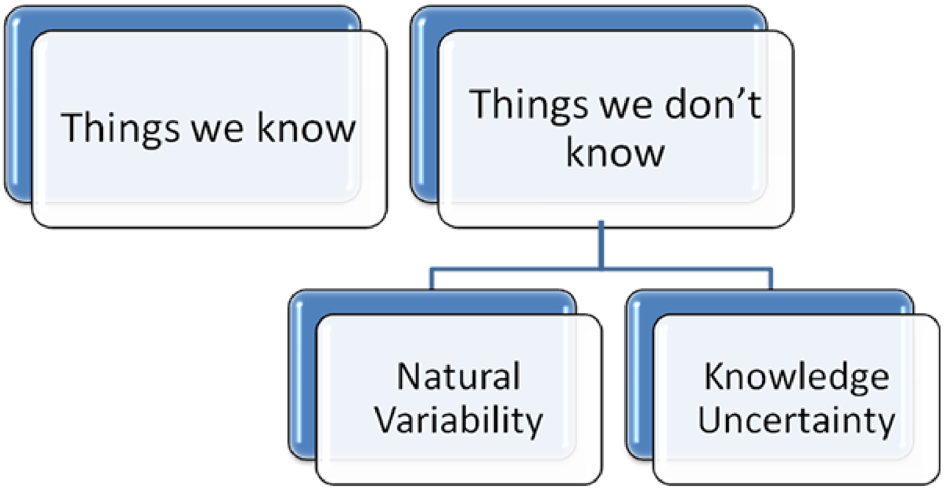


Figure 12. Subsets of Things We Don't Know

Natural variability, once we have the basic facts, the means, the standard deviations – or whatever the relevant facts are – is relatively straightforward for us to address. We do it all the time in our H and H work, for example. So, let's focus on the knowledge uncertainty here, and parse that a little further.

We could identify three broad categories of things that are subject to this knowledge uncertainty.

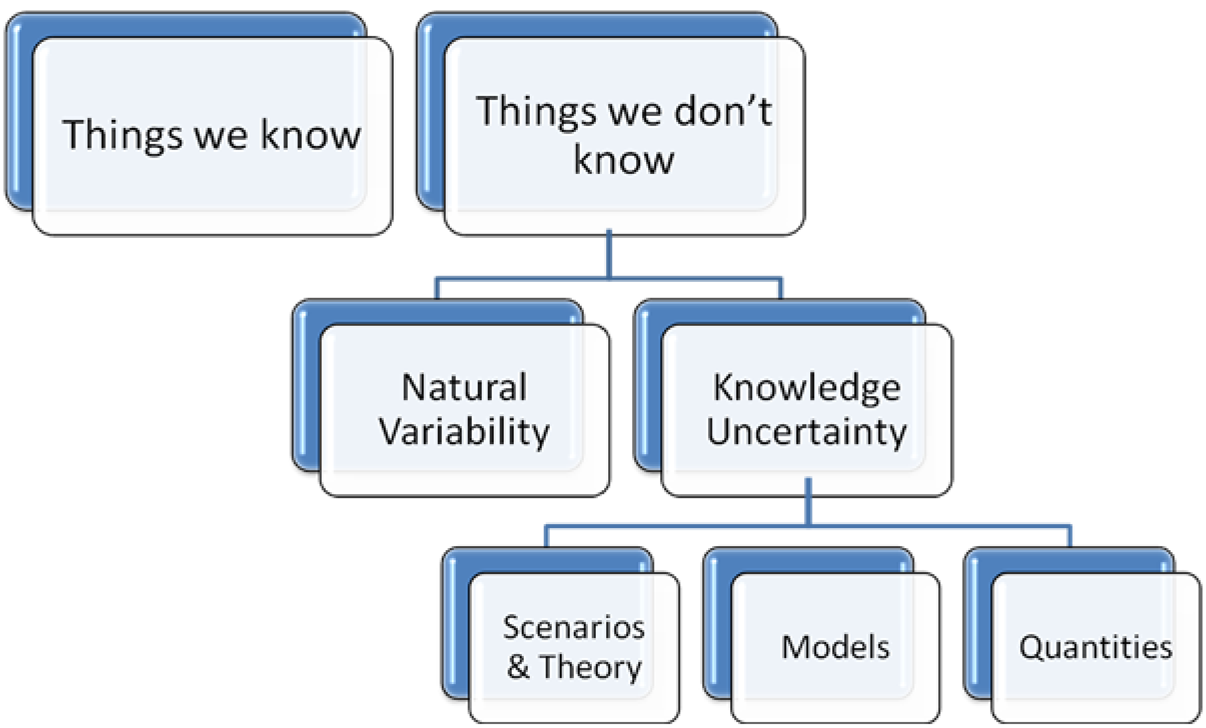


Figure 13. Types of Knowledge Uncertainty

One category would be our scenarios and theory. The basic stories we tell about how these complex ecosystems, natural systems and water systems work are sometimes subject to uncertainty; we just don't understand the workings well enough. Sometimes that's because our theory is incomplete. So, think of this first category as just general knowledge and understanding.

A second category of things which tend to be subject to knowledge uncertainty are our models themselves, the ways we structure our scenarios, and the equations that we use. The resolution of our models and all those sorts of things tend to be subject to a great deal of knowledge uncertainty. Take the habitat suitability index that we use frequently for habitat units. There's a great deal of uncertainty in there, in the life requisites that are identified, and how those life requisites contribute to the suitability of a particular habitat for a species.

The third and most familiar type of thing that is subject to knowledge uncertainty would be quantities. These are the variables that we deal with in our studies. These are the cost variables, the values, the value of a house, the cost of that cubic yard of concrete, the amount of rock in the bottom of a channel – all those sorts of quantitative variables can be subject to knowledge uncertainty.

What's eventually going to be important to us, and what we're going to do here in the remainder of this module, is to take those quantities and break them down even further into the categories that you see here:

- Output criterion
- Model domain parameter
- Index variable
- Value parameter
- Decision variables
- Defined constants
- Empirical quantities

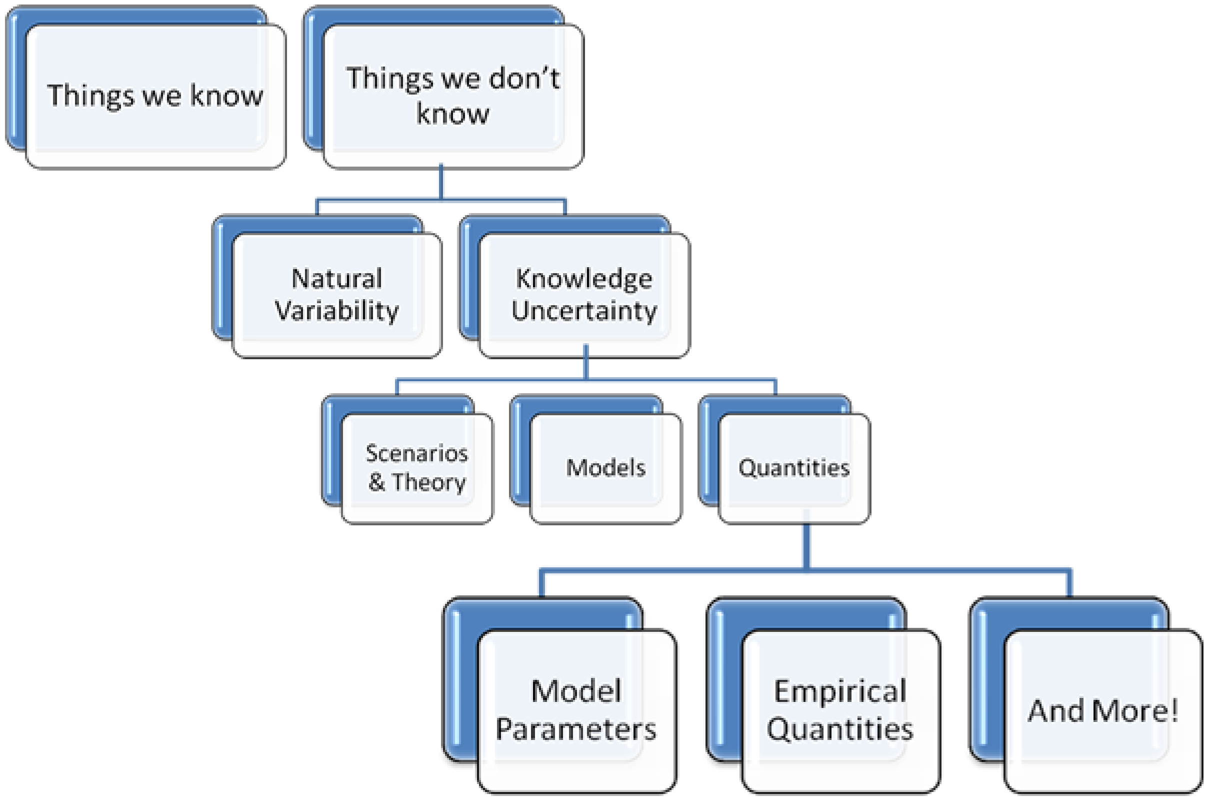


Figure 14. Subsets of Knowledge Uncertainty

The importance of understanding why a quantity might fall in one or the other of these categories has everything to do with helping us to decide the proper tool to use in order to try to reduce that uncertainty.

So, going all the way back to that set of things that we don't know - in risk analysis we want to be intentional about addressing that set of things that we don't know. There is a toolbox - the risk assessor's toolbox - that is available to us. It has many different tools in it. The proper tool depends upon the nature and the cause of the uncertainty.

For natural variability we tend to use probabilistic methods or classical statistics. For knowledge uncertainty, there is a broader array of tools. For example, if there is a defined constant that you don't know – how many square feet in an acre of land or the number of gallons in an acre foot of water – you wouldn't use such a probabilistic method. Instead, you would look it up.

We're going to be spending a little bit of time parsing these quantities so that we can understand what's in that set of things that we don't know. Ideally, we'd like to take that set of things that we don't know in the work that you're doing and break them down into the finest resolution of subsets.

Chapter 2 - Why is it Important?

2.1 A Closer Look at Knowledge Uncertainty

Natural variability occurs in most of the physical systems with which the Corps interacts.

Variability occurs:

- In the natural world
- In the performance of the infrastructure systems and assets the Corps manages

Natural variability can often be addressed in a straightforward manner using:

- Probabilistic methods
- Classical statistical techniques

A Closer Look

Knowledge uncertainty is more challenging than natural variability. It is found in:

- Scenarios, theories and knowledge
- Models
- Quantities

The figure below illustrates sources of knowledge uncertainty and how they are related.

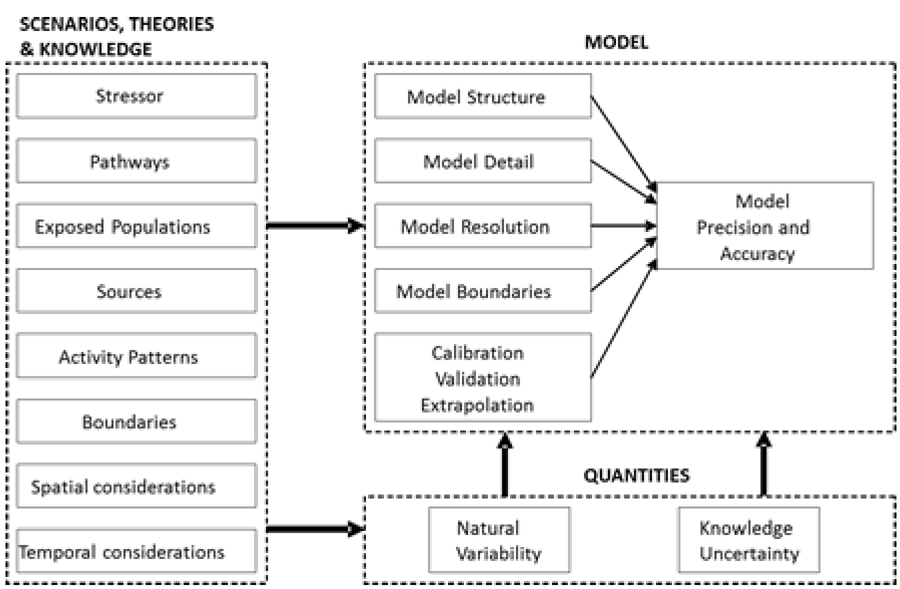


Figure 15. Knowledge Uncertainty

Scenarios, Theories and Knowledge

We may not know how the components of our scenarios relate to one another. Our characterizations of these components may be incomplete or even incorrect. The theories of our disciplines rarely provide a complete understanding of the systems with which we work. Finally, our knowledge of facts is often incomplete.

Model Uncertainty

Model uncertainty is one of the most common and persistent sources of knowledge uncertainty. All of our models are more or less flawed simplifications of reality. Model uncertainty is rarely addressed in a serious way in most analyses. The Corps has, indirectly, sought to address this problem through its model certification program (https://planning.erdc.dren.mil/toolbox/current.cfm?Title=Model%20Certification&ThisPage=ModelCert&Side=No).

The most common sources of knowledge uncertainty are the quantities the Corps analysts work with on an ongoing basis. These include inputs to models as well as the data used for day-to-day decision making. Notice the quantities source in the figure includes reference to natural variability so it does not get lost in the discussion. Let's look at the quantities a little more carefully.

Quantities

A quantity can be a fact, a parameter in a model, a parameter of a population, a statistic, a variable, data or any other form of numerical information. Morgan, Henrion, and Small (1990)[1] provided the following classification of uncertain quantities:

- Empirical quantities
- Defined constants
- Decision variables
- Value parameters
- Index variables
- Model domain parameters
- Outcome criteria

Some of these quantities have a true or factual value while others do not. When there is a true value, assessors seek it. When there is not a true value, assessors seek the best or most appropriate value.

The search for a true value is objective; the search for a best value is subjective. Analysts can often look up, measure or estimate a true value. Generally, best or most appropriate values are chosen.

Examples

True values

Quantities that have a true value include:

- Population of a floodplain
- Number of bridges crossing a waterway
- Percentage of the channel bottom that is rock
- Mean strength of materials in a structure
- Mean daily stream flow
- Average weight of a miter gate
- Current price of a yard of concrete
- Mean beam width of a class of ships
- Beam width of a specific vessel
- Number of tainter gates at a specific dam

Best values

Best values, arrived at subjectively, include:

- Discount rate
- Value of a life
- Money to be allocated for operations and maintenance
- Operations and maintenance funds a given project should receive
- Design flow for a levee
- Design vessel for a channel
- Useful life of a project
- Planning horizon
- Mitigation goal
- Length of levee to be blown up to protect a city

2.2 Types of Knowledge Uncertainty in Quantities

Often Corps analysts, risk managers, and others who work with data are faced with knowledge uncertainty about different types of quantities. The following sections describe the type of quantity and types of data it represents.

Empirical Quantities

Empirical quantities are the most common quantities encountered in Corps analyses. They have true values that may or may not be known but they are measurable in principle. Empirical quantities are things that can be measured or counted. This includes:

- Stream flows
- PH of a stream reach
- Dissolved oxygen in a stream reach
- Number of native plant species
- Distances
- Times
- Sizes
- Statistics

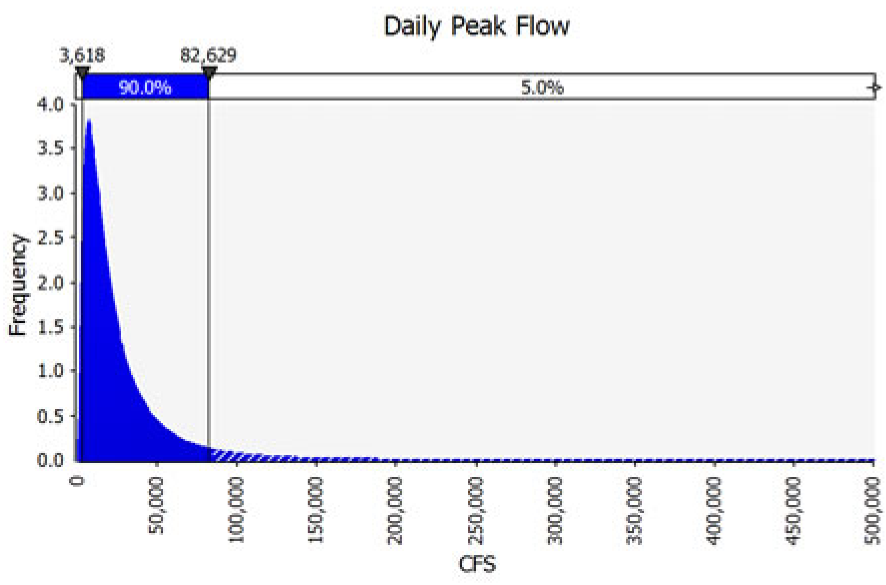


Figure 15. Example of an empirical quantity

Defined Constants

Defined constants are quantities that have a true value that is fixed by definition. When they are unknown by the analyst, the solution is to look them up.

Examples of defined constants include:

- Pi (numerical constant that occurs whenever the circumference of a circle is divided by its diameter and equals 3.14159)
- E (numerical constant that is equal to 2.71828)
- 43,560 square feet in one acre
- 325,851 gallons of water in one acre-foot of water

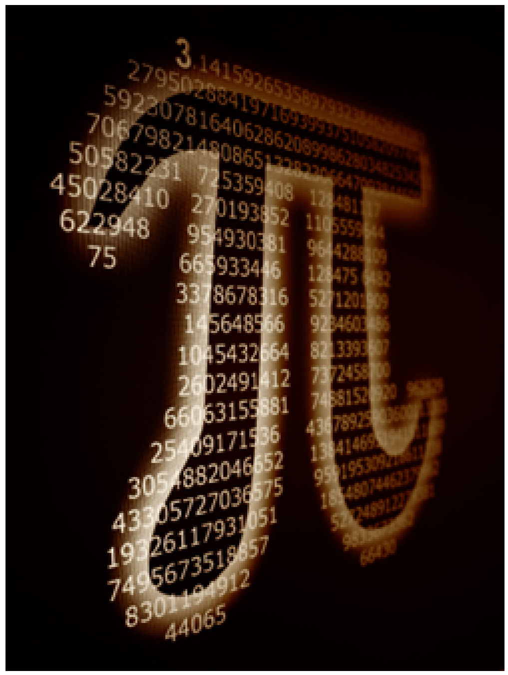


Figure 16. Illustration of pi.

Decision Variables

Decision variables do not have a true value. Decision makers exercise direct control over decision variables. Corps personnel must choose or decide many variables during the course of their work. These decisions are subjective to varying extents. For example:

- Risk managers decide the Dam Safety Action Class (DSAC) rating for a dam
- Planners decide the depth for channel deepening
- Analysts and project managers decide reasonable costs, mitigation goals, tolerable levels of risk, etc.

Figure 17. Photo of dam.

Value Parameters

Subjective assessments of social values are referred to as value parameters; they do not have true values. Such values express aspects of social values that emerge from the macro levels of uncertainty in the day-to-day work at the Corps. These values represent aspects of the decision makers' preferences and judgments and include:

- Discount rates
- Value of a life
- Weights used in multi-criteria decision analyses

Figure 18. Photo of money.

Index Variables

An index variable identifies an element of a model, a point in time, or a location within a spatial domain. Often, index variables do not have true values. Occasionally, though, a very specific point in time or place in space is needed; when this happens, there may be a true value. Examples of index values include:

- Project years 10, 20, and 30
- A representative grid cell in a GIS model

- The position of an object in a model where a sequence of events is initiated (e.g. the location of a vessel on a waterway when a lock begins to open).

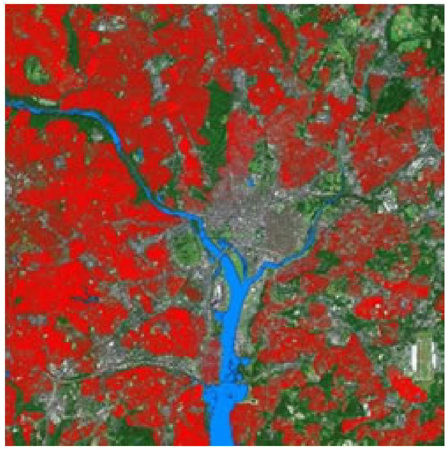


Figure 19. Aerial photo of waterway.

Model Domain Parameters

Model domain parameters are quantities that specify and define the scope of systems within a decision problem. For the Corps, these include:

- Definitions of study areas
- Impact areas
- Tributary areas to a port
- Regional sediment systems

On a smaller scale, the domains of specific models might include:

- Floodplain delineations
- Land areas for habitat unit calculations
- Sea level change boundaries

These parameters often describe the geographic, temporal and conceptual boundaries (domain) of a model and define the resolution of its inputs (minutes, hours, days, weeks) and outputs. They may or may not have true values. They usually reflect judgments regarding the model domain and the resolution needed to assess risks adequately.

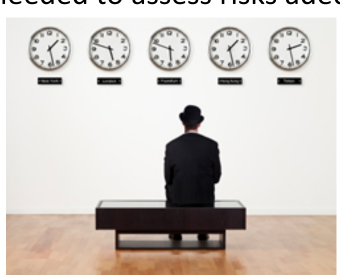


Figure 20. Photo of man and clocks.

Outcome Criteria

Outcome criteria are output variables such as:

- Benefit cost ratios
- Net benefits
- Habitat units
- Probabilities of unsatisfactory performance

They are variables that are outputs of models and calculations. Their values are determined by the models used and the quality of the model's input quantities.

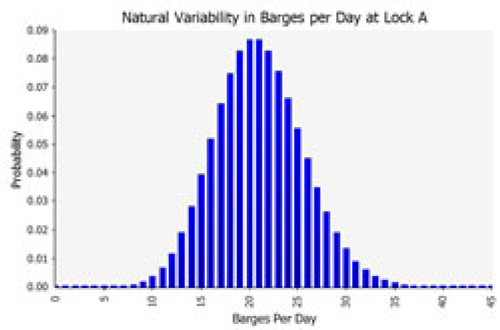


Figure 21. Example of outcome criteria

2.3 Causes of Uncertainty in Empirical Quantities

Quantities with true values that are subject to knowledge uncertainty may be the most important quantities for Corps analysts. Most are empirical quantities. To more effectively address this uncertainty, it is important to be able to identify the source or cause of the uncertainty. The following sections provide a description of causes of uncertainty in empirical quantities.

Random Error and Statistical Variation

Sample data is often the only data available for empirical quantities. Not all samples are valid probability samples; they are subject to random and statistical error. Classical statistical techniques provide a wide array of techniques and tools for quantifying this type of uncertainty, including:

- Estimators
- Standard deviations
- Confidence intervals
- Hypothesis testing
- Sampling theory
- Probabilistic methods

Table 6a. Knowledge Uncertainty Is Gone

Calculation	Value
Count	26,834 cfs
Mean	27,175 cfs
SD	32,953 cfs

Table 6b: Example of Statistical Data
USGS 01554000 Susquehanna River at Sunbury, PA

agency_cd	site_no	datetime	Flow
5s	15s	20d	
USGS	1554000	10/1/1937	4300
USGS	1554000	10/2/1937	4140
USGS	1554000	10/3/1937	3970
USGS	1554000	10/4/1937	3860
USGS	1554000	10/5/1937	3750
USGS	1554000	10/6/1937	3750
USGS	1554000	10/7/1937	3970
USGS	1554000	10/8/1937	3860
USGS	1554000	10/9/1937	3650
USGS	1554000	10/10/1937	3860
USGS	1554000	10/11/1937	3800
USGS	1554000	10/12/1937	3860
USGS	1554000	10/13/1937	3860
USGS	1554000	10/14/1937	4080
USGS	1554000	10/15/1937	4080
USGS	1554000	10/16/1937	4080
USGS	1554000	10/17/1937	4140
USGS	1554000	10/18/1937	3920
USGS	1554000	10/19/1937	4300

Systematic Error and Subjective Judgment

Systematic errors arise when the measurement instrument, the experiment or the observer are biased. Examples of such bias and ways to prevent it include:

- If a scale is not zeroed or a datum point is off, the solution is a better calibration of the instrument or data.
- If the observer tends to over- or under-estimate values, a more objective means of measurement is needed or the observer needs to be calibrated or recalibrated.

The Corps analyst is challenged to try to reduce systematic error. The best solution is to avoid, minimize or correct the bias. Please see the discussion in the example below for a few more thoughts on this problem.

Figure 22. Photo of scale.

Example: Systematic Error and Subjective Judgment

Much data are collected outside a laboratory and under less than ideal conditions. Where in the stream does the investigator insert the meter to read dissolved oxygen? How do you estimate the percent of mid-day shade on a stream? How carefully does the inspector inspect?

Subjective judgments like these are notoriously suspect under uncontrolled conditions. Although the methods discussed here tend to be analytical ones, uncertainty can sometimes be addressed by very practical means. Calibrating all measurement devices, including the people taking the measurements is often one of the best hedges for uncertainty. How good are your first floor elevations if estimated by interns using hand levels, a six-foot rule and topographic maps? It is more difficult to correct for biases that are unknown or merely suspected.

Linguistic Imprecision

Communication is still humankind's number one challenge, despite all our years on the planet. We often use the same words to mean different things and different words to mean the same things. We are linguistically imprecise.

Communicating about complex matters involving risk is especially challenging. Consider:

- If we say flooding occurs frequently or a risk of infrastructure failure is unlikely, what do these words really mean? These terms are imprecise.
- Tasked with measuring the percentage of mid-day shade on a stream for a habitat suitability model, a group of environmentalists engaged in a lengthy discussion about when mid-day occurs and how dark a surface must be to be considered shade.

The most obvious solution to this kind of ambiguity is to carefully specify all terms and relationships and to clarify all language before or as it is used. Use quantitative rather than qualitative definitions whenever possible.

Figure 23. Photo of girl holding a backwards sign.

Natural Variability

Many quantities vary over time or space or from one individual or object in a population to another. For example, an oil spill kills some fish but not others. This variability is inherent in the system that produces the population of things we measure. Frequency distributions based on samples or probability distributions for populations, if available, can be used to estimate the values of interest. Other probabilistic methods may be used as well.

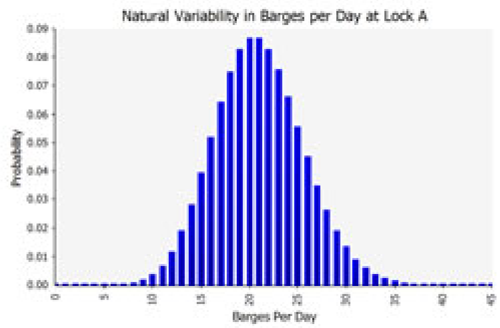


Figure 24: Natural variability outcome criteria

Randomness and Unpredictability

Inherent randomness is irreducible in principle. These events are not predictable in practice at the current time. Examples include:

- Where the next major flood will occur
- When a lock gate will fail to operate
- How the next major marine casualty will occur

These events can be treated as random processes. Uncertainty about such quantities can be addressed by a full range of methods from narrative descriptions through probabilistic methods.

Figure 25. Photo of question marks scattered randomly.

Disagreement

Experts do not always see eye-to-eye on matters of uncertainty. Neither do organizations. Examples include:

- Planning studies are cost shared. There have been many spirited debates over whether to use the water authority's hydraulics and hydrology or that of the Corps.
- Widely disparate views of the problem can exist.
- Different technical interpretations of the same data can give rise to disagreements.
- Conscious or unconscious motivational bias can exist.

Disagreements are often resolved through negotiation and other issue resolution techniques. Another option is to allow disagreements to coexist. Sensitivity analysis can be used to examine the effect of the different arguments on decision criteria.

Figure 26. Photo of thumbs up and thumbs down.

Approximation

Sometimes conditions can only be approximated. For example, ecosystem restoration in the Comprehensive Everglades Restoration Plan sought to restore the timing, quantity, and quality of water closer to conditions that existed many decades before data were systematically collected. At best, they could only approximate these conditions through the spotty data that were available. Often this type of uncertainty is manifested in model uncertainty. Methods for dealing with this source of uncertainty will depend on the specific limitations of the approximation.

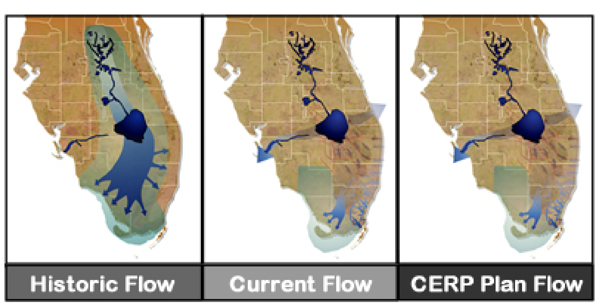


Figure 27. Graphics of historic, current and CERP plan flow.

2.4 Being Intentional About Uncertainty

It is important to intentionally take steps to effectively address uncertainty. Below are nine steps you can take to become intentional about uncertainty in decision making:

1. Recognize that uncertainty exists in your decision problem.
2. Identify the specific things that are uncertain and the sources of that uncertainty.
3. Identify those uncertainties that are important to your decision problem. These are the uncertainties that have the potential to have a significant effect on your decision criteria or on decision outcomes.
4. Acknowledge this significant uncertainty and make stakeholders aware of its existence.
5. Choose appropriate tools and techniques to reduce or otherwise address each significant source of uncertainty.
6. Complete your analysis incorporating these tools and techniques.
7. Understand the results of your analysis.
8. Identify any options for further reducing remaining uncertainty.
9. Convey your results, the significance of the uncertainty, and any options for further reducing uncertainty to decision makers.

Chapter 3 - Let's Pull It All Together

3.0 Let's Pull It All Together

This course has provided you with a brief overview to the importance of considering uncertainty in the Corps decision making processes. Risk analysis is a framework that has evolved specifically to make decisions under conditions of uncertainty. It is, therefore, impossible to understand risks or risk analysis without understanding uncertainty.

In this final section, we pull the various concepts together and illustrate them in a series of four examples that discuss aspects of common decision making processes at the Corps.

3.1 Example: Probability, the Language of Uncertainty

Probability is the language of uncertainty. It's the language we use to express our knowledge uncertainty of true facts, and it's the language we use to describe the natural variability in the world around us. Let's wrap-up with a fairly simple example. Imagine that we had 100,000 houses and we're interested in estimating the cost of purchasing these 100,000 houses. Consider it the world's biggest nonstructural flood risk management project. We need to know the mean value of houses. This is a number that is a true fact - it exists, but we just don't know what it is.

Table 7. Examining Probability

Description	Value
100,000 houses	100,000
Estimated Mean Value of House	$ 152,474
Estimated Value of 100,000 Houses	$ 15,247,409,274
True Mean	$ 149,894
True Standard Deviation	$ 42,975
True Value of 100,000 Houses	$ 14,984,900,000
Estimated Value of 1 House	$ 187,165

The mean value of the houses is a significant uncertainty, and we're going to represent that as a probability distribution. Figure 28 below shows how we can conceptualize this issue; the software used is not relevant for now.

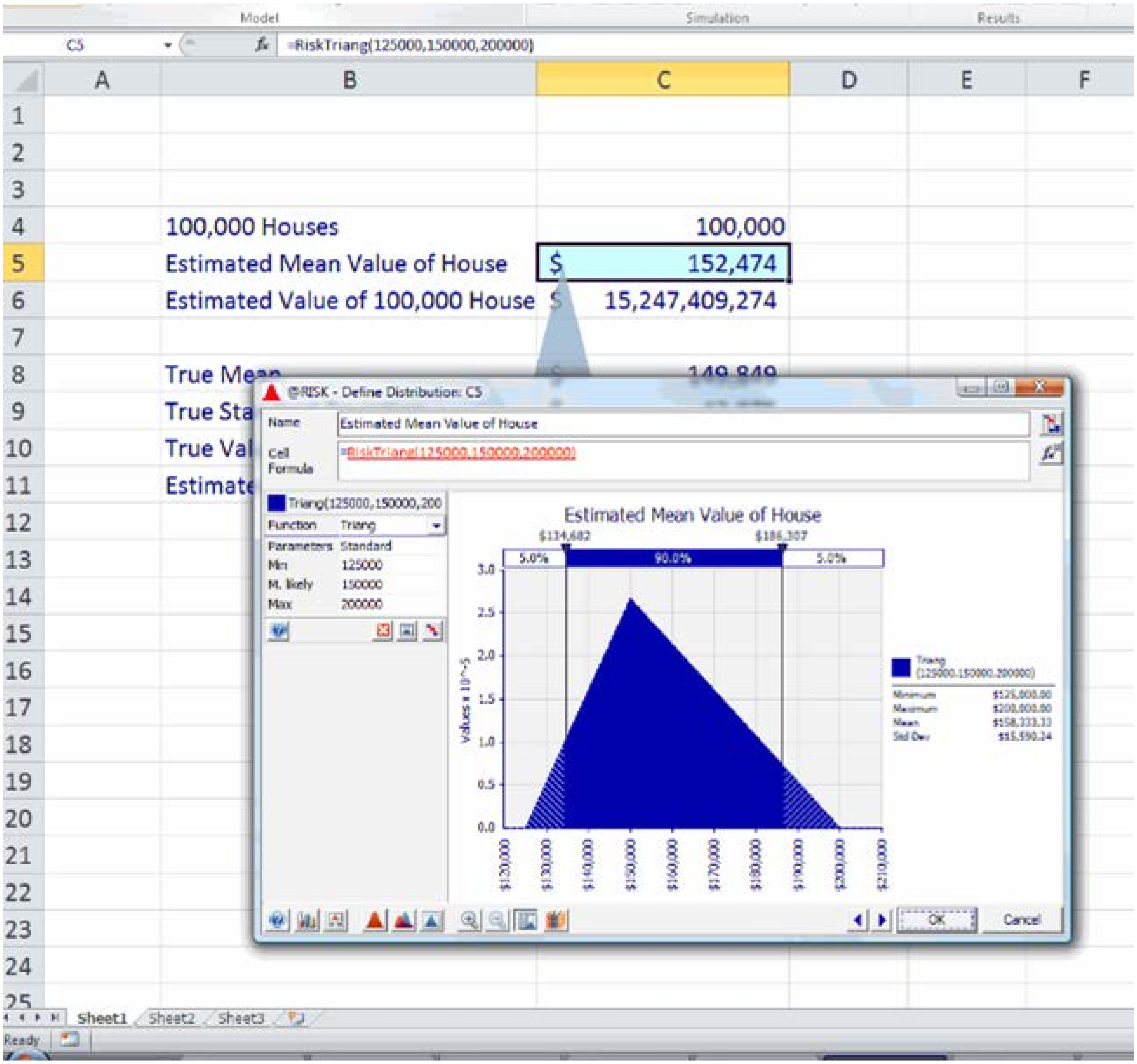


Figure 28. Estimating the Mean Value of a House

The distribution above expresses the knowledge uncertainty about the mean value of a house in this area. Consider the number line on the graph in the figure. It suggests the mean value of a house is not less than $125,000 nor is it more than $200,000. This is a defined segment of the number line and suggests the true value is somewhere within the segment. Now imagine some very smart software that asks me the question “do you know anything else about the mean value?” Suppose some realtors provided information that suggested that the mean was probably about $150,000. The distribution you’re looking at grabs that number 150,000, lifts it up, and gives us three points on a triangle: a minimum, a most likely, and a maximum value.

This probability distribution is a representation of uncertainty. The fat parts of the graph show that numbers near 150,000 are more likely to be the true mean than other values.

When you look at the spreadsheet in Table 9, there's a value there now; $183,000 is the estimate. This is a number randomly chosen from that segment of the number line following the rule assigned with that triangular distribution.

Table 9. Estimating the Mean Value of a House

Description	Value
100,000 houses	100,000
Estimated Mean Value of House	$ 183,512
Estimated Value of 100,000 Houses	$ 18,351,234,330
True Mean	$ 149,894
True Standard Deviation	$ 42,975
True Value of 100,000 Houses	$ 14,984,900,000
Estimated Value of 1 House	$ 108,048

If we recalculate, the estimate of the 100,000 houses changes to 15.8 billion:

Table 10. Estimating the Mean Value of a House

Description	Value
100,000 houses	100,000
Estimated Mean Value of House	$ 158,333
Estimated Value of 100,000 Houses	$ 15,833,333,333
True Mean	$ 149,894
True Standard Deviation	$ 42,975
True Value of 100,000 Houses	$ 14,984,900,000
Estimated Value of 1 House	$ 149,849

If we recalculate again, it then changes to 16.9 billion:

Table 11. Estimating the Mean Value of a House

Description	Value
100,000 houses	100,000
Estimated Mean Value of House	$ 169,431
Estimated Value of 100,000 Houses	$ 16,943,114,963
True Mean	$ 149,894
True Standard Deviation	$ 42,975
True Value of 100,000 Houses	$ 14,984,900,000
Estimated Value of 1 House	$ 133,030

Recalculate again, it then changes to 15.8 again:

Table 12. Estimating the Mean Value of a House

Description	Value
100,000 houses	100,000
Estimated Mean Value of House	$ 158,333
Estimated Value of 100,000 Houses	$ 15,833,333,333
True Mean	$ 149,894
True Standard Deviation	$ 42,975
True Value of 100,000 Houses	$ 14,984,900,000
Estimated Value of 1 House	$ 149,849

We've used probability to express our uncertainty about a true value. Now imagine that we went out and actually calculated the true mean and found it to be $149,849, with the standard deviation shown below in Table 13.

Table 13. Estimating the Mean Value of a House

Description	Value
100,000 houses	100,000
Estimated Mean Value of House	$ 169,439
Estimated Value of 100,000 Houses	$ 16,943,914,477
True Mean	$ 149,849
True Standard Deviation	$ 42,975
True Value of 100,000 Houses	$ 14,984,900,000
Estimated Value of 1 House	$ 163,392

It's quite a simple matter to calculate the cost of 100,000 houses. That happens to be the nearly $15 billion.

Suppose we're interested not in the total value of all these houses or in the mean value. These two would be constants. Suppose instead we are interested in the cost of a single randomly chosen house. We would use probability to describe this number. But this is no longer a true value; it changes depending on what house we pull. So imagine that big lottery cage with ping pong balls with 100,000 ping pong balls, each one representing a house. Each time we draw a ball it could be a house with a different value.

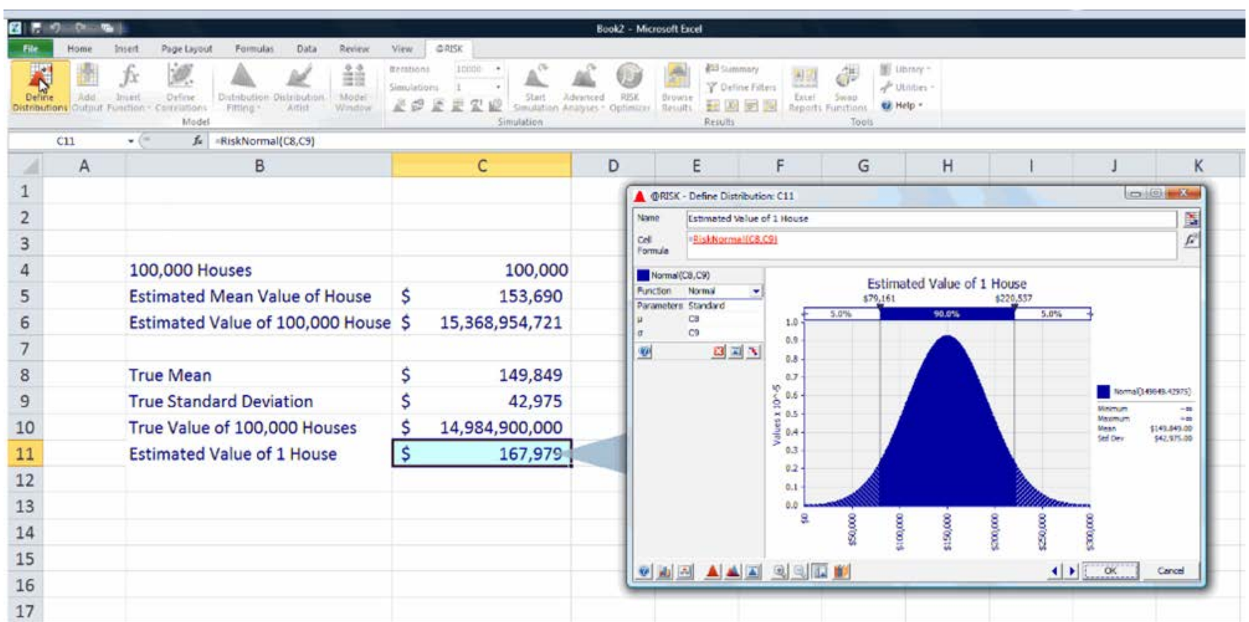


Figure 29. Estimating the Mean Value of a House

We've used probability to represent our uncertainty. In this case, we have a normally distributed value of houses with a mean of $149,000 and a standard deviation of almost $43,000. The takeaway point from this is very simple — if you're going to be intentional about how you address the uncertainty in your decision making problems, someone is going to have to learn some basic probability concepts. If we're dealing with risk, which is defined as a probability times a consequence, there's an essential reason to understand probabilities. This only becomes more essential when we understand that these two elements of a risk are themselves subject to uncertainty.

3.2 Example: CDEP (Corps Dredge Estimating Program) - The Dredging Model

The following sections demonstrate the ubiquitous nature of uncertainty in Corps work using a cost estimating model developed to estimate unit costs of dredging. You're looking at CDEP, the Corps Dredge Estimating Program, below. This is a spreadsheet model that was developed by and is maintained by the Walla Walla District of the US Corps of Engineers. It's used for estimating the unit costs of dredging projects.

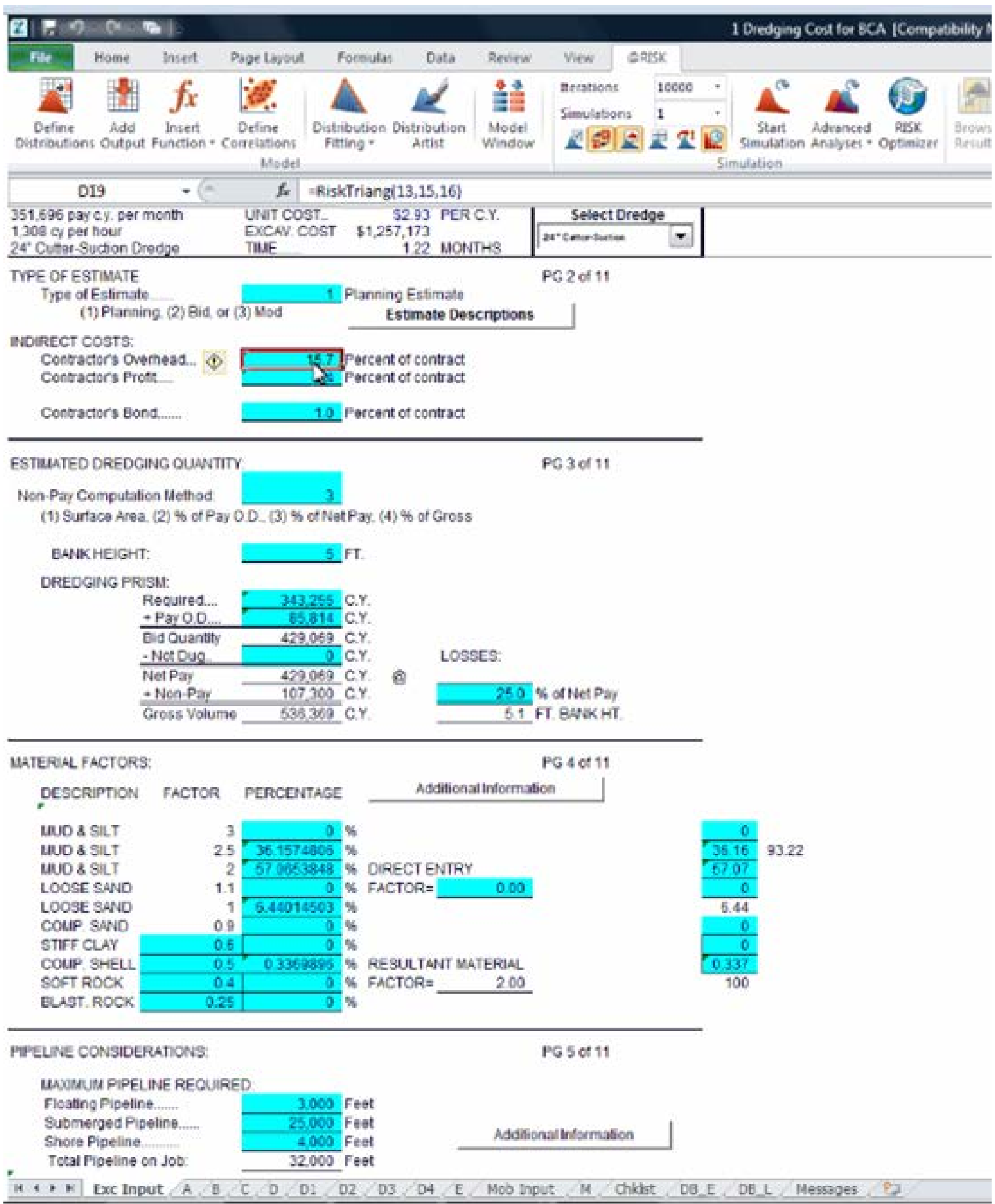


Figure 30. CDEP Spreadsheet Model

We notice in the unit cost in this particular calculation is $2.93 per cubic yard. Without going into the details of this model, we might understand that there are a lot of numbers here that are true values that we simply don't know or are naturally variable.

For example, we can examine what percent of the contract the 'Contractor's Overhead' will constitute. The estimators have put in 15.7 percent. But is that in the set of things that we know or the set of things that we don't know?

It turns out that this is in the set of things that we don't know. When pressed, our estimators said this value isn't going to be less than 13 percent because in recent history it's never been less than 13 percent. Likewise, it's not going to be more than 16 percent because in recent history it has not risen above that number. Most likely though, between 13 and 16, their best estimate was 15 percent. So once again you can imagine that we grab the number 15, raise it up above the number line, get out our ruler and connect the dots between 13 and 15 percent raised up above the number line and 16 percent back down on the number line, giving us a very neat triangular distribution (shown in the figure below in blue) that represents our uncertainty.

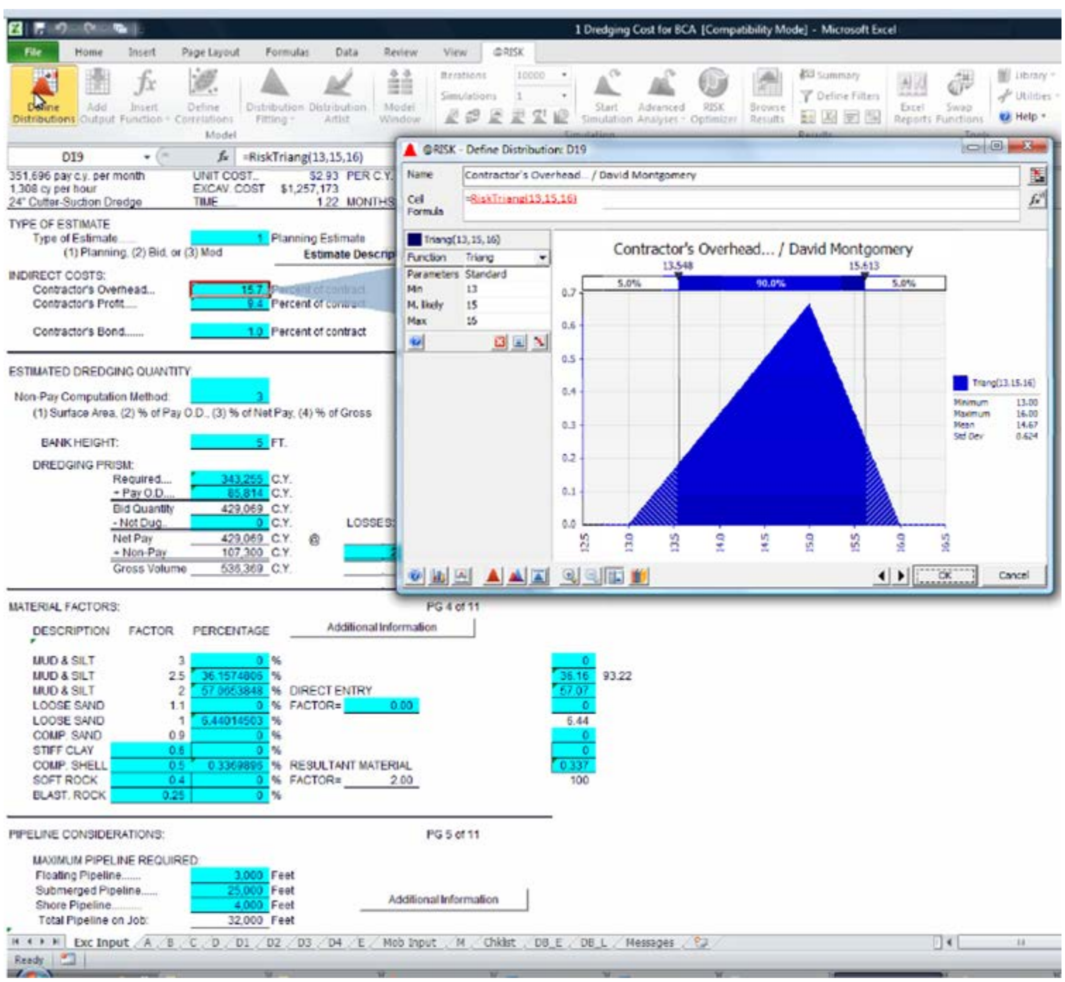


Figure 31. CDEP Spreadsheet Model using @Risk

We are not exactly sure what the "Contractor's Overhead"could be; it could be any of these numbers.

The 'Bid Quantity' for example, is somewhere between 302,000 cubic yards and 495,000 cubic yards. That's quite a range with quite a bit of uncertainty. Most likely it's 391,590 cubic yards.

Just to take one more example: what percent of the month will the pipeline be available?

Our analysts say it's going to be somewhere between 40 and 70 percent of the time, most likely about 64 percent. So experts are expressing their beliefs about those things that are uncertain. They're expressing their understanding of the natural variability in the system.

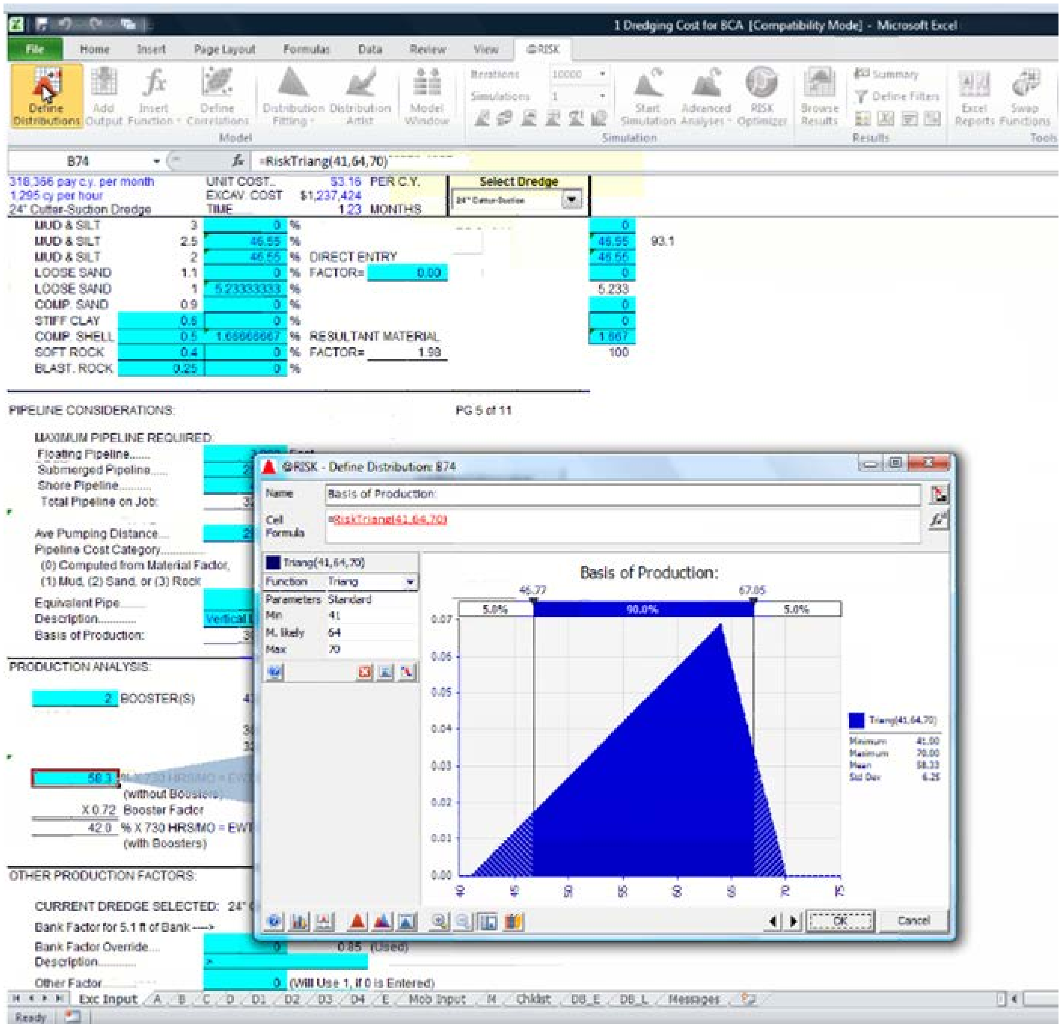


Figure 32. CDEP Spreadsheet Model Inputs

But what's important to understand is that as this model goes through its calculations, this value (the unit cost) is uncertain. So we'll quickly do a simulation. Instead of clicking the dice a great number of times, we're going to let this software calculate that number 10,000 times for us.

This will give you an example of a quantitative characterization of the uncertainty in the value.

It didn't take too long to complete but we now have a result that says 'I don't know for sure what the unit cost of dredging is going to be.' We're now being honest about the things that we don't know. And one of the reasons I'm not sure is because I don't know what the 'Contractor's Overhead' is going to be. And I don't know what the Bid Quantity is going to be. And I don't know what percent of the 730 hours in a month our pumps are going to be working. How often will they be clogged with rocks, or have to go down, be taken apart for other reasons? And when you take all of this into account, what it means is the unit cost of dredging is somewhere between $2.31 (and that would be in the circumstance where everything that could break in the good direction breaks in that good direction; all the uncertainty resolves itself in favor of a low cost estimate) and at the other extreme it could be $4.06. And that's just the opposite, where all the uncertainty, all of those factors would resolve themselves in the least favorable way in terms of the cost.

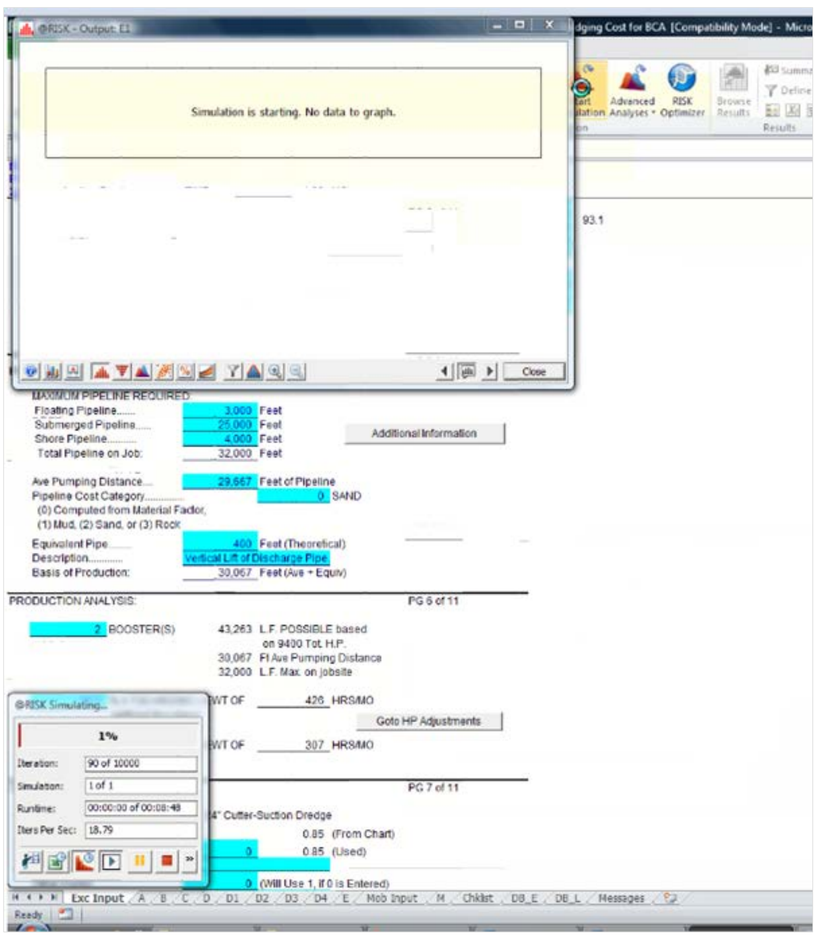


Figure 33. CDEP Spreadsheet Model Inputs

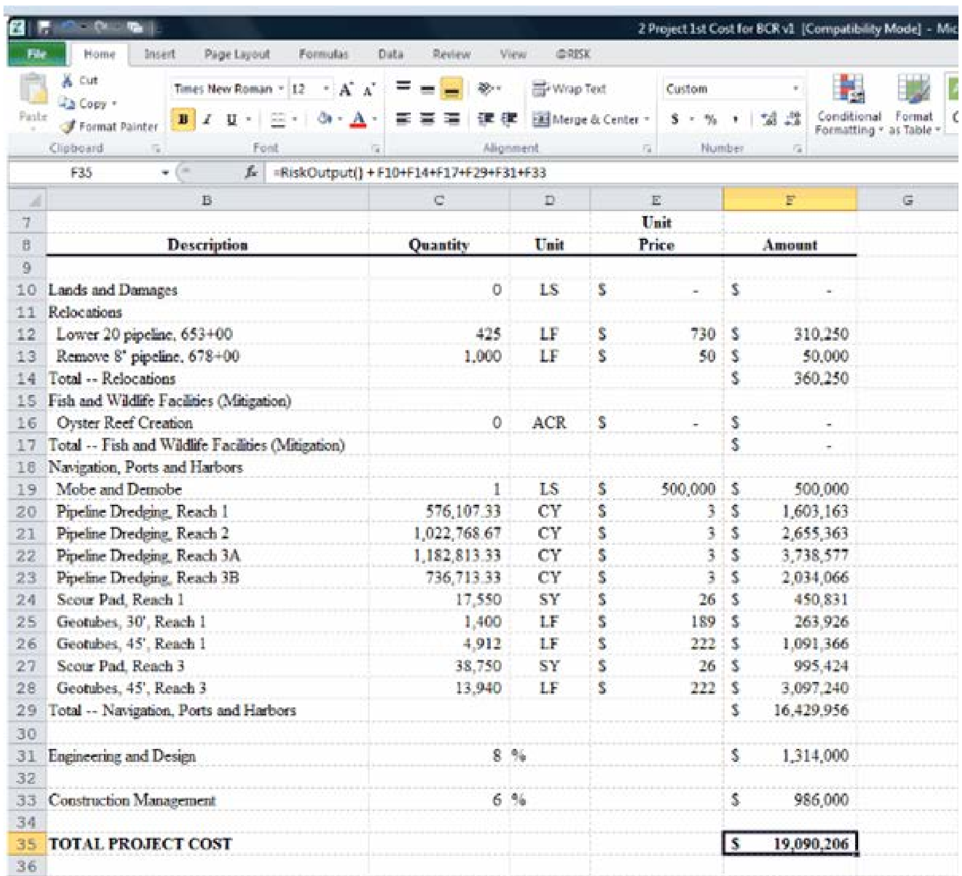
Our best guess: take all 10,000 cost estimates, add them up and divide by 10,000, which means an average cost of $3.07. This distribution shown, in the earlier figures, are the output of the CDEP model that would become an input to our cost estimate.

This is one way we can all understand costs even if we don’t understand the CDEP model. We all understand the notion of costs and this shows you one way where all of that probability that we used to express our uncertainty, whether it’s knowledge uncertainty or natural variability, can be propagated to give us a range of possible outcomes for a model.

3.3 Example: Cost Estimating – Project Cost Model

Using outputs from the CDEP model as inputs to the project cost model, costs are estimated under uncertainty. The following sections demonstrate how assessors can be intentional about uncertainty.

It's the risk assessor's, or Corps analyst's, job to characterize the uncertainty in the inputs to decision making. We previously examined a CDEP model where we derived a distribution of cost of dredging.



	Description	Quantity	Unit	Unit Price	Amount
10	Lands and Damages	0	LS	$ -	$ -
11	Relocations				
12	Lower 20 pipeline, 653+00	425	LF	$ 730	$ 310,250
13	Remove 8" pipeline, 678+00	1,000	LF	$ 50	$ 50,000
14	Total -- Relocations				$ 360,250
15	Fish and Wildlife Facilities (Mitigation)				
16	Oyster Reef Creation	0	ACR	$ -	$ -
17	Total -- Fish and Wildlife Facilities (Mitigation)				$ -
18	Navigation, Ports and Harbors				
19	Mobe and Demobe	1	LS	$ 500,000	$ 500,000
20	Pipeline Dredging, Reach 1	576,107.33	CY	$ 3	$ 1,603,163
21	Pipeline Dredging, Reach 2	1,022,768.67	CY	$ 3	$ 2,655,363
22	Pipeline Dredging, Reach 3A	1,182,813.33	CY	$ 3	$ 3,738,577
23	Pipeline Dredging, Reach 3B	736,713.33	CY	$ 3	$ 2,034,066
24	Scour Pad, Reach 1	17,550	SY	$ 26	$ 450,831
25	Geotubes, 30', Reach 1	1,400	LF	$ 189	$ 263,926
26	Geotubes, 45', Reach 1	4,912	LF	$ 222	$ 1,091,366
27	Scour Pad, Reach 3	38,750	SY	$ 26	$ 995,424
28	Geotubes, 45', Reach 3	13,940	LF	$ 222	$ 3,097,240
29	Total -- Navigation, Ports and Harbors				$ 16,429,956
31	Engineering and Design	8 %			$ 1,314,000
33	Construction Management	6 %			$ 986,000
35	**TOTAL PROJECT COST**				$ 19,090,206

Figure 34. CDEP Spreadsheet Model (Costs of Dredging)

Those distributions have been captured (Figure 34) and entered into the model in Figure 35. What you see before you is a cumulative distribution. This is just another form in which the data can be presented.

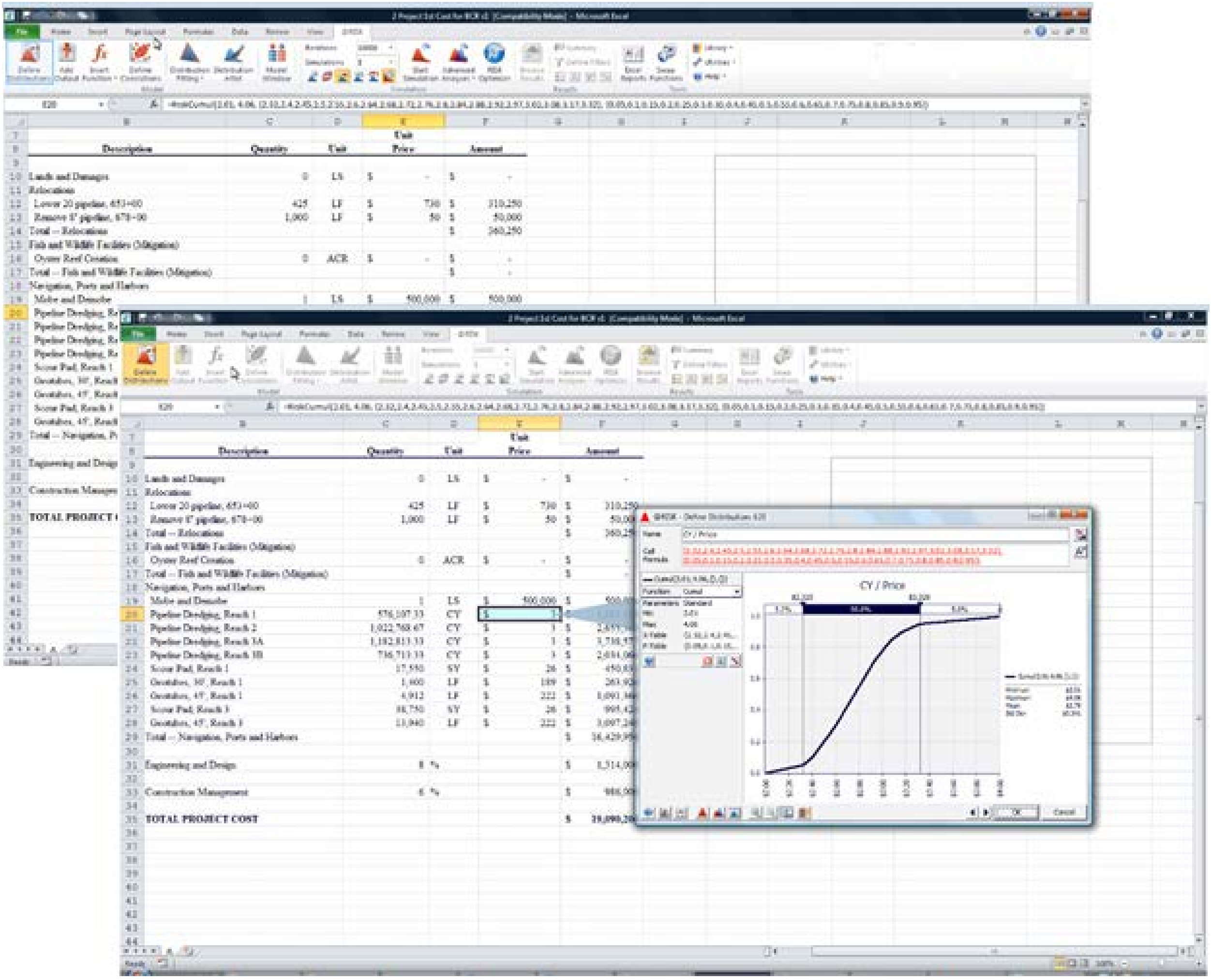


Figure 35. Cumulative Distribution

Below in Figure 36 is a more familiar probability density function, without all the blocks and the details you saw before.

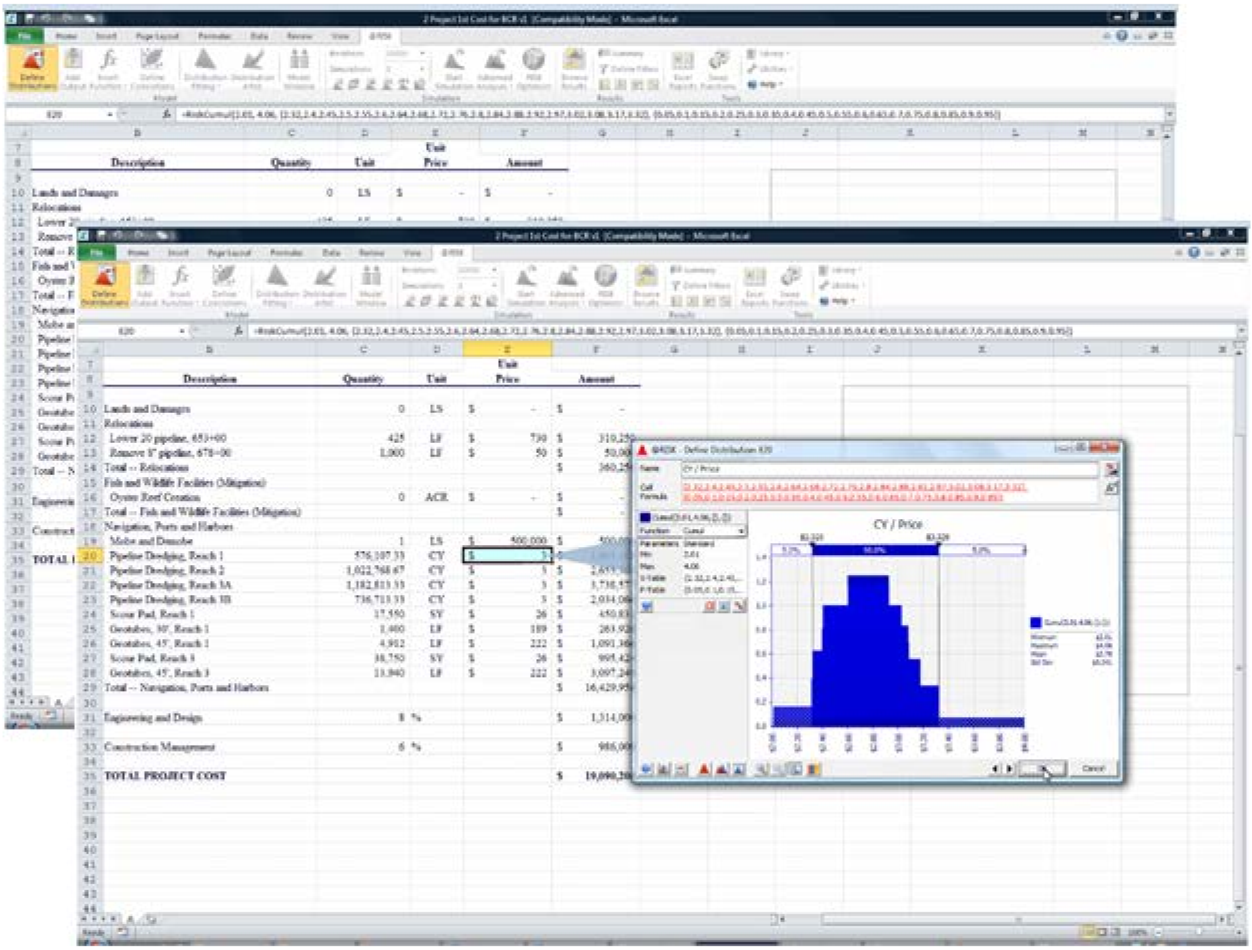


Figure 36. Probability Distribution

But, every distribution input here is described appropriately as a probability distribution, and what this does is enables us to look at the entire range of potential project costs (Figure 37).

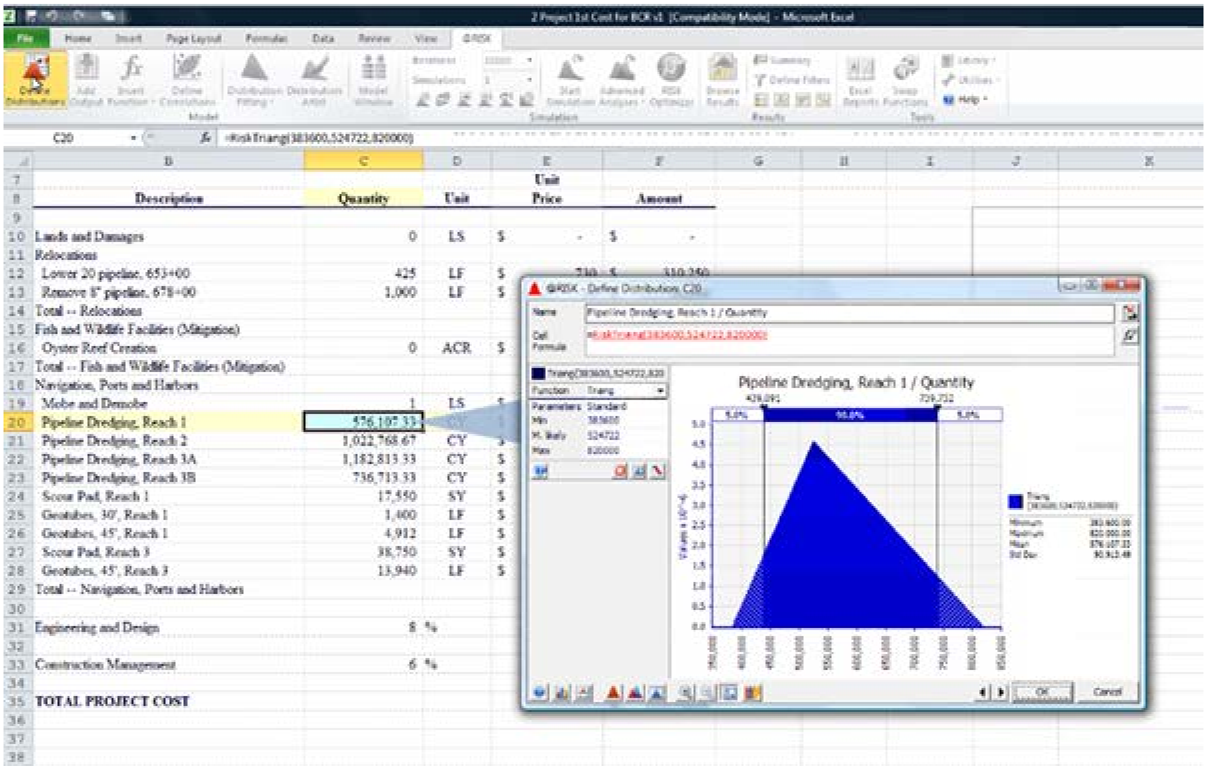


Figure 37. Using Uncertainty in a Simulation

If we're interested ultimately in project costs, we'll select that number and run another simulation similar to the one you saw before, but this is taking into account all of the uncertainty in the inputs. Here, you see the outcome where we estimate that this projects costs somewhere between $13 and $26 million dollars. Mostly likely the cost is $19 million dollars.

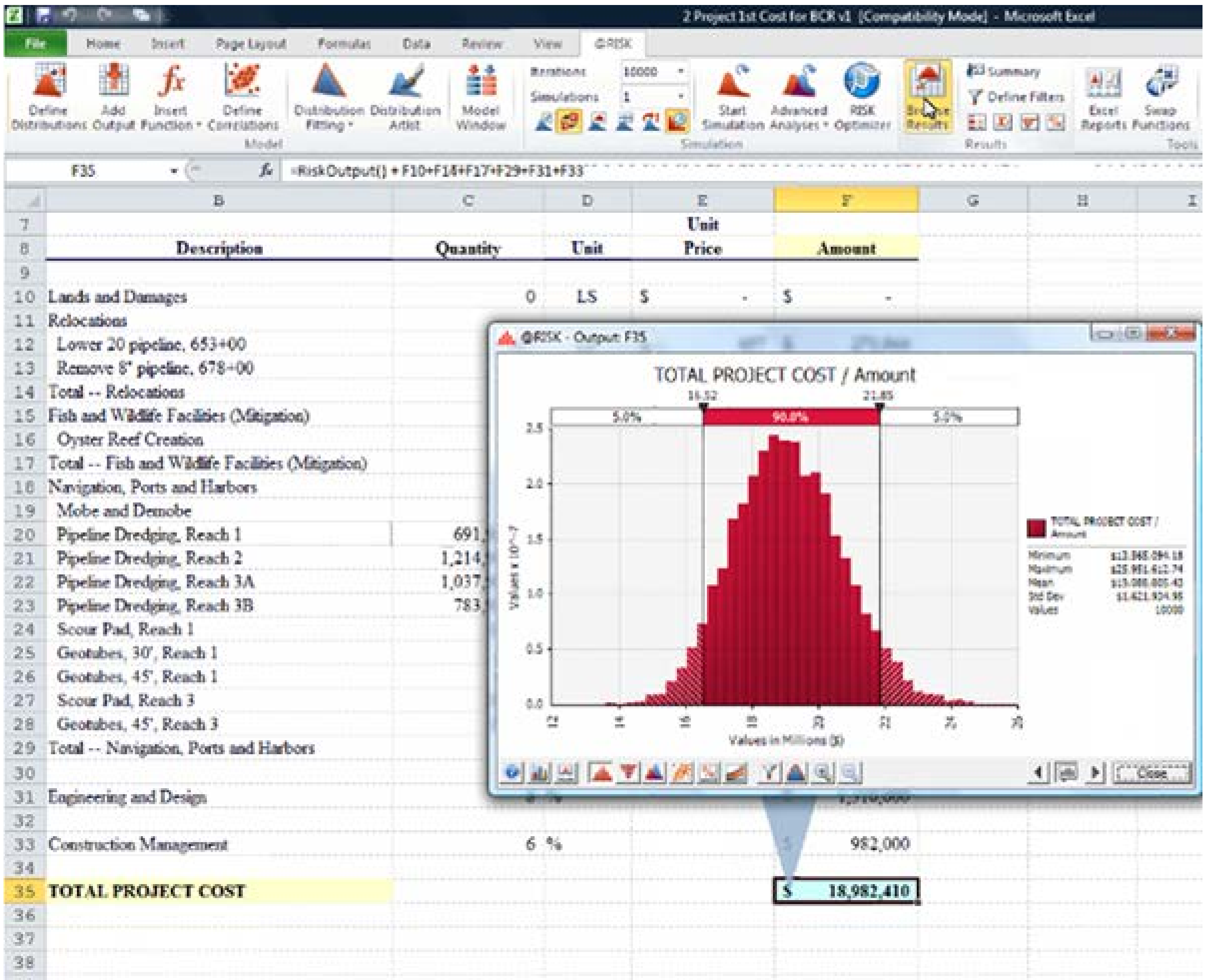


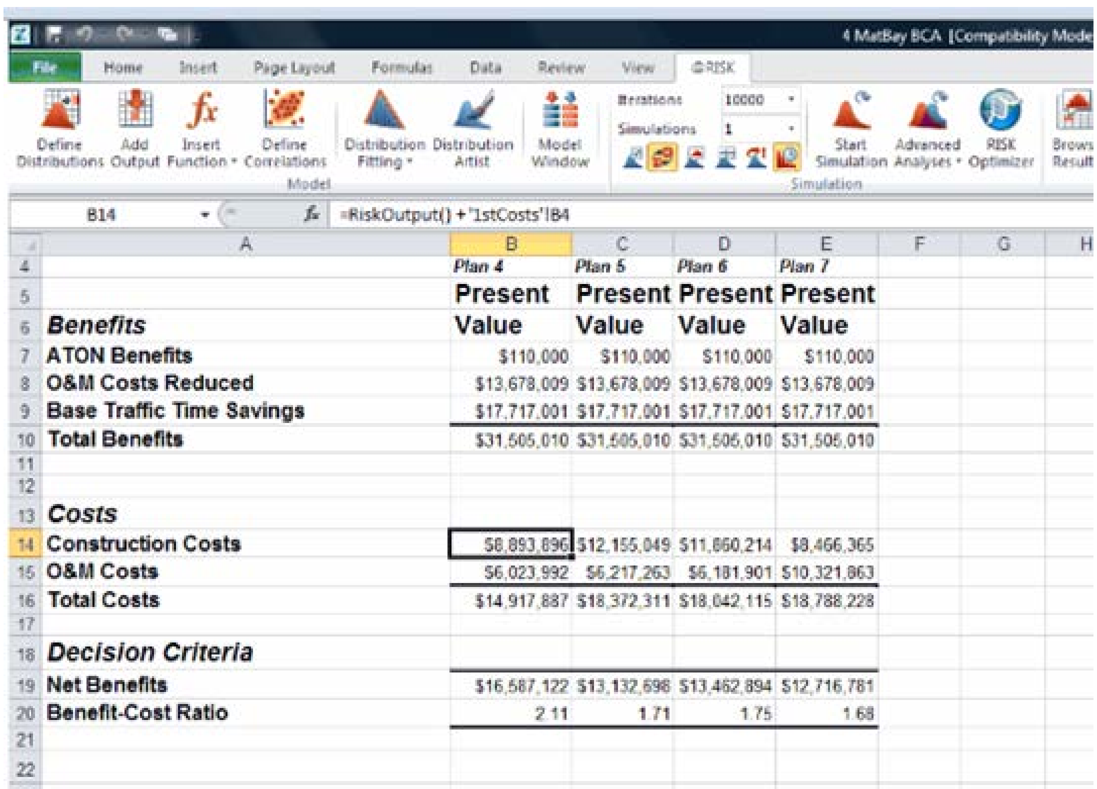
Figure 38. Using Uncertainty in a Simulation

It is the risk assessor's job to characterize the decision-making criteria in this way and to be intentional about the uncertainty.

3.4 EXAMPLE: DECISION MAKING – BENEFIT-COST RATIO MODEL – BENEFITS

Using outputs from previous models as inputs to the benefit-cost ratio model, the uncertainty about the benefit-cost ratios can be characterized and compared across plans. Addressing such uncertainty in decision criteria is the risk manager's responsibility.

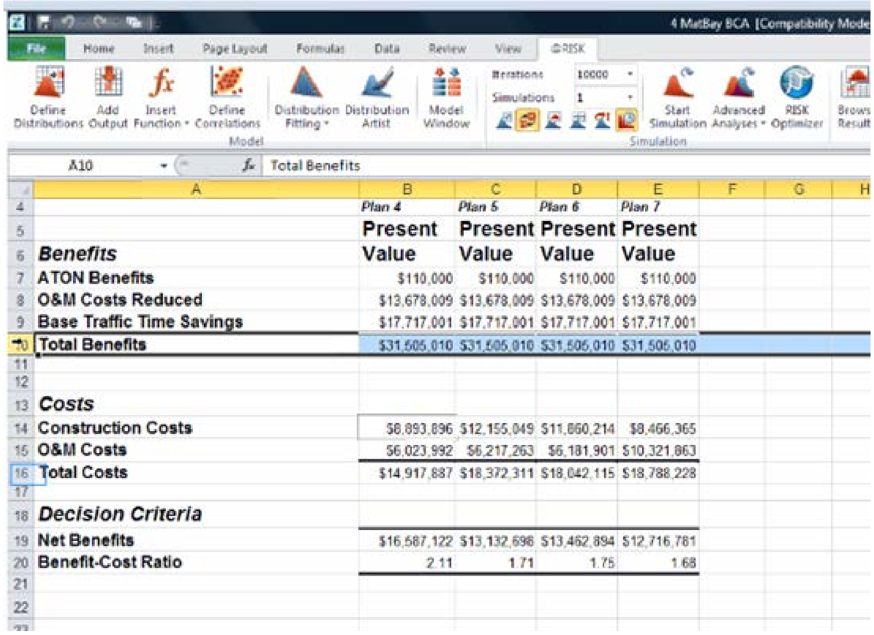
To bring this full circle, let's consider the risk manager's responsibility with respect to this uncertainty. First, we saw how the uncertainty in the dredging cost was handled in the CDEP model. Then we looked at how all the uncertainties related to the cost were handled in the cost-estimating model. A similar effort has been undertaken in estimating the benefits, although we won't show that here.



	Plan 4	Plan 5	Plan 6	Plan 7
Benefits	**Present Value**	**Present Value**	**Present Value**	**Present Value**
ATON Benefits	$110,000	$110,000	$110,000	$110,000
O&M Costs Reduced	$13,678,009	$13,678,009	$13,678,009	$13,678,009
Base Traffic Time Savings	$17,717,001	$17,717,001	$17,717,001	$17,717,001
Total Benefits	$31,505,010	$31,505,010	$31,505,010	$31,505,010
Costs				
Construction Costs	$8,893,896	$12,155,049	$11,860,214	$8,466,365
O&M Costs	$6,023,992	$6,217,263	$6,181,901	$10,321,863
Total Costs	$14,917,887	$18,372,311	$18,042,115	$18,788,228
Decision Criteria				
Net Benefits	$16,587,122	$13,132,698	$13,462,894	$12,716,781
Benefit-Cost Ratio	2.11	1.71	1.75	1.68

Figure 39. Distribution of Costs and Benefits

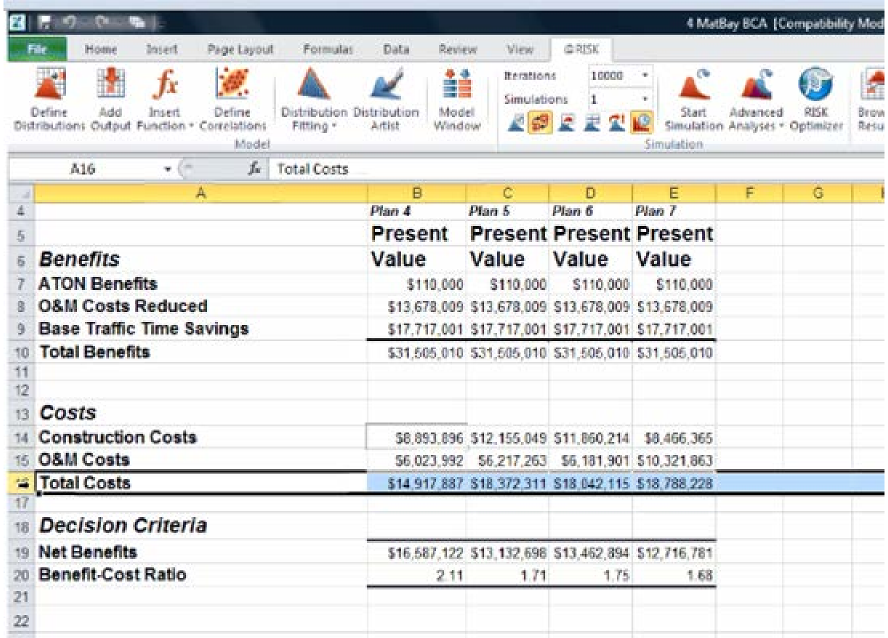
As an example, the image below shows the estimated benefits for four specific plans.



	Plan 4	Plan 5	Plan 6	Plan 7
	Present Value	Present Value	Present Value	Present Value
Benefits				
ATON Benefits	$110,000	$110,000	$110,000	$110,000
O&M Costs Reduced	$13,678,009	$13,678,009	$13,678,009	$13,678,009
Base Traffic Time Savings	$17,717,001	$17,717,001	$17,717,001	$17,717,001
Total Benefits	$31,505,010	$31,505,010	$31,505,010	$31,505,010
Costs				
Construction Costs	$8,893,896	$12,155,049	$11,860,214	$8,466,365
O&M Costs	$6,023,992	$6,217,263	$6,181,901	$10,321,863
Total Costs	$14,917,887	$18,372,311	$18,042,115	$18,788,228
Decision Criteria				
Net Benefits	$16,587,122	$13,132,698	$13,462,894	$12,716,781
Benefit-Cost Ratio	2.11	1.71	1.75	1.68

Figure 40. Estimating Benefits

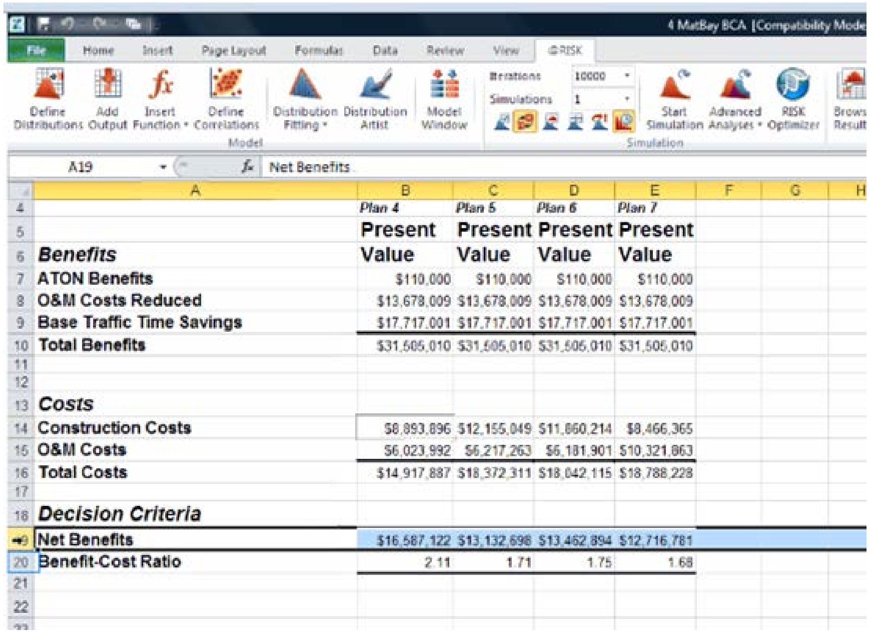
Figure 41 illustrates the varied estimated costs for those four plans.



	Plan 4	Plan 5	Plan 6	Plan 7
	Present Value	Present Value	Present Value	Present Value
Benefits				
ATON Benefits	$110,000	$110,000	$110,000	$110,000
O&M Costs Reduced	$13,678,009	$13,678,009	$13,678,009	$13,678,009
Base Traffic Time Savings	$17,717,001	$17,717,001	$17,717,001	$17,717,001
Total Benefits	$31,505,010	$31,505,010	$31,505,010	$31,505,010
Costs				
Construction Costs	$8,893,896	$12,155,049	$11,860,214	$8,466,365
O&M Costs	$6,023,992	$6,217,263	$6,181,901	$10,321,863
Total Costs	$14,917,887	$18,372,311	$18,042,115	$18,788,228
Decision Criteria				
Net Benefits	$16,587,122	$13,132,698	$13,462,894	$12,716,781
Benefit-Cost Ratio	2.11	1.71	1.75	1.68

Figure 41. Estimating Costs

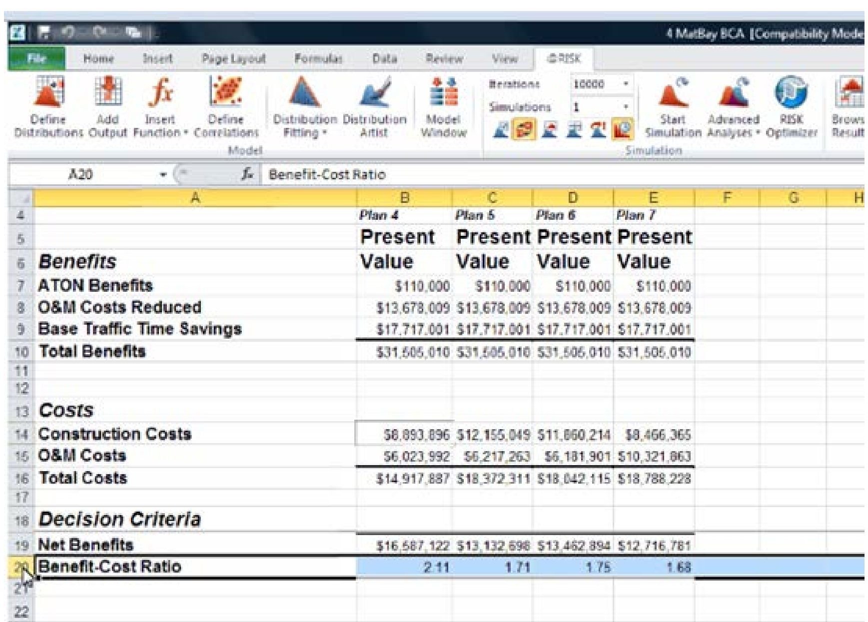
Figure 42 illustrates the varied net benefits expressed in present value.



	Plan 4	Plan 5	Plan 6	Plan 7
	Present Value	Present Value	Present Value	Present Value
Benefits				
ATON Benefits	$110,000	$110,000	$110,000	$110,000
O&M Costs Reduced	$13,678,009	$13,678,009	$13,678,009	$13,678,009
Base Traffic Time Savings	$17,717,001	$17,717,001	$17,717,001	$17,717,001
Total Benefits	$31,505,010	$31,505,010	$31,505,010	$31,505,010
Costs				
Construction Costs	$8,893,896	$12,155,049	$11,860,214	$8,466,365
O&M Costs	$6,023,992	$6,217,263	$6,181,901	$10,321,863
Total Costs	$14,917,887	$18,372,311	$18,042,115	$18,788,228
Decision Criteria				
Net Benefits	$16,587,122	$13,132,698	$13,462,894	$12,716,781
Benefit-Cost Ratio	2.11	1.71	1.75	1.68

Figure 42. Estimating Net Benefits

Figure 43 illustrates the various benefit-cost ratios. Now because we don't know what the contractor's overhead is, because we don't know the bid quantities, because there are so many quantities that are uncertain along the way, we can't possibly know the benefit-cost ratio.



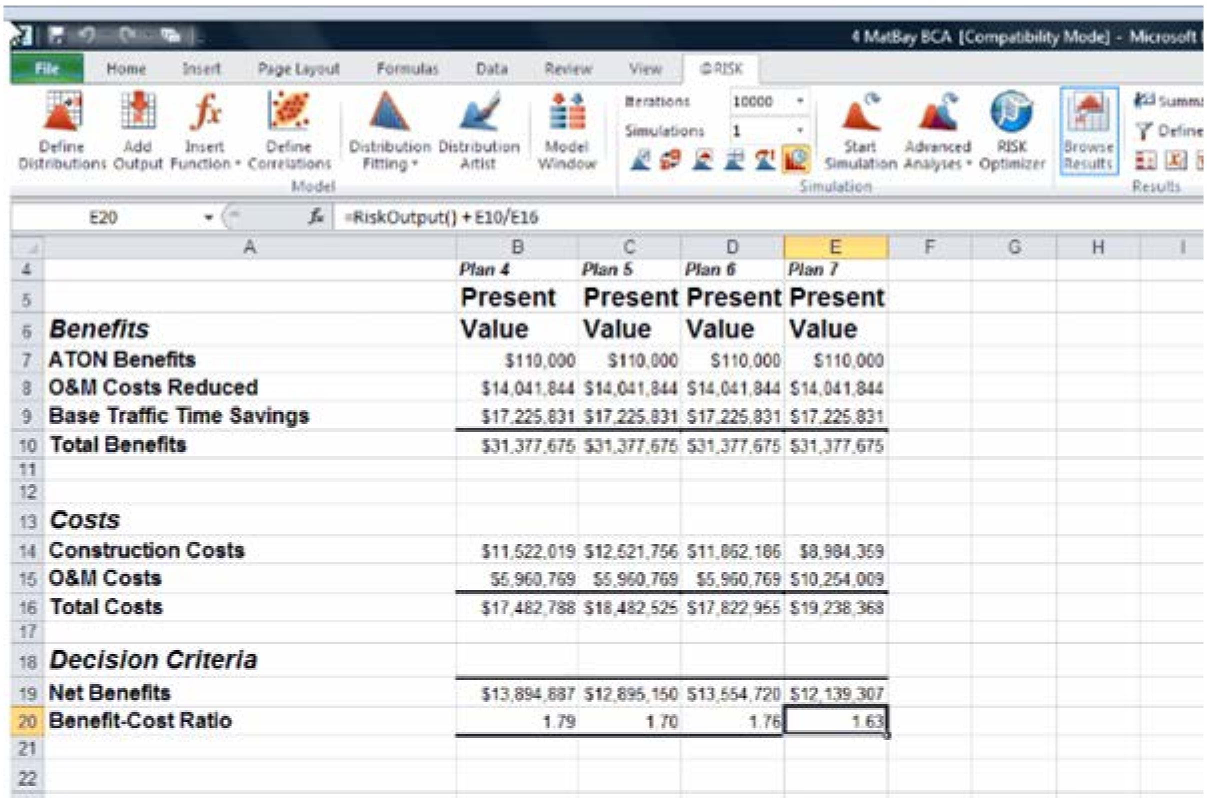
	Plan 4	Plan 5	Plan 6	Plan 7
	Present Value	Present Value	Present Value	Present Value
Benefits				
ATON Benefits	$110,000	$110,000	$110,000	$110,000
O&M Costs Reduced	$13,678,009	$13,678,009	$13,678,009	$13,678,009
Base Traffic Time Savings	$17,717,001	$17,717,001	$17,717,001	$17,717,001
Total Benefits	$31,505,010	$31,505,010	$31,505,010	$31,505,010
Costs				
Construction Costs	$8,893,896	$12,155,049	$11,860,214	$8,466,365
O&M Costs	$6,023,992	$6,217,263	$6,181,901	$10,321,863
Total Costs	$14,917,887	$18,372,311	$18,042,115	$18,788,228
Decision Criteria				
Net Benefits	$16,587,122	$13,132,698	$13,462,894	$12,716,781
Benefit-Cost Ratio	2.11	1.71	1.75	1.68

Figure 43. Benefit-Cost Ratios

3.5 Example: Decision Making – Benefit-Cost Ratio Model – Simulation

In the previous example, the net benefits are changing with the inputs. So, it becomes a challenge to really understand what the best project should be. So, a different tool, called the Monte Carlo process, can be used to run a simulation. It is a tool to analyze risks and uncertainties. There are a number of different possibilities.

Below in Figure 44, you see the results of 10,000 calculations of the benefit-cost ratio made with a Monte Carlo process.



	Plan 4 Present Value	Plan 5 Present Value	Plan 6 Present Value	Plan 7 Present Value
Benefits				
ATON Benefits	$110,000	$110,000	$110,000	$110,000
O&M Costs Reduced	$14,041,844	$14,041,844	$14,041,844	$14,041,844
Base Traffic Time Savings	$17,225,831	$17,225,831	$17,225,831	$17,225,831
Total Benefits	$31,377,675	$31,377,675	$31,377,675	$31,377,675
Costs				
Construction Costs	$11,522,019	$12,521,756	$11,862,186	$8,984,359
O&M Costs	$5,960,769	$5,960,769	$5,960,769	$10,254,009
Total Costs	$17,482,788	$18,482,525	$17,822,955	$19,238,368
Decision Criteria				
Net Benefits	$13,894,887	$12,895,150	$13,554,720	$12,139,307
Benefit-Cost Ratio	1.79	1.70	1.76	1.63

Figure 44. Estimating the Benefit/Cost Ratio

You'll notice in Figure 45 that in this graph each plan has a different color:

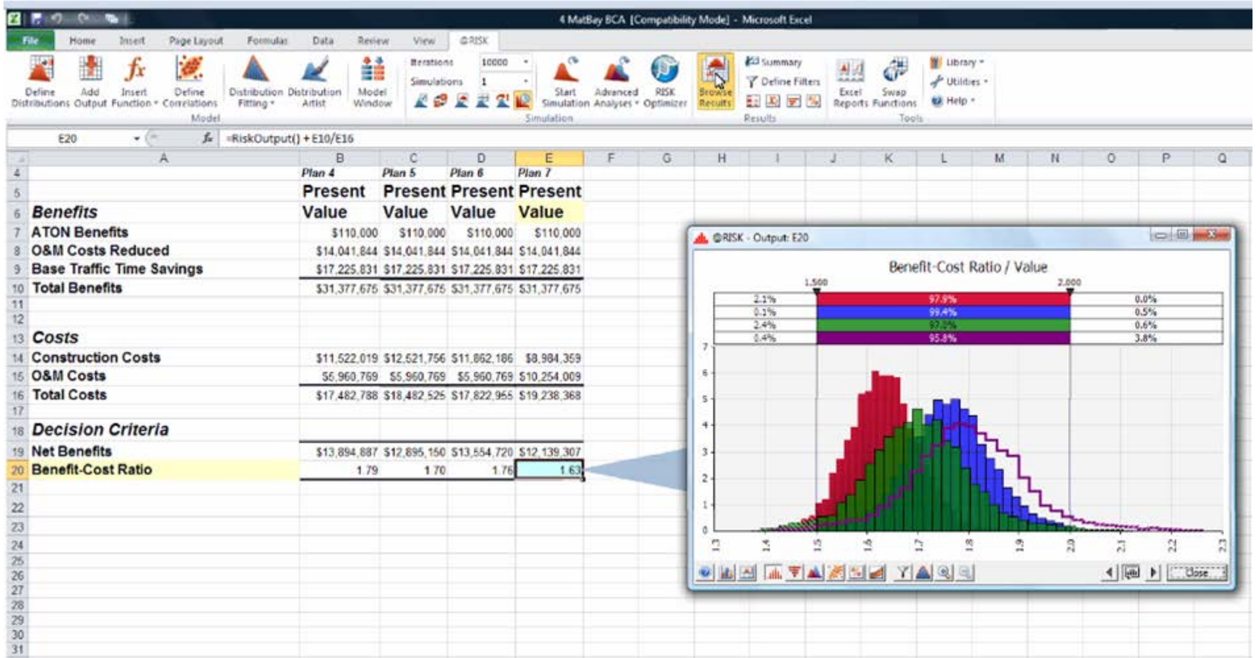


Figure 45. Estimating the Benefit/Cost Ratio

"Plan 7"is represented by the red distribution, and every other plan is represented by a different colored distribution. Notice that on the number line, some distributions seem to be oriented more toward the left-hand side of the number line, or lower benefit-cost ratios (specifically, plan 7). That would appear to have the lowest benefit-cost ratio, and the purple distribution seems to be slanted more toward the right-hand side of the line. The delimiters we see at the top indicate that 2.1% of all the values calculated for the red plan are below 1.5 in the benefit-cost ratio; and, none of the benefit-costs ratio calculations were above a 2.0 for the red plan. As we look at the purple plan, we see a different result. There's a 3.8 percent chance that the purple plan will yield a benefit-cost ratio in excess of 2, and only a 0.4% probability that it will yield a benefit-cost ratio of less than 1.5.

Considering information like this decision criteria, the benefit-cost ratio itself is uncertain, because of the uncertainty we encountered in the estimation of it. The risk manager must deal with this uncertainty.

Earlier in this module we presented an orientation to the idea of uncertainty, describing natural variability and knowledge uncertainty as two distinct sources of uncertainty. What it comes down to at the end for the risk manager is the variation and the (multiple) benefit-cost ratios.

If the uncertainty is due to knowledge uncertainty, then it's possible that by doing more research, spending more money, collecting more data, we could reduce the uncertainty and squeeze those distributions closer together to get a better estimate of the benefit-cost ratio.

But, to the extent that the distributions you are looking at are due to the natural variability in all those inputs that go into calculating all those benefit-cost ratios, then there is nothing further that can be done to squeeze the distributions. That's just the natural variability in the world.

The reason we try hard to distinguish the two sources of uncertainty is so that decision makers can understand whether we have the capability to provide them with better numbers.

Chapter 4 - Self Assessment

4.0 Self Assessment

1. With additional data it is possible to reduce the natural variability in a system.

   - True
   - False

2. There is no reason for knowledge uncertainty to remain at the time a decision is made.

   - True
   - False

3. Probability is the language of uncertainty.

   - True
   - False

4. The percent of harbor bottom that is hard bottom is an example of what kind of quantity?

   - a. Value parameter
   - b. Defined constant
   - c. Empirical quantity
   - d. Decision variable

5. Your design engineer drove to the floodplain and estimated there were 20 utilities crossing the channel. The cause of the uncertainty in this estimate is most likely which of the following?

   - a. Statistical variation
   - b. Subjective judgment
   - c. Approximation
   - d. Natural variability

4.1 Self Assessment - Answers

1. With additional data it is possible to reduce the natural variability in a system.

   True INCORRECT. *Natural variability is an intrinsic attribute of a population that cannot be reduced with additional data.*

   False **CORRECT**. *Natural variability is an intrinsic attribute of a population that cannot be reduced with additional data.*

2. There is no reason for knowledge uncertainty to remain at the time a decision is made.

   True INCORRECT. *Although knowledge uncertainty is reducible in principle it may be impractical (e.g. , too costly) or impossible to do so.*

   False **CORRECT**. *Virtually every decision of even modest complexity will involve some amount of knowledge uncertainty.*

3. Probability is the language of uncertainty.

   True **CORRECT**. *Probability statements and probability distributions can be used to model the natural variability in a system or to express our degree of belief about an uncertain value.*

   False INCORRECT. *Probability statements and probability distributions are the most common tools used to express quantities subject to either knowledge uncertainty or natural variability.*

4. The percent of harbor bottom that is hard bottom is an example of what kind of quantity?

   a. Value parameter INCORRECT. *A value parameter represent aspects of the decision makers' preferences and judgments.*

   b. Defined constant INCORRECT. *This is a constant by definition and there is no definition for channel hard bottom.*

   c. Empirical quantity **CORRECT**. *This is a quantity that can be measured. c. Empirical quantity*

   d. Decision variable INCORRECT. *No one decided how much hard bottom exists.*

5. Your design engineer drove to the floodplain and estimated there were 20 utilities crossing the channel. The cause of the uncertainty in this estimate is most likely which of the following?

   a. Statistical variation INCORRECT. *The estimate is not based on sample data.*

   b. Subjective judgment **CORRECT**. *This is your engineer's expert opinion but uncertainty about the true value remains.*

- c. Approximation INCORRECT. *The use of this language in risk analysis approximation is more appropriate for models than quantities like utility crossings.*
- d. Natural variability INCORRECT. *There is no variability on the number of utility crossings. It is a constant.*

References

[1]Morgan, M. G., Henrion, M., and Small, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, UK: Cambridge University Press.