A REVIEW OF CONTEMPORARY METHODS FOR THE PRESENTATION OF SCIENTIFIC UNCERTAINTY

K.A. Makinson*, D.M. Hamby†, J.A. Edwards‡

Acknowledgements

This work was made possible in part by funding provided from the Defense Threat Reduction Agency under research grant #HDTRA1-08-1-0044 with Oregon State University.

ABSTRACT

Graphical methods for displaying uncertainty are often the most concise and informative way to communicate abstract concepts. Presentation methods currently in use for the display and interpretation of scientific uncertainty are reviewed. Numerous subjective and objective uncertainty display methods are presented, including qualitative assessments, node and arrow diagrams, standard statistical methods, box-and-whisker plots, robustness and opportunity functions, contribution indexes, probability density functions, cumulative distribution functions, and graphical likelihood functions.

KEY WORDS

Risk Communication; Risk Estimates; Public Information; Reviews

INTRODUCTION

Data analyses across the many fields of science and engineering often contain investigations into parametric uncertainty for specific modeling applications (Hamby 1993; Hunink 2005; Halpern et al. 2006; Gallagher and Doherty 2007; Neumann et al. 2007a; Neumann et al. 2007b; McMichael and Hope 2007; Vandenberghe et al. 2007; Monni et al. 2007). Model sophistication varies from the mundane to the extremely complex. The methods of uncertainty analysis in these models cover a wide range from simple one-at-a-time parametric studies (Hamby 1994) to the estimation of model structural uncertainty (van der Sluijs et al. 2003; Lindenschmidt et al. 2007). Many analyses include Monte Carlo estimates of uncertainty (Hunink 2005; Gallagher and Doherty 2007; Neumann et al. 2007a; Monni et al. 2007; Lindenschmidt et al. 2007; Escuder-Gilabert et al. 2007; Smith and Heath 2001), with several authors using Latin Hypercube Sampling (LHS) for enhanced statistical power (Hamby 1993; Harvey and Hamby 2001; Harvey et al. 2003; Halpern et al. 2006; Helton 2009; Vandenberghe et al. 2007; Gottschalk et al. 2007). Uncertainty measures are of interest because they provide guidance on the allocation of limited resources, direct future research, or help build confidence in model-driven decisions.

A variety of scientific fields use uncertainty analysis, varying from trauma imaging (Hunink 2005) to species occupancy (Nichols et al. 2007) to radiation dosimetry (Hamby 1993). Some of the approaches used to express confidence or uncertainty include probability, predictive domain, and reasoning (Doull et al. 2007). The majority of uncertainty analyses used in these fields are straightforward and common in technique, but the presentation thereof varies considerably.

The International Organization for Standardization (ISO) (1995) defines uncertainty as the "parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand." The parameter may be, for example, a multiple of the standard deviation, or the half-width of an interval having a stated level of confidence. This parameter is typically presented either numerically (e.g. data tables, confidence intervals), graphically (e.g. box-and-whisker plots, error bars), or used to generate distributions to predict further outcomes (e.g. probability density functions, scatter plots).

The ISO (1995) characterizes two types of uncertainties (A and B) based on the method used to evaluate them. Type A uncertainty evaluations are carried out by the statistical analysis of a series of observations, i.e., repeated measurements. For example, in a Type A evaluation of normally distributed data, one would use the estimated standard deviation, equal to the positive square root of the estimated variance, as an expression of uncertainty. Type B uncertainty is evaluated by means other than the statistical analysis of a series of observations, for example, scientific judgment, manufacturer's specifications, or calibration data. It has been recommended (Braudaway 2003) that uncertainty components from Type A sources can be combined with uncertainty components from Type B sources in quadrature. This recommendation caused considerable concern in some scientific and mathematical circles because it did not appear to follow common uncertainty combination approaches. However, this recommendation has been strongly supported by organizations (ISO 1995) as well as experimental measurements of various types (Braudaway 2003).

A common misconception is that uncertainties from Types A and B sources are meant to be descriptors of random and systematic variability, respectively. This is not true. The uncertainty of a correction for a known systematic effect, for example, may in some cases be obtained by a Type A evaluation, while in other cases by a Type B evaluation (ISO 1995).

Tung and Yen (2005) state that uncertainty, in general, can be attributed to information limits regarding problem definition and resolution. They classify uncertainty in two main categories, *objective uncertainties* associated with any random process or deducible from statistical samples, and *subjective uncertainties* for which no quantitative factual information is available. They also conclude that overall uncertainty in engineering can be the result of at least four different sub-categories of uncertainty, including:

- *model formulation uncertainty* which reflects the inability of model or design procedures to represent true physical behavior;
- *parameter uncertainty* resulting from our inability to quantify model inputs and parameters with accuracy;
- *data uncertainties* which include measurement errors, inconsistency and non-homogeneity of data, data handling and transcription errors, and inadequate representation of data samples due to time and space limitations; and
- *operational uncertainty* including that associated with construction, manufacture, procedure, deterioration, maintenance and human activities.

Mathematical modeling, in which probabilistic results are generated, requires that input data be parameterized in terms of probability functions (**Fig. 1**). Once these data are generated, their visual presentation is paramount to the translation of uncertainties. Stochastic input propagated through mathematical models result in probabilistic outputs, each of which could be described in various ways. For example, scatter plots, impact matrices, probability density functions, complementary cumulative distribution functions, and coefficients of variation all provide the decision maker with a different representation of uncertainty as to the predictive power of a given model. The decision maker may, in fact, come away with a different understanding of the data depending on its presentation.

The effectiveness of predictive models can, in certain cases, be limited by an inability to adequately quantify prediction uncertainty, but more importantly by an inadequate evaluation or interpretation of those uncertainties. Many studies over the past decade have sought to quantify uncertainty in various areas of human consequence following the release of hazardous substances (Bauer and Hamby 1991; Hamby 1993; Hamby and Benke 1999; Harvey et al. 2003; Harvey et al. 2006). Probabilistic estimation is generally carried out to provide the decision maker with a sense of the range of potential consequence, for identifying the strength of a particular point estimate, or for the sake of identifying sensitive parameters in a given model to aid in directing limited resources. It is rare that probabilistic model results get compared to probabilistic consequence limits; rather, uncertainty output is often weighed against deterministic regulatory restrictions (radiation dose limits, for example).

The literature in the area of judgment and decision making (Kahneman and Tversky 1974) contains a multitude of examples of how individuals err when asked to incorporate and use probabilistic information (Dawes 1998; Kahneman et al. 1982). Conversely, the social-judgment literature is sparse on recommendations of how uncertainty information should be presented for effective decision making. There are exceptions (e,g. Gigerenzer and Hoffrage 1995; MacGregor and Slovic 1986; Wickens et al. 2000), but little related to consequence or risk analysis.

Research on optimizing the presentation of uncertainty information for decision makers has usually focused on presenting probability data in different ways. For example, researchers have tried presenting probability information as color variations, verbal expressions, frequencies, odds, visual objects with varying degrees of degradation, and graphical presentations (Schwartz and Howell 1985; Ibrekk and Morgan 1987; Kirschenbaum and Arruda 1994; Johnson and Slovic

1995; Wickens et al. 2000; Schapira et al. 2001). Overall, such efforts have met with equivocal success, although certain trends have been uncovered. For instance, line graphs lead to easier perception of trend whereas bar graphs facilitate perception of individual magnitudes or comparisons between magnitudes (Pinker, 1990). Research has also shown (Kirschenbaum and Arruda 1994; Stone et al. 1997) that graphical displays of uncertainty information are superior to verbal descriptions. However, the use of displays and graphs often requires a relatively sophisticated understanding of how to interpret the information presented.

Several methods of presenting uncertainty have emerged from different disciplines. An overview is presented of the various contemporary methods of quantifying subjective uncertainty, objective uncertainty, and uncertainty relating to modeling and sensitivity analysis.

UNCERTAINTY PRESENTATION METHODS

Subjective Uncertainty

Qualitative Assessment. A qualitative assessment, defined by Chen et al. (2007) as "non-quantitative", only speaks to potential uncertainties of input parameters and how those uncertainties might propagate through a given model. Qualitative uncertainty assessments focus on the total uncertainty present and are based inherently on human judgment. Qualitative assessments are particularly useful when quantitative data are scarce, and experts are familiar with the behavior of the system. For example, the United States Environmental Protection Agency (EPA) requires a qualitative analysis of uncertainty for Superfund human health risk assessments. In its National-Scale Air Toxics Assessment (NATA 2007), the EPA notes that uncertainties in emission parameters cannot be estimated quantitatively, and so mandates the use of qualitative judgment, based on previous experience, to determine a level of confidence for each emission source. Uncertainty estimates for a variety of parameters necessary for the NATA assessment are given "a factor of" some value, i.e., the true value lies within a "factor of" the calculated value. In some instances, uncertainties are simply expressed as having a "higher," "medium," or "lower" confidence.

Often, organizations like the EPA use impact matrices (**Fig. 2**) to portray a qualitative indication of risk perception. Environmental forces positioned on the matrix highlight the fact that not all forces are equally important or uncertain. Of the "important" cells (top section of the matrix), the "low uncertainty" forces are the relative future certainties to prepare for, while the “high uncertainty” forces are the potential shapers of entirely different futures and their corresponding preparations.

Another example is provided in a recent study by Napier et al. (2007) in which soil-to-plant concentration ratios were determined for various plant, soil, and contaminant replicates at a number of sites across the United States. A short, qualitative analysis of the uncertainties in concentration ratios determined that “variations of over a factor of 2 are not unusual.” The authors also stated that “the variability may be reduced” when averaging over a large harvest area. These qualitative statements rarely present uncertainty in detail, but can still be useful to provide a general indication of the precision of the presented data, or when presenting to a non-scientific audience.

Radar and Kite Diagram. A notational system known as NUSAP (Numeral Unit Spread Assessment Pedigree) was proposed by van der Sluijs et al. (2003) to increase the usefulness of both qualitative and quantitative uncertainty assessments for policy issues. Using NUSAP, van der Sluijs et al. (2003) highlight key uncertainties in the TIMER (Targets Image Energy Regional) model, which is an energy model that was used in the development of the 2001 greenhouse gas emission scenarios from the Intergovernmental Panel on Climate Change (IPCC).

The NUSAP method combines expert judgment with radar (**Fig. 3a**) and kite (**Fig. 3b**) diagrams. A group of experts rate their level of confidence about the numerical value of input variables and model components that were identified as most sensitive with regard to model output (projected CO_2 emissions in the TIMER case). Each variable is given its own axis and is rated on a scale of 0 to 4, 0 representing crude speculation, and 4 meaning that a large sample of direct measurements had gone into the estimate. The center of the polygon is given a value of 0 and the corners have a value of 4. In a radar diagram, the scores of the experts on each axis are connected with a line. The kite diagrams follow a traffic light analogy. The minimum scores in each group span the green kite; the maximum scores span the amber kite. The remaining area is red. The width of the amber band represents expert disagreement on the scores. This method is useful for parametric analyses and model assumption uncertainty (van der Sluijs et al. 2003). It also gives the reader a more informative view of qualitative uncertainty.

Node and Arrow Diagram. In an effort to quantify the sources of uncertainty in a water quality management program, Chen et al. (2007) use a node and arrow diagram (**Fig. 4**). Node and arrow diagrams are similar in theory to radar and kite diagrams, but attempt to show a pathway of influence for each variable. For example, a radar and kite diagram could be constructed from the six middle nodes in **Fig. 4** (problem identification, source investigation, etc).

Two tables accompany the diagram (not shown here). The first table identifies each of the outermost nodes in the diagram and asks specific questions as to the experts' confidence in their answer(s). The second table contains a written description of how uncertain the experts are, along with their associated confidence value, a number between zero and one. For example, if the expert feels he or she is 80% confident with the background data, that expert would assign a confidence value of 0.80 to that node and enter it in the second table. Once confidence is decided for each node, the confidence values are summed and divided by the number of outer nodes. **Fig. 4**, for instance, shows fifteen outer nodes used to yield an average confidence value associated with the result of a total maximum daily loads (TMDL) program, which is used as a surrogate for water quality management. If expert judgment indicates that one node should carry more weight than the others, then a weighted average can be used. The main assumption in using this approach is that each node is independent of the others. Node and arrow diagrams should be used when it is important to make each influencing factor explicit to the reader. However, it does not yield a graphical display of uncertainty; rather, it displays the *influential factors* graphically and the uncertainty numerically.

Objective Uncertainty

Standard Statistical Methods. With a given data set, one can use traditional descriptive statistical measures such as the mean, median, or standard deviation to communicate uncertainty. Typically, researchers will simply generate a table to display these values (Counil et al. 2005; Neumann et al. 2007b; Nichols et al. 2007). However, occasionally these values can be misleading for non-normal distributions. For example, a 95% confidence interval (discussed in the next section) for a log-normal distribution is $\sigma^{1.96}$ instead of σ *x* 1.96. If not specified, it is common for the reader to assume that a normal distribution is appropriate, although this assumption may not always be accurate. When basic statistical measures are not sufficient, extensions of these derived values can lead to more enlightening displays.

Aside from measurements of mean, median, and standard deviation, the statistical descriptors of central tendency, dispersion, asymmetry, and peakedness are common (Tung and Yen 2005). These four statistical product-moments are often used to describe the distributions of random variables, and, as an extension thereof, may be useful in describing or comparing uncertainties of probabilistic model output (discussed later).

Confidence Intervals. A confidence interval (CI) expresses a range (percentiles) about an estimated mean of an observed quantity (i.e., calculated from observations) used to produce numeric or graphical representations of possible alternatives. In other words, if a researcher uses a 95% confidence interval, he is saying that if the same population were randomly sampled in a new experiment, there is a 95% chance that the true mean lies within the confidence interval around the estimated mean.

One of the more common confidence intervals includes the 5^{th} and 95^{th} percentiles, as displayed by Vandenberghe et al. (2007) (**Fig. 5**). This figure displays a 90% confidence interval, as there is 5% of the data not represented on either end. Vandenberghe et al. use CIs to show the uncertainty in the time prediction of NO_3 exceeding 3 mg L^{-1} from upstream to the mouth of the Dender River in seven Belgian cities.

Confidence intervals have the advantage of being very versatile, in that they can be applied in diverse disciplines. For example, Jones and Kay (2007) provided uncertainty analysis in the form of the variances of two errors: a rainfall-catchment calibration error and a generalization error, one that accounts for random variations from the "true" model parameter values. These uncertainties are presented as bounds, at the 90^{th}, 95^{th}, and 99^{th} percentiles. Clough et al. (2005) illustrated uncertainty in their results with a CI while determining the methylmercury mass fraction. Su et al. (2007) used CIs to demonstrate the accuracy of radiation dose delivered to the prostate during brachytherapy. Conclusions were then made regarding the number of radioactive seeds necessary, their strength, their placement, and the prostate volume.

Confidence intervals are not always represented graphically. McMichael and Hope (2007) use a Generalized Likelihood Uncertainty Estimation (GLUE) which is a Bayesian Monte Carlo-based approach. This approach computes 5% and 95% CIs that capture the uncertainty in model predictions arising from uncertainties in model parameterization. In a study on the emissions of CH_4 and NO_2 from farm animals in Finland, Monni et al. (2007) use CIs for several input parameters. They also use confidence intervals for conclusions based on statistical distributions stating that, for example, "a horse emits 18 kg of CH_4 per year, ± 50% based on a normal distribution."

While confidence intervals are very enlightening in the case of normal distributions, they can lack power when used with other distributions. For example, typical confidence intervals would not be appropriate for displaying binomial data. Another drawback is the lack of density information. For example, the estimated mean of three data points expressed with a confidence interval could look identical to the estimated mean and interval of three-hundred data points distributed in a similar fashion. For the same reason, clusters of points are also impossible to discern.

Standard Error Bars. Similar to confidence intervals, **s**tandard error bars are a simple way to represent statistical variability in a mean estimate. They generally provide an estimate of the standard deviation of data, or some multiple thereof. Chao et al. (2007) use standard error bars (**Fig. 6**) to depict the standard deviation of time relative to radioactive dating of plant species. This predicted uncertainty in time is a function of radiological detection methods, specifically, counting time and amount of radioactivity present in the sample.

Standard errors are often used as a quick indication of uncertainty because they can easily be generated in most statistical processing programs. For normal distributions, standard error bars correspond to a 68% confidence interval. A rule of thumb when hypothesis testing the difference in two estimated means of normally distributed data is if the standard errors overlap, the difference in means is not significant.

Authors will occasionally use multiples of standard error. For example, one author could use error bars to illustrate two standard deviations (~95% for a normal distribution), while others may use them to denote one (~68%), or three (~99%). When interpreting standard error bars, it is commonplace to assume one standard deviation unless otherwise noted.

Standard error bars have similar shortcomings to error bars: lack of density and distribution information. Standard error bars also may interfere with data legibility if the data points are too close together as Pliel et al. (2007) showed in modeling pharmacokinetics for inhalation exposure to methyl tertiary butyl ether.

Box-and-Whisker Plots. Box-and-whisker plots, often referred to simply as "box plots," are generated to provide a visual indication of the amount of variability in deterministic output, describing the variability as mean, median, 25^{th}, 75^{th}, and sometimes 95^{th} or 99^{th} percentiles. The edges of the box are usually defined as the upper and lower quartile; the middle line in the box is the median. Often there is a cross or dot in the center to indicate the mean. The "whiskers" of the box show the largest and smallest values of no more than 1.5 box-lengths from the upper and lower quartile, although they are sometimes used to express the 95% confidence intervals (Makinson 2009), or minimum and maximum values (Whyatt et al. 2007). If the 1.5 box-length whiskers are used, there could also be extreme points denoted that lie more than 1.5 box-lengths from the upper and lower quartile (Ramsey and Shafer 2002). Whyatt et al. (2007) use box-and-whisker plots overlain on maps of the UK to demonstrate uncertainty in modeled concentrations of sulfate- and ammonium-containing aerosols (**Fig. 7**). These figures allow the reader to quickly assess regional uncertainty. Gottschalk et al. (2007) use a form of the box-and-whisker plot to visualize measurement uncertainty and global uncertainty (standard deviation), relative to deterministic estimates of grassland net ecosystem exchange.

Box plots are one of the simplest and most informative uncertainty display methods, which explains their relative frequency in the literature. And, although box plots do a better job at describing the distribution shape than confidence intervals and standard error bars, their greatest drawback is their lack of density information.

Scatter Plots. A generally straightforward way to demonstrate the variability associated with an output parameter is with scatter plots. Often, scatter plots are generated from a Monte Carlo analysis; for example, Gallagher and Doherty (2007) use scatter plots to illustrate uncertainty in estimating watershed model parameters using a random-sampling approach. Storlie and Helton (2006) use scatter plots to show the uncertainty in a quadratic regression model for a sensitivity analysis (**Fig. 8**). Notice how easy it is to detect the cluster of points between $10^{-11.5}$ and $10^{-11.0}$ m^2 that would otherwise go unnoticed with box plots, confidence intervals, or standard error bars. When multiple groups of data are represented, research shows (Lewandowsky and Spence 1989) scatter plots that vary groups by color perform better than those using different symbols, especially when perceivers are under time pressure.

Scatter plots are particularly useful in showing the difference between predicted and measured values. For example, Neal et al. (2007) use this method to show the difference between predicted and measured values in flood simulation. Neumann et al. (2007b) use scatter plots to show the strength of their model by displaying the difference between modeled and measured values of ozone concentration. Scatter plots are useful when there are more data points than can be easily viewed numerically. Usually in scatter plots, the exact position of each point in the plot is not as important as the general trend(s) and clustering indicated by the data.

Scatter plots are also used in regression analysis to detect departures from normality (normal quartile-quartile (Q-Q) plots), high leverage or influential outliers (Cook's Distances), and to validate equal variance assumptions (residuals vs fitted values) (Ramsey and Schafer 2002). Unlike box plots, confidence intervals, and error bars, scatter plots do show density. However, sometimes trends are difficult to detect without further statistical tools such as t-tests, chi-square tests, or other measures of goodness of fit.

Probability Density Functions (PDF). Probability density functions (**Fig. 9**), sometimes called density functions, or for discrete cases, probability mass functions (PMFs), provide quantitative output measures that indicate the most likely events, bounding events, or the relative likelihood of intermediate events. Probability density functions illustrate the density of data in a particular data range (sometimes called a bin). When normalized, the density becomes the probability of a randomly selected data point falling in a particular bin. Wackerly et al. (2002) define PDFs by two characteristics, stating that $f(y)$ is a probability density function if:

$$f(y) \geq 0 \text{ for any value of y; and} \quad (1)$$

$$\int_{-\infty}^{\infty} f(y)\, dy = 1, \text{ or } \sum_{-\infty}^{\infty} f(y_i) = 1 \text{ for discrete cases.} \quad (2)$$

In other words, PDFs are always positive and after normalization, have probabilities which add to unity.

As with confidence intervals, standard error bars, and boxplots, PDFs are used across many disciplines of research. In a study on the exposure of various organisms to ozone, PDFs are used to determine how many organisms are expected to remain active after a given dose of ozone (Neumann et al. 2007a). Neumann et. al. (2007b) also use PDFs to show the variability of a rate constant while predicting performance of water treatment techniques.

Because of the overall readability of PDFs, they are generally preferred as the illustration of choice for methods involving Monte Carlo predictions. For example, Gallagher and Doherty (2007) use PDFs to illustrate probability and uncertainty in watershed model parameters while using a Monte Carlo analysis. Likewise, An et al. (2007) use PDFs from Monte Carlo simulations for microbial risk of E. coli exposure. In addition to using confidence intervals to report contamination variability in food with ochratoxin A, Counil et. al. (2005) use a three-dimensional PDF to show variability of exposures in the population, with parameters estimated through a Monte Carlo simulation.

Probability density functions are also relevant to Bayesian analysis. In Bayes' Theorem, the posterior probability, or the probability of a true hypothesis for a given data set is (after normalization) a probability density function. Graphical presentation of the posterior distribution can be used to help interpret the distribution, as in **Fig. 9**. In addition, Bayesian PDFs can include information on the Highest Density Region (HDR), also known as a Baysian confidence interval. A 95% HDR denotes the area containing 95% of the highest posterior density.

PDFs are not without their shortcomings, Ibrekk and Morgan (1987) showed that when subjects were asked to estimate the mean from a PDF, most subjects incorrectly chose the mode. PDFs can be deceiving because unless otherwise specified, the mean is not displayed. An easy work-around is to simply place a dot at the mean, as shown later in **Fig.11**.

Cumulative Distribution Function. The cumulative distribution function (CDF), also called the distribution function, or cumulative probability plot, displays the probability of a randomly chosen data point being less than a particular value. The CDF is often thought of as the integral of the PDF. However, most statisticians prefer to instead define the PDF as the derivative of the CDF. Wackerly et al. (2002) define the CDF using three properties. A CDF, $F(y)$, is a function in which:

$$F(-\infty) = \lim_{y \to -\infty} F(y) = 0\,; \quad (3)$$

$$F(\infty) = \lim_{y \to \infty} F(y) = 1\,; \text{ and} \quad (4)$$

$F(y)$ is a non-decreasing function of y.

In a study of marine reserve spacing, Halpern et. al. (2006) use CDFs to show the persistence criterion, q, at four possible reserve spacing distances, d (km) (**Fig. 10**). The persistence criterion is approximately equal to the annual persistence probability (or the annual probability of a species avoiding extinction) of spatially separated, single-species populations. While the persistence criterion is empirically derived and not identical to the probability of persistence, maximizing the persistence criterion will approximately maximize the probability of persistence. The cumulative probability shows the chance that the persistence criterion will have a value up to that specified on the horizontal axis. Conservative attitudes to risk (risk-averse) operate at the lower end of the CDFs, while optimistic (risk-taking) operate at the upper end. A q-value of at least 0.95 is guaranteed for all q-values due to the lower bounds of the dispersal distances. It is easy to see that there is no distance which consistently maximizes the q-value. If read from left to right (conservative to optimistic risk attitudes), it is observed, from the values on the right axis of **Fig. 10** that the there is an 18% chance that d=25 km results in the highest q-value, 25% chance at d=50 km, 29% chance at d=100 km, and 28% chance at d=200 km.

McCarthy et al. (2005) use CDFs to show optimal reserve configurations for habitat persistence, and earlier they (McCarthy et al. 1996) illustrated the risk of quasi-extinction as a function of threshold population. The probability of illness relative to mean countertop storage time after cooking is demonstrated using CDFs by Mokhtari et al. (2006). Gallagher and Doherty (2007) present both the PDF and CDF next to each other to illustrate the interrelation of the functions. The CDF is sometimes preferred over the PDF because it can lead to greater legibility when applied to larger amounts of information.

Unfortunately, CDFs are just as unreliable at communicating the mean as PDFs. Ibrekk and Morgan (1987) commented that "many subjects incorrectly chose the maximum" when asked to identify the mean. The CDF was shown to be the second most effective (behind error bars) at communicating a confidence interval, as the information can be directly read from the display. Consequently, they suggest that CDFs and PDFs be presented simultaneously with the mean clearly marked on each (**Fig. 11**).

Complementary Cumulative Distribution Function. The complementary cumulative distribution function (CCDF) is essentially the mathematical opposite of the CDF, or the probability of a randomly chosen data point being greater than a particular value. Thus, if $F_c(y)$ is the CCDF and $F(y)$ is the CDF, then:

$$F_c(y) = 1 - F(y)\ . \tag{5}$$

The CCDF has similar properties as the CDF except that it is a non-increasing function. The complementary cumulative distribution function is the preferred uncertainty representation in risk analyses because it provides an answer to the question, "How likely is it to be this bad or worse?" (Helton et. al. 2006). Complementary CDFs can be used to present both uncertainty and acceptability in a single plot (**Fig. 12**), where the CCPF is the complementary cumulative plausibility function, and CCBF is the complementary cumulative belief function, which can be viewed as the upper and lower bounds of possible probabilities. If the CCDF overlaps the CCPF or CCBF, the possibility then exists for events to occur that may lead to regulatory limits being exceeded or potential consequences being deemed unacceptable.

Graphical Likelihood Functions. A graphical likelihood function (GLF) is a two- or three-dimensional plot where color gradation is used to indicate confidence levels. Bourennane et. al. (2007) generated spatial uncertainty estimates (**Fig. 13**) used to compare the spread of standard deviation achieved by a soil water content simulation. In this case, standard deviation was represented by a light-to-dark gradient of a single color. Color GLFs printed in gray scale can be difficult to interpret without modification. Rodi (2007) modified his figures appropriately in the error analysis of seismic event location (**Fig. 14**). He used likelihood functions to estimate epicenter regions, bounded by colored bands indicating confidence levels. The black circle marks the maximum-likelihood estimate for the event location, and the white circle marks the measured location.

Color graphics appear to have some benefits for decision makers, especially with regards to speeding decision performance (Benbasat, Dexter, and Todd, 1986) and increased recall (Gremillion and Jenkins, 1981). Color information displays are especially beneficial to decision making when the decision maker is under time pressure (Benbasat and Dexter, 1986). Choice of color can also have an impact on the effectiveness of the graph. Graphs which use colors that vary in brightness lead to quicker perception of the differences portrayed in the graph whereas graphs that use colors that vary in hue, saturation, and brightness tend to lead to more accurate perceptions (Spence, Kutlesa, and Rose, 1999).

Modeling and Sensitivity Analysis

Accuracy Plots. Accuracy plots are computed as a measure of simulation algorithm accuracy. The accuracy plot indicates an algorithm's performance by comparison of its modeled probability interval with the proportion of results that lie in that interval. Points above the one-to-one bisector indicate an accuracy of model prediction uncertainty; points below indicate where the model prediction uncertainty is inaccurate. The goodness statistics (G and D) are average numerical indicators of accuracy. A smaller G-value denotes larger deviations from model predictions. A G-value of 1 indicates perfect model-data agreement. Likewise, the value of D is a measure of accuracy of model prediction uncertainty, but gives equal weight to accurate and inaccurate information.

Bourennane et al. (2007) use accuracy plots (**Fig. 15)** to compare simulations from two predictive models of soil water content in soil cores from 10 hectare areas in the Picardie region, France. Later, they (Bourennane et al. 2010) mapped enrichment factors (the normalized ratio between concentrations in topsoil versus deep horizon soil) of anthropogenic trace elements (e.g., bismuth, cadmium, copper, zinc, lead, and others) by sampling numerous sites in the Nord-Pas de Calasis region, France. Accuracy plots successfully demonstrated model accuracy of enrichment factors at unsampled locations by comparing predicted enrichment factors with sampled factors. In a similar fashion, Peters et al. (2009) use accuracy plots to verify spatial model accuracy while modeling species and vegetation distributions in Doodle Bemde, Belgium.

Accuracy plots are somewhat similar in appearance and utility to normal quantile-quantile (Q-Q) plots. Respectively, they tell the user how closely a model fits the data, as opposed to how closely a normal distribution fits the data. Accuracy plots, assisted by the goodness statistics, allow the reader to quickly and easily assess model-data agreement.

Sensitivity Analysis. Sensitivity analysis techniques are quite similar to parametric uncertainty methods. The technique determines the amount of influence each input parameter has on model output. Parameters that are ‘sensitive’ are those that result in the greatest amount of uncertainty in their own value propagated through the model to the uncertainty of the output. Sensitivity is not only a function of input uncertainty, but is also quite dependent on model structure and complexity.

A thorough sensitivity analysis can provide useful information when estimating overall uncertainty or determining one’s level-of-confidence in parametric assessments (Bauer and Hamby 1991; Hamby and Tarantola 1999; Hamby 2002; van der Sluijs et al. 2003). However, sensitive parameters do not necessarily translate to increased output uncertainty unless the parameter is under-characterized (Hamby 1994; Saltelli 2000). Under-characterization in sensitivity analysis is expressed as the lack of knowledge about a particular parameter. For example, if only a range was known for a particular parameter, then one would define a uniform distribution about the range instead of a single value. If one knew a range and a most likely value, a triangular distribution would be used. However, if one knew more about the parameter, for example, the mean, median, mode, distribution type, etc… then the parameter would be well-characterized.

Investigators (van der Sluijs et al. 2003) have used sensitivity data in a number of different ways, primarily with the intent of demonstrating how model input variability leads to output variability. **Fig. 16** is a diagnostic diagram adapted from key uncertainties in the TIMER (Targets Image Energy Regional) model (van der Sluijs et al. 2003). The sensitivity axis measures normalized importance of quantitative parameter uncertainty. The strength axis displays the normalized average pedigree scores from a group of experts. Error bars indicate one standard deviation about the average expert value, to reflect disagreement of scores. The strength axis has 1 at the origin and zero on the right. This way, the more “dangerous” variables lie in the top right quadrant of the plot (high sensitivity, low strength). In **Fig. 16**, the most dangerous variables would be structural change, population scenario, and Autonomous Energy Efficiency Improvement (AEEI). These parameters can then be targeted for further research in an effort to reduce overall model uncertainty.

Contribution Index. A contribution index, proposed by Vose (2000), is expressed as the normalized percentage change in standard deviation with respect to the standard deviation of global uncertainty. In other words, it is an importance measurement of each factor in the overall model output uncertainty. The contribution index is calculated by running a Monte Carlo simulation to sample input factors from predefined distributions based on known information (such as bounds, or most likely values) about the factors. The parametric effect of uncertainty is determined by two separate simulations: one with the parameter defined precisely, and one with the parameter defined as a distribution; in both cases, all other factors are allowed to vary within their defined range. The difference in model uncertainty between the two simulations represents the effect of that parameter on total uncertainty. The simulation is then repeated for the number of assessed input factors, using the pair of simulations described above. The ratio of the individual contribution to the total uncertainty is expressed as a percentage. These can be positive or negative, but by definition, the sum of all contributions is 100%. A positive contribution indicates an increase in global uncertainty when the factor is varied; a negative value indicates a decrease in global uncertainty. The contribution index, c_i is calculated using:

$$c_i = \frac{\sigma_g - \sigma_i}{\sum_{i=1}^{i_{\max}} \left(\sigma_g - \sigma_i\right)} \text{x} 100\% \quad , \tag{6}$$

where c_i is the contribution index in percent of factor i, i_{max} is the total number of model input factors considered, i is the specific input factor of interest at a time, σ_g is the standard deviation of the global uncertainty, and σ_i is the standard deviation of the simulations while setting factor i to its default value. Smith and Heath (2001) use the contribution index to examine uncertainties in forest carbon budget model (FORCARB) estimates of carbon levels within discrete pools. The individual estimates of carbon pools were summed to determine total forest carbon level per unit area, with associated parametric uncertainty expressed as a contribution index.

Gottschalk et al. (2007) use polar plots of the contribution index, also referred to as “radar plots” (**Fig. 17**), to show the importance of input variables, relative to output (global) uncertainty, for a simulation of grassland Net Ecosystem Exchange (NEE). The effect of measurement uncertainties in the main input factors for climate, atmospheric CO_2 concentration, soil characteristics, and management on output uncertainty of NEE prediction was displayed using these plots. The contribution index is particularly useful for establishing quantitative uncertainty estimates for deterministic models scarce in initial information. More generally, the contribution index is used to quantitatively illustrate the percentage contribution of each input parameter on global uncertainty.

Multiple PDFs. In complex probabilistic assessments, researchers may have an interest in evaluating the impact or sensitivity of assumptions made about parameter uncertainty. This evaluation can take place by varying input uncertainty measures such that output uncertainty is generated many times to produce multiple probability density functions. Then, the variability between individual PDFs can be observed, rather than investigating the performance of one particular PDF. Plots of multiple PDFs are sometimes referred to as spaghetti plots, cobweb plots, or multiple epistemically uncertain curves. Halpern et. al. (2006) use multiple PDFs to show possible minimum and maximum distances for marine reserve spacings. Neal et. al. (2007) use multiple PDFs to show the variability in depth of water as modeled by a Monte Carlo analysis of flood probability.

Helton (2009) uses multiple PDFs to illustrate solutions to the first 50 elements of a Latin hypercube model (**Fig. 18a**). One can observe the density and spread of the curves to obtain the same information as a singular PDF at a specific time (e.g., 0.10 sec). As shown in **Fig. 18b**, a summary plot is often provided to help digest the data with 5^{th} and 95^{th} percentiles, mean (expected value), and median. Storlie and Helton (2006) also generate spaghetti plots to illustrate time dependent two-phase fluid flow results during a sensitivity analysis. Hamby and Tarantola (1999) use cobwebs plots for parameter sensitivity analysis. The density of a group of lines in spaghetti or cobweb plots, as well as the overall trend of the data, is more meaningful than the paths of individual lines.

The biggest advantage of these plots is the virtually limitless variation of input parameters. However, this advantage also comes at a cost: too much variation leads to a lack of legibility.

Bivariate Plots. A bivariate plot (**Fig. 19**) is used by Escuder-Gilabert et al. (2007) to assess whether uncertainty intervals for the bias (E±U(E)) and intermediate precision (RSD_i±U(RSD_i)) are included within prefixed limits. The authors start by fixing N_r replicate analyses during N_s runs to decide on an experiment design. A preliminary simulation is run (by varying N_r and N_s) to determine the optimal values of N_r and N_s that provide the greatest precision and accuracy. This method is intended for validating accuracy of results when only a single method-validation experiment was conducted. The researcher is left with a single estimate of bias and precision (respectively, relative error, $E = \frac{\mu - \mu_0}{\mu_0} x100\%$, where μ_0 is the accepted mean and μ is the measured mean, and relative standard deviation, $RSD_i = \left(RSD_r{}^2 + RSD_{run}{}^2 \right)^{0.5}$, where RSD_r is the relative standard deviation under repeatability conditions and RSD_{run} is the relative standard deviation from random run effects), but no estimate of their variability if more validation experiments were conducted. The authors suggest the use of Monte Carlo methods for quantifying uncertainty about the relative standard deviation and error as opposed to simply checking whether E or RSD_i estimates are below limiting values. If regulatory limits are exceeded for a given value of N_r or N_s, the authors recommend simulating whether more duplicate analysis (larger N_s values) can solve the problem. If this does not help, the laboratory must consider further method development. **Fig. 19** was adapted from a case study involving the accuracy assessment of a particular method for monitoring nitrate levels in drinking water. The authors assert that this plot "facilitates visual interpretation even for unqualified laboratory staff."

This approach is similar in philosophy to interval hypothesis testing, except that uncertainty intervals (rather than statistical confidence intervals) are estimated from Monte Carlo simulations. Bivariate plots are useful for simultaneously presenting bias assessment and intermediate precision, enabling the reader an overall view of accuracy.

Information-Gap Decision Analysis. Information-gap (info-gap) decision analysis is useful in situations of severe uncertainty, where sensitive parameters are not well characterized, such as in system modeling. Robustness and opportunity functions are used by Ben-Haim (2004) as part of an info-gap analysis to demonstrate the uncertainty severity degree. These functions provide decision makers with a tool to assess modeling uncertainty and its impact on the decision making process.

The critical stability parameter, r_c, also known as the critical level of decision-error, shown as the horizontal axis in **Fig. 20**, is defined as the "decision maker's minimal requirement for stability of the algorithm." The windfall aspiration parameter r_w, also displayed on the horizontal axis, represents a small level of decision fluctuation (much smaller than r_c) that is not essential, but would be a highly desirable windfall. D(x) represents a decision based on data "x." D(x) may be continuous and real-valued (to represent a choice of time, location, etc…) or integer-valued to represent distinct options (such as act, don't act; options a, b, or c, etc…).

The robustness info-gap function, denoted by $\hat{\alpha}(D, r_c)$, describes the "immunity to failure." A larger $\hat{\alpha}$ implies that the data can vary significantly without unacceptably influencing the decision. A larger value of $\hat{\alpha}$ is better than a smaller value: a large $\hat{\alpha}$ implies that the data can vary significantly without exceeding the critical stability parameter; conversely, a smaller $\hat{\alpha}$ implies the data are dangerously unstable. The monotonic increase of $\hat{\alpha}$, shown in **Fig. 20**, expresses the trade-off between instability and significant immunity to failure.

The opportunity info-gap function, denoted $\hat{\beta}(D, r_w)$, describes the "immunity to windfall" (i.e., robustness to failure) of decision stability. In other words, the opportunity function assesses the degree of propitious variability of the initial data, with respect to the decision maker's windfall aspiration, r_w. A small opportunity function implies the decision algorithm is opportune, i.e, windfall is an immediate possibility, although the windfall will also be small in this case. A small opportunity function is desirable because small variations from the estimated value result in windfall. The monotonic decrease of $\hat{\beta}$ as shown in **Fig. 20** illustrates the trade-off between great aspiration and certainty of windfall.

A simple example will help. For the sake of this example, we assume linear robustness and opportunity functions, although this is not always the case. Say, a new intern was told by his company that they would pay for his travelling expenses; the company, however, did not specify exactly how much they are willing to pay. The intern thinks that around 100 dollars would be reasonable. However, if the company only pays him 40 dollars, he will not be able to afford rent when he arrives. Forty dollars is defined as the critical value, or r_c. This means that the robustness function, $\hat{\alpha}$, is 60 dollars. In other words, the actual amount he is paid can vary by 60 dollars from his best estimate and he will still be able to afford rent. However, if the company pays him 130 dollars, he will be able to not only afford rent, but can also pay for an oil change. 130 dollars is the windfall aspiration parameter, r_w. A 30 dollar increase from the estimate is now defined as the opportunity function, or $\hat{\beta}$.

Fig. 20 uses info-gaps to represent the propagation of uncertainty in the Sandia Algebraic Challenge Problem from the 2002 Epistemic Uncertainty Workshop. If a decision maker's minimum requirement for stability, r_c is 0.28, the fractional error must be less than 0.28; then by observing the non-linear robustness of the data, $\hat{\alpha}$, is less than 0.5. If the requirement is less demanding, the data fluctuation is not problematic because the robustness far exceeds the observed variation of the data. However, if the decision maker aspires to a much lower prediction error, the data are not robust and unable to accommodate decisions with this level of prediction error. A similar reasoning can be applied to the opportunity function, $\hat{\beta}$. **Fig. 20** illustrates that windfall performance is possible, though not guaranteed over the r_w range evaluated. For example, if the decision maker aspires to a prediction error of 0.05, it is possible with an opportunity function of 0.22.

One engineering application is the vibration analysis of a cracked beam (Wang 2004), where the influential factors to vibration dynamics such as location, size, shape, and orientation of the crack are not well characterized. Info-gap analysis allows one to determine the robustness degree of outputs (e.g., vibration amplitude, natural frequencies, and natural modes of vibration) to the uncertain influential parameters. Halpern et. al. (2006) use info-gap analysis to assist in deciding optimal marine reserve spacing. Troffaes and Gosling (2011) compare info-gap analysis to imprecise probability theory for the detection of exotic infectious diseases in animal herds. They surprisingly conclude that the set of maximal options can be inferred partly, and sometimes entirely, from an info-gap analysis.

Ben-Haim proposed that info-gap theory should be used in situations of severe uncertainty. However, Sniedovich (2007) criticized info-gap theory for this very reason. Sniedovich argued that under severe uncertainty, one should not start from a point estimate, because by definition, it is assumed to be seriously flawed. Instead, one should consider the whole universe of possibilities, not subsets thereof (i.e., under severe uncertainty, one should use global decision theory instead of local decision theory). In general, however, any modeling that considers only a subset of the entire range of possibilities will be susceptible to low probability, high impact outcomes. Even as models become increasingly complex, it is naïve to expect models to predict every possible outcome. Therefore, all modeling deserves this criticism to some degree. When sensitive parameters are not well characterized, one can make insights about the robustness of the data that might otherwise go unnoticed without info-gap analysis.

SUMMARY AND DISCUSSION

Probabilistic estimation is generally carried out to provide the end-user with a sense of the range of potential consequences for identifying the strength of a particular point estimate, or for the sake of identifying sensitive parameters in a given model to help direct limited resources. The presentation of uncertainty can be important to decision makers to assist in directing future research or for building confidence in model-driven decisions. This assistance is the primary reason for the development of techniques to present uncertainty.

However, the various techniques do not convey the message equally well in all situations. Displaying confidence intervals and standard error bars are appropriate when the density and distribution shape are not as important as a quick indication of the median or mean value combined with a display of the variability. For those seeking more information about the shape of the data, such as how it divides into quartiles, box-and-whisker plots are more appropriate. When mean and median values are not as valuable, scatter plots do an excellent job of visually conveying both distribution shape and density information for two parameters. However, if the user was interested in the same information (i.e., highlighting density and distribution shape, but not median and mean) about only one parameter, PDFs, CDFs, and CCDFs display this information optimally. To decide between using PDFs, CDFs, or CCDFs, one simply needs to identify the question being asked. If it is most beneficial to show distribution shape, a PDF should be used; if the question regards the probability of a point being less or more than a particular value, CDFs or CCDFs, respectively, are suitable. However, the user should not feel limited to displaying PDFs, CDFs, or CCDFs. Often, PDFs are presented side-by-side with either CDFs or CCDFs to increase legibility of the data. When uncertainty measures about more than two parameters are needed, two-, or three-dimensional GLFs are often the only way to effectively communicate these assessments. Graphical likelihood functions are also often used to illustrate spatial uncertainties.

Uncertainty estimates are also regularly displayed visually during mathematical modeling and sensitivity analysis. Accuracy plots show how an algorithm agrees with theory or experimental results. Sensitivity analysis illustrates how the variation of input parameters leads to output uncertainty. Diagnostic diagrams are often a valuable tool for displaying many and quickly identifying sensitive parameters. If it is beneficial to display fewer input parameters in greater detail, one could use contribution index radar plots. Furthermore, if the distribution shape of one particular sensitive parameter was of interest, a multiple PDF could be generated to show more precisely the shape of the variation. In cases of deficient information, such as when only a single method-validation experiment was conducted, bivariate plots can be generated to show if uncertainties are within predetermined limits. In cases of even more highly deficient information, such as when multiple sensitive parameters are not well characterized, info-gap analysis can be used. Robustness and opportunity functions are displayed as a decision making tool to illustrate the robustness of the data for aspired levels of prediction error.

Although increased information helps to give a more complete picture of the data, it also can make interpretation of the presented material more cognitively difficult for the reader and thereby increase the potential for confusion and interpretational error. These problems are magnified under conditions that limit decision makers' ability to think effortfully about the material. Such conditions include time pressure, environmental distraction, multitasking, and stress. Therefore, for a given decision, it is prudent to choose a method that minimizes the number of parameters displayed, leaving only the information pertinent to the decision at hand. Displays showing fewer parameters demand less cognitive effort for the decision maker and are less likely to be affected by conditions that limit effortful thought. In addition, it is possible that the amount of experience that a decision maker has with a given display type can affect how cognitively tasking the method is for the person to use, and thus the likelihood of interpretational errors. Novice users will need to have enough detailed information and attentional capacity to understand the logic of the display, and thus may be especially affected by conditions that make it difficult for effective understanding. Expert users, on the other hand, may have automatized the cognitive process underlying perception of the graph through repeated exposure, rendering them relatively immune to conditions that hinder effortful cognition. Therefore, the choice of display method will depend on both the type of end user and the conditions under which the end user is likely to use the display. Ultimately, it is the user's responsibility, assisted by the above discussion, to decide which information and display method are appropriate for a given decision.

REFERENCES

An, Y.J.; Yoon, C.G.; Jung, K.W.; Ham, J.H. Estimating the microbial risk of *E. coli* in reclaimed wastewater irrigation on paddy field. *Environmental Monitoring and Assessment*. 129:53-60; 2007.

Bauer, L.R.; Hamby, D.M. Relative sensitivities of existing and novel model parameters in atmospheric tritium dose estimates. *Radiation Protection Dosimetry*. 37(4):253-260; 1991.

Benbasat, I., and Dexter, A. s. An experimental evaluation of graphical and color-enhanced information systems under varying time constraints. *Management Information Systems Quarterly*. 10(1), 59-81; 1986.

Benbasat, I., Dexter, A. S., and Todd, P. The influence of color and graphical information presentation in a managerial decision simulation. *Human-Computer Interaction*. 2, 65-92; 1986.

Ben-Haim, Y. Information-gap Decision Theory: Decision Under Severe Uncertainty. Academic Press: San Diego, CA. 2001.

Ben-Haim, Y. Uncertainty, probability, and information-gaps. *Reliability Engineering and System Safety*. 85:249-266; 2004.

Bourennane, H.; King, D.; Couturier, A.; Nicoullaud, B.; Mary, B.; Richard, G. Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: An empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information. *Ecological Modeling*. 205:323-335; 2007.

Bourennane, H.; Douay, F; Sterckeman, T;, Villanneau, E; Ciesielski, H; King, D; Baize, D. Mapping of anthropogenic trace elements inputs in agricultural topsoil from Northern France using enrichment factors. *Geoderma*. 150:165-174; 2010.

Braudaway, D. Uncertainty Specification for Data Acquisition Devices (DAQ). Paper presented at the annual Instrumentation and Measurement Technology Conference (IMTC). Vail, Co, USA. May 20-22, 2003.

Chen, C.F.; Ma, H.; Reckhow, K.H. Assessment of water quality management with a systematic qualitative uncertainty analysis. *Science of the Total Environment.* 374:13-25; 2007.

Chao, J.H.; Niu, H.; Chiu, C.Y.; Lin, C. A potential dating technique using ^{228}Th/^{228}Ra ratio for tracing the chronosequence of elemental concentrations in plants. *Applied Radiation and Isotopes*. 65:641-648; 2007.

Clough, R.; Belt, S.T.; Fairman, B.; Catterick, R.; Evans, E.H. Uncertainty contributions to single and double isotope dilution mass spectrometry with HPLC-CV-MC-ICP-MS for the determination of methylmercury in fish tissue. *Journal of Analytical Atomic Spectrometry*. 20:1072-1075; 2005.

Counil, E.; Verger, P.; Volatier, J.-L. Handling of contamination variability in exposure assessment: A case study with ochratoxin A. *Food and Chemical Toxicology.* 43:1541-1555; 2005.

Dawes, R. M. Behavioral decision making and judgment. In D. T. Gilbert, S. T. Fiske, and G. Lindzey (Eds.), *The handbook of social psychology* (4th Ed., Vol. 1, pp. 497-548). Boston: McGraw-Hill. 1998.

Doull, J.; Borzelleca, J.F.; Becker, R.; Datson, G.; DeSesso, J.; Fan, A.; Fenner-Crisp, P.; Holsapple, M.; Holson, J.; Llewellyn, G.C.; MacGregor, J.; Seed, J.; Walls, I.; Woo, Y.-

T.; Olin, S. Framework for use of toxicity screening tools in context-based decision-making. *Food and Chemical Toxicology*. 45:759-796; 2007.

Escuder-Gilabert, L.; Bonet-Domingo, E.; Medina-Hernandez, M.J.; Sagrado, S. A diagnostic tool for determining the quality of accuracy validation: Assessing the method for determination of nitrate in drinking water. *Annals of Bioanalytical Chemistry*. 387:619-625; 2007.

Funtowicz, S.O.; Ravetz, J.R. *Uncertainty and Quality in Science for Policy*. Kluwer. Dordrecht. 1990.

Gallagher, M.; Doherty J. Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modeling and Software*. 1000-1020; 2007.

Gottschalk, P.; Wattenbach, M.; Neftel, A.; Fuhrer, J.; Jones, M.; Lanigan, G.; Davis, P.; Campbell, C.; Soussana, J.F.; Smith, P. The role of measurement uncertainties for the simulation of grassland net ecosystem exchange (NEE) in Europe. *Agriculture, Ecosystems and Environment*. 121:175-185; 2007.

Gigerenzer, G.; Hoffrage, U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review.* 102:684-704; 1995.

Gill, J. Bayesian Methods: A Social and Behavioral Sciences Approach. Boca Raton, FL: Chapman and Hall/CRC. Pp. 69-70; 2002.

Gremillion, L. L., and Jenkins, A. A. The effects of color-enhanced information presentations. *Proceedings of the second international conference on information systems*. Cambridge, MA; 1981.

Hall, E. J.; Giaccia A. J. (2006). *Radiobiology for the Radiologist* (6th ed.). Philadelphia, PA. Lippincott Williams and Wilkins

Halpern, B.S.; Regan, H.M.; Possingham H.P.; McCarthy M.A. Accounting for uncertainty in marine reserve design. *Ecology Letters.* 9:2-11; 2006.

Hamby, D.M. A probabilistic estimation of atmospheric tritium dose. *Health Physics*. 65(1):33-40; 1993.

Hamby, D.M. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental Monitoring and Assessment*. 32:135-154; 1994.

Hamby, D.M.; Benke, R.R. Uncertainty of the iodine-133 ingestion dose conversion factor. *Radiation Protection Dosimetry*. 82(4):245-256; 1999.

Hamby, D.M.; Tarantola, S. Exploring sensitivity analysis techniques for the assessment of an environmental transport model. Proceedings of ESREL '99 – The Tenth European Conference on Safety and Reliability. Munich-Garching, Germany. September 13-17, 1999.

Hamby, D.M. The Gaussian atmospheric transport model and its sensitivity to the joint frequency distribution and parametric variability. *Health Physics*. 82(1):64-73; 2002.

Harvey, R.P.; Hamby, D.M. Uncertainty in particulate deposition for 1 mm AMAD particles in an adult lung model. *Radiation Protection Dosimetry*. 95(3):239-247; 2001.

Harvey, R.P.; Hamby, D.M.; Benke, R.R. Age-specific uncertainty of the I-131 ingestion dose conversion factor. *Health Physics*. 84(3):334-343; 2003.

Helton J.C.; Johnson, J.D.; Oberkampf W.L.; Storlie C.B. A Sampling-Based Computational Strategy for the Representation of Epistemic Uncertainty in Model Predictions with Evidence Theory. Sandia National Laboratory. Report #5557. Albuquerque, NM: 23-37. October 2006.

Helton, J.C. Conceptual and Computational Basis for the Quantification of Margins and Uncertainty. Sandia National Laboratory. Report #3055. Albuquerque, NM. June 2009

Helton J.C.; Sallaberry C.J. Illustration of Sampling-Based Approaches to the Calculation of Expected Dose in Performance Assessments for the Proposed High Level Radioactive Waste Repository at Yucca Mountain, Nevada. Sandia National Laboratory, Report #1353. Albuquerque, NM. April, 2007.

Henderson, R. A. The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clinica Chimica Acta*. 359: 1-26; 2005.

Hunink, M.G.M. Decision making in the face of uncertainty and resource constraints: Examples from trauma imaging. *Radiology*. 235:375-383; 2005.

Ibrekk, H.; Morgan, M. G. Graphical communication of uncertain quantities to non-technical people. *Risk Analysis.* 7:519-529; 1987.

International Organization for Standardization (ISO), 1995. Guide to the Expression of Uncertainty in Measurement (GUM). ISO Geneva

Johnson, E.J.; Payne, J.W.; Bettman, J.R Information displays and preference reversals. *Organizational Behavior and Human Decision Processes.* 42:1-21; 1988.

Jones, D.A.; Kay, A.L. Uncertainty analysis for estimating flood frequencies for ungauged catchments using rainfall-runoff models. *Advances in Water Resources*. 30:1190-1204; 2007.

Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica.* 47:263-291; 1979.

Kahneman, D.; Slovic, P.; Tversky, A. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press. 1982.

Kirschenbaum, S. S., and Arruda, J. E. Effects of graphic and verbal probability information on command decision making. *Human Factors.* 36:406-418; 1994.

Klawiter, B. *An investigation into the potential for eochemical/geoarchaeological provenance of Prairie du Chien cherts*. Master's Thesis. University of Minnesota. Minneapolis, MN. 2000.

Lewandowsky, S., and Spence, I. Discriminating strata in scatterplots. *Journal of the American Statistical Association.* 84, 682-688; 1989.

Lindenschmidt, K.E.; Fleischbein, K.; Baborowski, M. Structural uncertainty in a river water quality modeling system. *Ecological Modeling*. 204:289-300; 2007.

MacGregor, D.; Slovic, P. Graphic representation of judgmental information. *Human-Computer Interaction.* 2:179-200; 1986.

Makinson, K. A. *Tissue Weighting Factors: Derivation and Parametric Analysis*. Master's Thesis. Oregon State University. Corvallis, OR. 2009.

McCarthy, M.A.; Burgman, M.A.; Ferson, S. Logistic sensitivity and bounds for extinction risks. *Ecological Modeling*. 86:297-303; 1996.

McCarthy, M.A.; Thompson, C.J.; Possingham, H.P. Theory for designing nature reserves for single species. *The American Naturalist*. 165(2): 250-257; 2005.

McMichael, C.E.; Hope, A.S. Predicting streamflow response to fire-induced landcover change: Implications of parameter uncertainty in the MIKE SHE model. *Journal of Environmental Management*. 84:245-256; 2007.

Mokhtari, A.; Moore, C.M.; Yang, H.; Jaykus, L.A.; Morales, R.; Cates, S.C. ; Cowen, P. Consumer-phase *salmonella enterica* serovar enteritidis risk assessment for egg-containing food products. *Risk Analysis*. 26(3): 753-768; 2006.

Monni, S.; Perala, P.; Regina, K. Uncertainty in agricultural CH_4 and NO_2 emissions from Finland – possibilities to increase accuracy in emission estimates. *Mitigation and Adaptation Strategies for Global Change*. 12:545-571; 2007.

Napier, B.A.; Fellows, J.R.; Krupka, K.M. Soil-to-plant concentration ratios for assessing food-chain pathways in biosphere models. Report No. PNNL-16741 (NUREG/CR-6941). Pacific Northwest National Laboratory. Richland, WA. August 2007.

NATA. National-Scale Air Toxics Assessment. Limitations in the 1996 national-scale air toxics assessment: uncertainty. U.S. Environmental Protection Agency. http://www.epa.gov/ttn/atw/nata/roy/page15.html. Accessed October 2007.

Neal, J.C.; Atkinson, P.M.; Hutton C.W. Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. *Journal of Hydrology*. 336:401-415; 2007.

Neumann, M.B.; Gunten U.V.; Gujer, W. Uncertainty in prediction of disinfection performance. *Water Research.* 41:2371-2378; 2007a.

Neumann, M.B; Gunten V.U.; Gujer, W. Sources of Parameter Uncertainty in Predicting Treatment Performance: The Case of Preozonation in Drinking Water Engineering. *Environmental Science Technology*. 41:3991-3996; 2007b.

Nichols, J.D.; Hines, J.E.; Mackenzie, D.I.; Seamans, M.E.; Gutierrez, R.J. Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*. 88(6):1395-1400; 2007.

Peters, J; Verhoest, N. E. C.; Samson, R; Van Meirvenne, M; Cockx, L; De Baets, Bernard. Uncertainty propagation in vegetation distribution models based on ensemble classifiers. *Ecological Modeling.* 220:791-804; 2009

Pinker, S. A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73–126). Hillsdale, NJ: Erlbaum; 1990.

Pleil, J.D.; Kim, D.; Prah, J.D.; Rapport, S.M. Exposure reconstruction for reducing uncertainty in risk assessment: example using MTBE biomarkers and a simple pharmacokinetic model. *Biomarkers.* 12(4):331-348; 2007.

Ramsey, F. L.; Schafer D.W. *The Statistical Sleuth*. 2nd edition. Pacific Grove: Duxbry. 2002.

Regan, H. M; Hope, B. K.; Ferson, S. Analysis and Portrayal of Uncertainty in a Food-Web Exposure Model. *Human and Ecological Risk Assessment* 8(7):1757-1777; 2002.

Rodi, W.L. Error analysis in the joint event location/seismic calibration inverse problem. 29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies. Denver, CO: pp. 454-462. September 21-24, 2007.

Romney, A. K. (1999). Culture consensus as a statistical model. *Current Anthropology, 40 (Supplement)*, S103–S115.

Saltelli, A. *Sensitivity Analysis*. John Wiley and Sons. Chichester, UK. 2000.

Schapira, M. M.; Nattinger, A. B.; McHorney, C. A. Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making.* 21:459-467; 2001.

Schwartz, D. R., and Howell, W. C. Optimal stopping performance under graphic and numerical CRT formatting. *Human Factors.* 27:433-444; 1985.

Smith, J. E.; Heath, L. S.; Identifying Influences on Model Uncertainty: An Application Using a Forest Carbon Budget Model. *Environmental Management*. 27:253-267; 2001.

Sniedovich, M 2007, The art and science of modeling decision-making under severe uncertainty, *Decision-Making in Manufacturing and Services,* Volume 1, Issue 1-2, pp. 109-134.Stone, E. R., Yates, J. F., and Parker, A. M. Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied.* 3:243-256; 1997.

Spence, I., Kutlesa, N., and Rose, D. L. Using color to code quantity in spatial displays. *Journal of Experimental Psychology: Applied*. 5: 383-412; 1999.

Storlie, C.B.; Helton, J.C. Multiple Predictor Smoothing Methods For Sensitivity Analysis. Sandia National Laboratory. Report #4693. Albuquerque, NM: 11-25, 48-67. September 2006.

Su, Y.; Davis, B.J.; Furutani, K.M.; Herman, M.G.; Robb, R.A. Dosimetry accuracy as a function of seed localization uncertainty in permanent prostate brachytherapy: increased seed number correlates with less variability in prostate dosimetry. Physics in Medicine and Biology. 52:3105-3119; 2007.

Tversky, A.; Kahneman, D. Rational Choice and the Framing of Decisions. *The Journal of Business*. 59:251-278; 1986.

Troffaes, M. C. M., & Gosling, J. P. (2011). Robust detection of exotic infectious diseases in animal herds: A comparative study of two decision methodologies under severe uncertainty. *7th International Symposium on Imprecise Probability: Theories and Applications.* Innsbruck, Austria.

Tung, Y.-K.; Yen, B.-C. *Hydrosystems Engineering Uncertainty Analysis*. New York: The McGraw-Hill Companies, Inc. New York, NY: 88; 2005.

Vandenberghe, V.; Bauwens, W.; Vanrolleghem P.A. Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns. *Environmental Modeling and Software.* 22:725-732; 2007.

van der Sluijs, J.; Kloprogge, P.; Risbey, J.; Ravetz, J. Towards a synthesis of qualitative and quantitative uncertainty assessment: Applications of the numeral, unit, spread, assessment, pedigree (NUSAP) system. Proceedings of the International Workshop on Uncertainty, Sensitivity, and Parameter Estimation for Multimedia Environmental Modeling. Rockville, MD. August 19-21, 2003.

Vose, D. *Risk Analysis: A Quantitative Guide*. Wiley. New York: NY. 2000.

Wackerly, D.D.; Mendenhall W.; Scheaffer R.L. *Mathematical Statistics with Applications.* 6th edition. Pacific Grove: Duxbry. 151-154; 2002.

Wang, Kaihong. *Vibration analysis of cracked composite bending-torsion beams for damage diagnosis*. Ph.D. Dissertation. Virginia Polytechnic Institute and State University. 2004

Weber, A.H.; Buckley, R.L.; Parker, M.J.; Harvey, R.P.; Hamby, D.M. The creation of an historical meteorological database for environmental dose assessment. *Environmental Monitoring and Assessment*. 83(3): 255-281; 2003.

Whyatt, J.D.; Metcalfe, S.E.; Nicholson, J.; Derwent, R.G.; Page, T.; Stedman, J.R. Regional scale modeling of particulate matter in the UK, source attribution and an assessment of uncertainties. *Atmospheric Environment.* 41:3315-3327; 2007.

Wickens, C. D.; Gempler, K.; Morphew, M. E. Workload and reliability of predictor displays in aircraft traffic avoidance. *Transportation Human Factors Journal.* 2(2):99-126; 2000.

William L.R. Error Analysis in the Joint Event Location/Seismic Calibration Inverse Problem. *29th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*. Denver, CO: September 24-27, 2007.

LIST OF FIGURE CAPTIONS

from Storlie and Helton 2006)

Fig. 9 In probability density functions, the Highest [Posterior] Density Regions (HDR's), also known as Bayesian confidence or credible intervals, are intervals containing a specified posterior probability. A β(α,β) density function is illustrated, where the mean (μ) is defined as: $\frac{\alpha}{\alpha+\beta}$ and variance (σ^2) is:

$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ (used with permission from Gill 2002; Romney 1999)

Fig. 10 Cumulative distribution functions of the persistence probability (q) at four possible marine reserve spacing values (d in km). Percentages on the right-hand side indicate the amount of cumulative probability in that interval (used with permission from Halpern et al. 2006)

Fig. 11 A CDF and PDF presented simultaneously with the same y-axis to assist readability; the mean is indicated by a dot (used with permission from Ibrekk and Morgan 1987)

Fig. 12 An illustration of how CCDFs can be easily incorporated into a diagram with upper and lower bounds, Complementary Cumulative Plausibility Functions (CCPFs), and Complementary Cumulative Belief Functions (CCBFs), respectively, in risk analysis. The vertical axis is the probability of the plausibility, belief, and probability being greater than the corresponding pF-value on the horizontal axis, or the probability of loss of assured safety (used with permission from Helton et. al. 2006)

Fig. 13 Graphical likelihood function used to represent spatial uncertainties in soil-water content modeling. Standard deviation is expressed as a light-to-dark gradient of a single color (used with permission from Bourennane et. al. 2007)

Fig. 14 Likelihood function and confidence regions for ground based nuclear explosion monitoring at the Pahute Mesa testing area. Epicenter confidence regions determined from the likelihood function. Confidence regions are shown for 90, 95 and 98% confidence (blue (B), green (G), and red (R), respectively). The black circle marks the maximum-likelihood estimate for the event location, and the white circle marks the measurement, or "ground truth"

(GT0), or measured location (used with permission from Rodi 2007)

Fig. 15 Accuracy plot and goodness statistics (G and D) from a predictive algorithm demonstrate model accuracy by comparison of the portion of results that lie in the interval predicted by the model (above the line) to results that lie outside the predicted interval (below the line) (used with permission from Bourennane et al. 2007)

Fig. 16 Diagnostic diagram used in sensitivity analysis for key uncertainties in model parameters of the Targets Image Energy Regional model. The sensitivity axis measures normalized importance of quantitative parameter uncertainty. The strength axis displays the normalized average pedigree scores from a group of experts (used with permission from van der Sluijs et al. 2003)

Fig. 17 Polar plots of contribution index used to show the importance of input variables, relative to output global uncertainty, for thirteen parameters. A positive value indicates an increase in global uncertainty when the factor is varied; a negative value indicates a decrease in global uncertainty (used with permission from Gottschalk et al. 2007)

Fig. 18 *a:* The variability of the first 50 Latin hypercube modeling solutions, illustrated as a multiple PDF. Q is the electrical charge (coulombs) at time, t (seconds) given a realization of uncertain inputs: a (aleatory) and e_M (epistemic)

b: An extrapolation of the 95 quantile, expected value, median, and 5^{th} quantile (from Helton 2009)

Fig. 19 Bivariate plot from a study involving the accuracy assessment of a particular method for monitoring nitrate levels in drinking water. The cross-hairs extending from the solid dot (representing the mean) indicate 2.5% and 97.5% of the relative standard deviation (RSDi) and 5% and 95% of the relative error (E) generated from 10^4 simulations. Acceptation criteria limits are shown as solid lines (used with permission from Escuder-Gilabert et al. 2007)

Fig. 20 Robustness and opportunity functions used as part of an information-gap analysis to demonstrate the uncertainty severity degree. D(x) represents a decision based on data x, the critical stability parameter, r_c, is defined as the decision maker’s minimal requirement for stability of the algorithm, and the windfall aspiration parameter r_w, represents a very small level of decision fluctuation that is not essential, but would be a highly desirable windfall (used with permission from Ben-Haim 2004)

[*]Department of Nuclear Engineering and Radiation Health Physics

Oregon State University

116 Radiation Center

Corvallis, OR 97331-5303

makinske@engr.orst.edu

Tel: (541) 737-2343

Fax: (541) 737-0480

[†] Department of Nuclear Engineering and Radiation Health Physics

Oregon State University

116 Radiation Center

Corvallis, OR 97331-5303

hambydm@engr.orst.edu

Tel: (541) 737-8682

Fax: (541) 737-4678

[‡] Department of Psychology

Oregon State University

212 Moreland Hall

Corvallis, OR 9733-5303

jedwards@oregonstate.edu

Tel: 541-737-1370

Fax: 541-737-3547