Purpose

The purpose of this application is to help users measure and improve their ability to assess uncertainty.

Rationale: Assessing uncertainty and why it is important

Life, business, and, in particular, the petroleum business, are full of uncertainty. Life and business decisions are made in the midst of uncertainty.

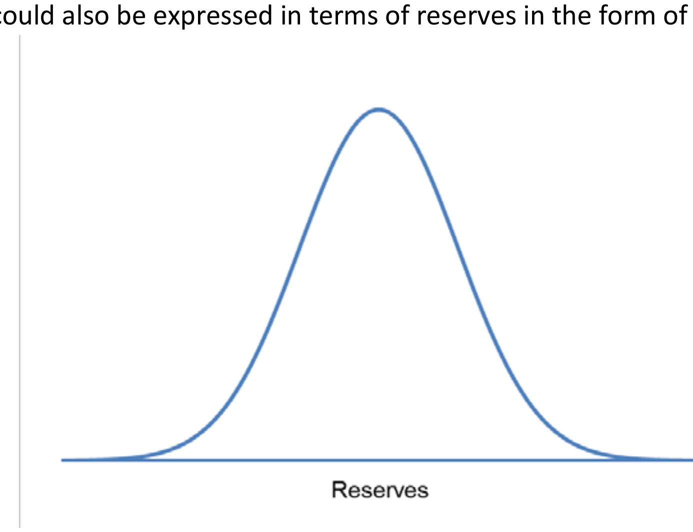
A decision consists of choices, uncertain outcomes, and values (payoffs) associated with each combination of choice and uncertain outcome. A simple decision for drilling a well is shown in the following table. For simplicity in this example, the uncertain outcome, well quality, is considered to have only two possible outcomes—good well vs bad well.

		NPV ($ millions) for different choices	
Possible outcome	Probability	Drill well	Don't drill well
Good well	70%	15	0
Bad well	30%	-10	0

The uncertainty in well quality in this example is expressed quantitatively in terms of probabilities. Quantifying uncertainty allows the decision maker to assess the upside potential and downside potential (risk) associated with the decision in order to make a good decision.

There are two fundamental types of probabilistic assessments: discrete and continuous. In the example above, the possible outcomes were expressed in terms of two discrete, exhaustive, mutually-exclusive outcomes. Exhaustive means these are the only two possible outcomes; thus, the probabilities must sum to 100%.

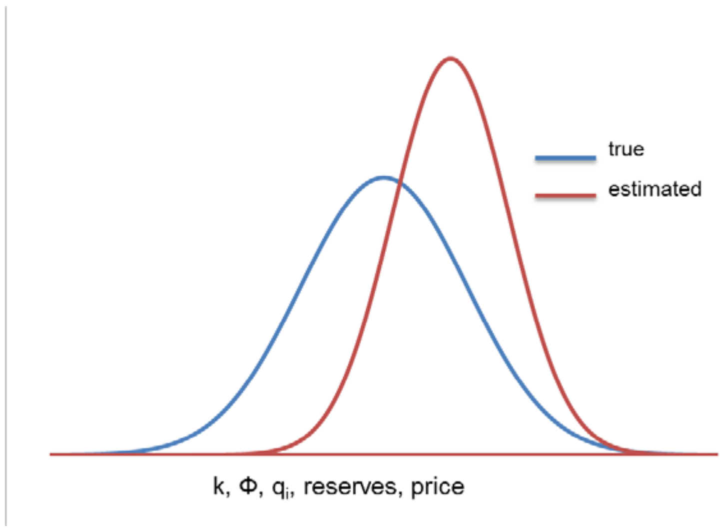
The outcome of the well could also be expressed in terms of reserves in the form of a continuous probability distribution.



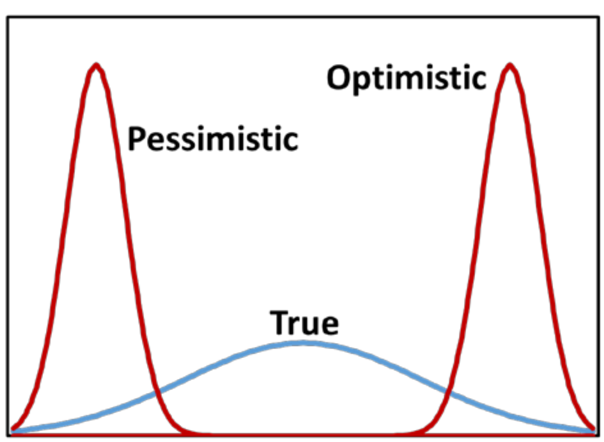
Assessing the uncertainty in possible outcomes is an important part of the decision-making process. Unfortunately, humans are poor at assessing uncertainty, i.e., we are biased. We have an almost universal tendency for overconfidence, or underestimation of uncertainty. This is well documented in many studies (e.g., Tversky and Kahneman 1974, Capen 1976). We are more certain than we have a right to be. Humans also have a general tendency for optimism, i.e., to think that things will turn out better than they do on average (e.g., Weinstein 1980, Merrow 2011). The tendency for optimism is not as strong or prevalent as the tendency for overconfidence (i.e., sometimes we overconfident and pessimistic).

For discrete probabilistic assessments, overconfidence and/or optimism results in overestimation of the probability of a desirable outcome. In the example, overconfidence and/or optimism might result in the analyst assigning a probability of 80% to a good well rather than 70%.

For continuous probabilistic assessments, overconfidence results in the estimated distribution being too narrow (too certain of the possible outcomes). Optimism results in the estimated distribution being shifted in the more desirable, or beneficial, direction (to the right in this case). The estimated distribution is both overconfident and optimistic in the following figure.



It should be obvious that if we are making decisions with unreliable uncertainty assessments, we will sometimes, if not often, make poor decisions. In general, if we are overconfident and optimistic, we will invest in some investments that we should not invest in. Likewise, if we are overconfident and pessimistic (figure below), we will not invest in some investments that we should invest in, resulting in missed opportunities.



The result of making poor decisions is reduced value. McVay and Dossary (2014) demonstrated that overconfidence and optimism in project selection can result in substantial disappointment in portfolio value. That is, the actual or realized portfolio value will be lower than the estimated portfolio value. They also demonstrated that overconfidence and pessimism also results in poor decision making, and results in portfolio values that are not high as they could be if these biases were not present.

The conclusions of McVay and Dossary are corroborated by industry performance. Many authors have reported that for decades the petroleum industry has underperformed due to chronic overconfidence and optimism in project evaluation (e.g., Capen 1976, Rose 2004, Brashear et al. 2001, Nandurdikar 2014). One key to optimizing decision making and value creation is eliminating biases, such as overconfidence and optimism, in uncertainty assessment.

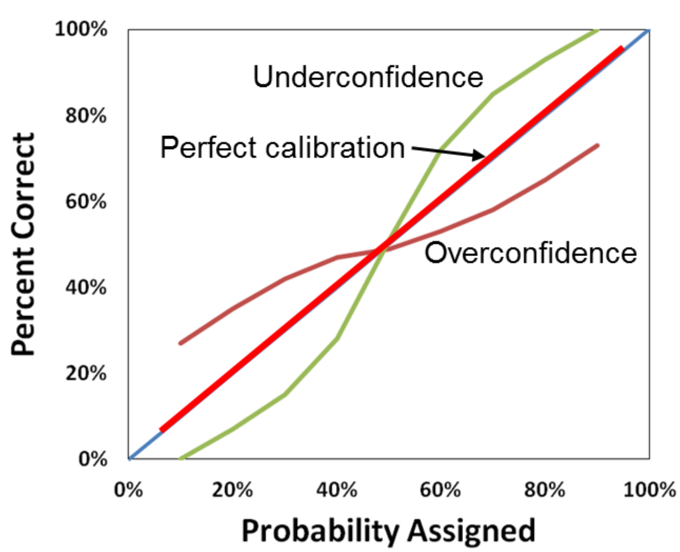
Go back to the top

How do we eliminate biases and improve uncertainty assessment?

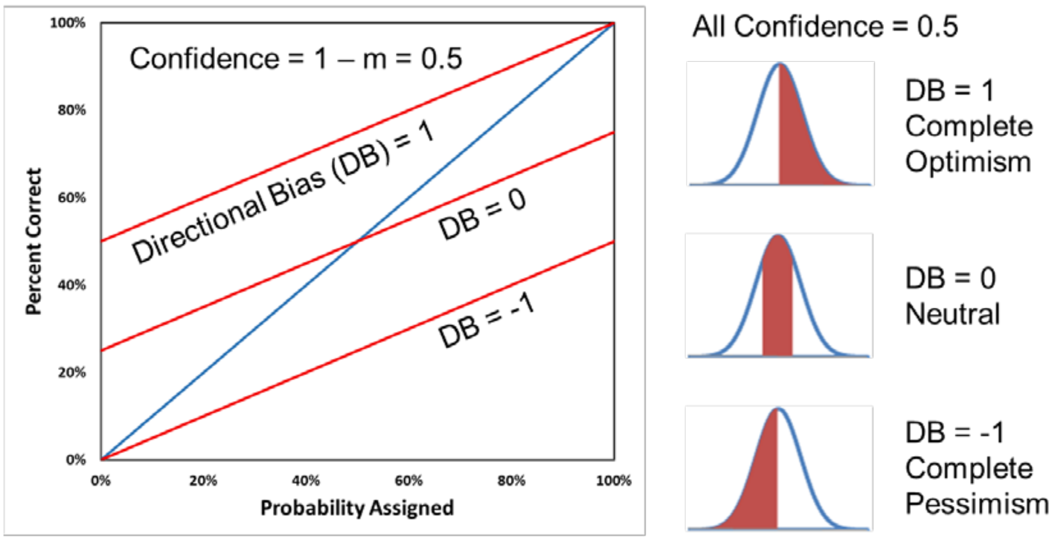
To eliminate biases, it is first necessary to measure them. The only way to measure biases is to make probabilistic assessments, or estimates, and then to compare the probabilistic assessments to the actual values when they become known.

If we are unbiased and are estimating uncertainty well, then outcomes will be consistent with the probabilities assigned. For example, with discrete assessments, if there are a 100 wells for which we have estimated a 70% probability of being a "good" well, then about 70 of these wells should turn to be "good" wells once they have been drilled and we know their performance. Continuous probabilistic assessments are often expressed in terms of cumulative probabilities. For example, a P10 means there is a 10% probability the actual value will be less than or equal to the P10 value (for the convention in which the P10 is the low number and the P90 is the high number). If our continuous probabilistic assessments are unbiased, or reliable, then the actual reserves values will be less than the P10 reserves estimates for about 10 wells out of 100 once the actual reserves are known.

The reliability of probabilistic assessments can be expressed on a calibration plot (figure below). A group of probabilistic assessments are unbiased, or reliable, if the calibration curve falls on the unit-slope line. A slope less than 1 indicates overconfidence. For example, for a group of P90 assessments, the actual value was less than the P90 estimate only 70% of the time (the analyst was overconfident by estimating the actual value would be less than the P90 value 90% of the time rather than 70%). Underconfidence (overestimation of uncertainty) is indicated by a slope greater than 1 (but this is rare).



We can also measure directional, or optimism-pessimism, biases from the intercept of the calibration plot. In the figure below, we show that we can measure confidence and directional biases quantitatively. Confidence bias (CB) ranges from -1 (complete underconfidence) through 0 (unbiased) to 1 (complete overconfidence, or a point estimate). Directional bias (DB) ranges from -1 (complete negative directionally, which is pessimism for value-based assessments) through 0 (unbiased directionally) to 1 (complete positive directionally, which is optimism for value-based assessments). See McVay and Dossary (2014) and Alarfaj and McVay (2016) for more detail on these concepts.



Biases can be measured with two different kinds of probabilistic assessments. The first is with non-forecast questions in which the correct answer is known, but uncertain to the estimator at the time of the assessment. These types of non-forecast assessments can be either discrete or continuous assessments. The advantage of these types of assessments is that, because the true answers are known, immediate feedback can be provided to help the estimator improve uncertainty estimation in a training context. Non-forecast questions can be about virtually any subject, such as science, literature, history, sport, and technology. Some may question the value of non-forecast questions in other subjects if their primary interest is improving their ability to assess uncertainty in oil and gas production forecasts. However, studies show that experts can be just as poor at assessing uncertainty as novices (Lichtenstein 1977), which indicates that assessing uncertainty is a different skill than petroleum engineering, astrophysics or any other subject. Thus, training that improves one's ability to assess uncertainty generally will benefit analysts in their subjects of expertise, vocation or interest (although the benefits may not be 1 to 1 in all subjects).

Biases can also be measured by making probabilistic forecasts of future events (with either discrete or continuous assessments), letting time pass until the actual values are known, and then measuring biases with calibration plots. One advantage of forecast questions is that the true answers are truly unknown, which makes them a better test of one's ability to assess uncertainty.. Forecast questions are also of value because the most common application of uncertainty assessment is to future events. If one's primary interest is improving ability to assess uncertainty for long-term forecasts that are years in the future (e.g., oil and gas production), it will take years to get the true answers to compare to forecasts for purposes of measuring one's ability to assess uncertainty. Answering short-term forecast questions in a variety of subjects can provide useful training that can be applied to assessing uncertainty in long-term forecasts.

Once quantitative measurements of confidence and directional biases are available, steps can be taken to improve uncertainty assessment by reducing or eliminating biases. This can occur in two ways. The first is by internal correction or improvement. Just by measuring and knowing the presence and magnitudes of biases, analysts learn to correct for them in future assessments. Usually the primary problem is overconfidence (directional, or optimism-pessimism, biases exist primarily because we are first overconfident). Learning that assessment distributions are usually too narrow, the analyst learns to widen the distributions, or ranges, with practice.

Biases can also be eliminated with external correction or improvement. Knowing the directions and magnitudes of biases, subsequent probabilistic assessments can be mathematically adjusted to correct for the biases (see Capen 1976 and Fondren et al. 2013 for examples). Ideally, the process of forecasting, measuring biases with calibration charts and scoring metrics, and using bias measurements to improve subsequent forecasts should be a continuous process over time.

Go back to the top

Making probabilistic assessments

Probabilistic assessments are made by completing an Assignment, which consists of one or more Questions. Assignments that are available for completion can be accessed from the To Do List that appears when you first log in. You can get back to the To Do List at any time by clicking on the logo at the upper left-hand corner of the screen. Assignments can also be accessed by going to Assignments>Show.

Questions can be of two types—discrete and continuous. Discrete questions have a fixed number of choices, as in True-False or Multiple-Choice questions. To complete a discrete question, you assign probabilities to all the choices such that the probabilities sum to 100%. You should assign probabilities that correspond to your relative certainty about the different choices. For example, if you are reasonably, but not completely, sure the correct answer to a True-False question is True, you might assign a 75% or 80% probability to True and 25% or 20% probability to False. If you are completely unsure, i.e., you have absolutely no knowledge about the choices, then you should assign equal probabilities—50% to each choice for a True-False question and 25% to each choice for a 4-choice Multiple-Choice question.

To save time, hitting the Auto Fill button will divide the remaining probability evenly among the blank probability fields. For example, for a True/False question, you could put 70 in one cell and hit the button, and it would put 30 in the blank cell. For a 4-choice multiple choice question, if you were reasonably sure about one choice but equally unsure about the other choices, you could put 70 in the one cell you are sure about and hit the button. It would assign 10 to each of the remaining 3 blank cells.

The answer to a continuous question is a number. You express your uncertainty about the numerical answer to the question by specifying points on a cumulative probability distribution. By default, the cumulative probabilities (percentiles) listed are 10%, 50% and 90% (P10, P50 and P90). You can change these values if you choose. You can also add and delete rows in order to specify as many points on the distribution as you want. You then assign numerical values to each of the percentiles. You can answer with either a cumulative distribution (by assigning the low number to the P10 and the high number to the P90) or a decumulative distribution (by assigning the high number to the P10 and the low number to the P90). The order of the percentiles does not matter (although if you leave the question and come back, the system will order the cumulative probabilities from low to high). However, the numerical values must either monotonically increase with increasing cumulative probabilities or monotonically decrease with increasing cumulative probabilities. If they do not, the system will flag this as an error and it must be corrected before you will be allowed to proceed.

Questions can be divided into two types in another way—forecast vs non-forecast. Non-forecast questions are questions about facts and events for which the answers are known, but likely unknown and therefore uncertain to most users. The primary advantage of non-forecast questions is that users can obtain instant feedback on their ability to assess uncertainty from calibration reports. Because the true answers are available, users must of course resist the temptation to look up the true answers when answering questions, because this defeats the entire purpose of non-forecast questions. Assignments consisting of non-forecast questions can be completed only once, because the true answers are available to the user in scoring reports.

Forecast questions are just what the name implies, forecasts of things that will occur in the future and, thus, are unknown now. One advantage of forecast questions is that the true answers are truly unknown; thus, it is impossible to cheat, intentionally or inadvertently. Assignments consisting of forecast questions can be completed as many times as desired until the date the true answer is known. A reason one might consider making multiple probabilistic assessments over time of the same future event is because their knowledge related to the future event may change over time, which means their uncertainty related to the future event may change as well.

Go back to the top

Scoring assessments

Probabilistic assessments are scored by going to Evaluation, then Scoring Report. The Scoring Options screen presents a variety of options for filtering assessment for scoring. If multiple filters are applied, a Boolean AND operation is applied. For example, selecting Forecast?: Yes and Category: Sports will result in a set of assessments that are forecasts AND have a category of Sports. Thus, each application of a filter will either reduce the number of assessments scored or leave the number of assessments the same (if there are no assessments to which the filter applies). After selecting the desired filters, hit the Submit button to generate the scoring report.

Both discrete and continuous assessments are ultimate reduced to a set of binary assessments, each of which consists of a probability assigned, an operator (= for discrete assessments, and ≤ or ≥ for continuous assessments), and a value assigned (real number for continuous assessments and an integer corresponding to choice number for discrete assessments). Reducing all assessments to a set of binary assessments allows both discrete and continuous assessments to be combined on the same calibration plot and in the same scoring metrics.

The Summary of Assessment page shows a table of Probability Assigned, Proportion Correct and Bincount, which is used to generate the calibration plot shown on the same page. Assessments are grouped into bins (currently 11) by probability assigned. For example, the first bin contains all assessments with probabilities assigned that fall between 0 and 0.0909 (1/11). For each bin, the average probability assigned is calculated for all the assessments in the bin; this value becomes the x-value on the calibration plot. For each assessment in the bin, the binary assessment is evaluated. For example, the binary assessment for a discrete assessment is actual_choice = assessed_choice. The binary assessment for a continuous assessment with a cumulative distribution convention is actual_value ≤

assessed value (e.g., the P5 value). Note that continuous assessments with a decumulative distribution convention are all converted to a cumulative distribution convention for scoring and plotting so the interpretation of biases is consistent. The evaluation of each binary assessment results in either True or False. The Proportion Correct is calculated by dividing the number of True binary assessments by the total number of binary assessments in the bin (Bincount). The Proportion Correct value becomes the y-axis value on the calibration plot for each bin.

The calibration plot shows Proportion Correct vs Probability Assigned by probability bin (maroon curve). Perfect calibration (unit-slope line) is indicated by the dashed blue line. A least-squares regression line (green), weighted by the number of assessments associated with each data point (Bincount), is fitted to the maroon calibration curve. This weighted regression line is used to calculate confidence and directional biases.

The table at the top of the Summary of Assessment page shows calculated scoring metrics. The first four columns show the Brier score and the three components of the Brier score—calibration, resolution and knowledge. The last four columns show the slope and intercept of the least-squares regression line and values of confidence and directional biases calculated from the line. Strictly speaking, calculation of confidence and directional biases applies to only continuous assessments. The interpretation of confidence and directional biases is unclear for a set of assessments that contains any discrete assessments. Also strictly speaking, resolution and knowledge apply only to discrete assessments. The interpretation of resolution and knowledge is unclear for a set of assessments that contains any continuous assessments. For a set of assessments that includes both discrete and continuous assessments, only calibration (measure of average distance of the calibration curve from the unit-slope line) and possibly the Brier score have any meaningful interpretation.

A more detailed report is available by selecting "Download log file" at the bottom right corner of the Summary of Assessment page. This downloads an Excel spreadsheet containing the assessment data, the true answers, intermediate calculations, and the scoring metrics that are presented on the Summary of Assessment page. This is used for debugging at the present, but at some point will be enhanced to become a more formal detailed report.

Go back to the top