Formats for Expressing Acceptable Uncertainty

Brandon J. Wilde and Clayton V. Deutsch

This short note aims to define a number of formats that could be used to express acceptable uncertainty. These formats are categorized as relative, absolute, and misclassification. Expressing an acceptable level of uncertainty typically requires specification of a volume, a +/- measure of uncertainty, the probability to be within the +/- measure, and the proportion of volumes required to meet the preceding criteria. A relative format is recommended in most cases. When there is an important threshold, a misclassification format is recommended.

1. Introduction

An important aspect of geostatistics deals with characterizing geologic uncertainty. This uncertainty is not an inherent feature of the geology; there is only one geologic truth. Uncertainty arises due to our limited sampling of the true distribution. Geostatistics uses all available site-specific and analogue knowledge to construct models of uncertainty, taking care to ensure that these models reasonably represent our state of incomplete knowledge. These models are often built to assist decision makers. For example, uncertainty is often used qualitatively to support decisions about collecting additional data or, perhaps, preferring locations where the uncertainty is less. In some cases, a maximum level of uncertainty is considered for classification of the area as measured or indicated. Classification is required for public disclosure. Codes for public disclosure suggest a quantification of the error associated with the estimation of the grade for classification. The petroleum resource/reserve classification scheme presented in NI 51-101 and the accompanying CIM guidelines mention specific probabilities. This is less common in the mining industry. There is no mention of probabilities in codes such as the JORC, code, SAMREC code, SEC Industry Guide 7 and NI 43-101; however, there is an expectation that the uncertainty would be quantified and used to support the final classification decision. There are different formats for expressing the acceptable level of uncertainty.

2. Uncertainty

Uncertainty is quantified by multiple high resolution geostatistical realizations. These must be constructed with care, checked with cross validation and reconciled with any production data. The uncertainty at the high resolution scale of the geostatistical model is rarely relevant for disclosure or expressing acceptable uncertainty. The high resolution models are scaled to a larger scale relevant for technical and economic decision making. Often, this larger scale represents a nominal time period for production such as a month, quarter or year. Each nominal volume in the area of interest has a distribution of uncertainty for the variable of interest. The variable of interest could be the mass fraction of an important component, a combined economic variable, or the material above a fixed economic threshold. The outcomes of multiple realizations describe the uncertainty for each nominal volume; however, it is necessary to summarize the uncertainty for comparison and reporting.

An acceptable level of uncertainty is defined for a particular purpose. This purpose is some type of decision making or classification. The decision whether some level of uncertainty is acceptable is made by a qualified person who deems the uncertainty to be acceptable for the problem at hand.

The domain for which an acceptable level of uncertainty is applicable must be defined. Within some global site, G, there could be a number of areas, $A_j, j \in G$ as shown in Figure 1. Each area would be characterized by a particular data spacing as shown for areas 3 and 4 in Figure 1. An acceptable level of uncertainty is defined for a given area. Within an area there are a number of volumes, $V_i, i \in A_j$ (area 1 in Figure 1). Each volume has a distribution of uncertainty. Establishing whether an area meets an acceptable level of uncertainty involves determining the proportion of volumes that meet a required level. The required proportion of volumes is part of the specification of acceptable uncertainty. The area can then be classified based on whether the acceptable level is met.

3. Formats

There are a number of formats for expressing acceptable uncertainty. The choice of format depends on the audience, local customs, the particular problem, transparency, and preferences of the practitioner. It is common for probabilistic uncertainty specification to include:

1. Identification of the population or sample being considered (the volume, V)
2. A measure of +/- uncertainty
3. The probability to be within the +/- measure of uncertainty
4. The proportion of volumes, V_i, within the area, A_j, required that meet the preceding criteria

An example is *the true grade of monthly production volumes will be within 15% of the predicted grade 19 times out of 20 for at least 90% of the volumes* $V_i, i \in A$. (Deutsch *et al.*, 2006). This statement of acceptable uncertainty includes a volume (monthly production volumes), a +/- measure of uncertainty (within 15%), a probability to be within the +/- measure of uncertainty (19 times out of 20, or 95%), and the proportion of volumes required to meet these criteria (90%). The second two parameters are illustrated for one volume, V_i, in Figure 2. This is just one way of specifying uncertainty. There is nothing special about monthly/15%/95%/90%, but values similar to these are commonly mentioned.

3.1. Absolute Uncertainty Formats

The sample format presented above is a relative measure of uncertainty; the measure of +/- uncertainty depends on the value of the predicted grade (*within 15% of the predicted grade)*. Another format for specifying acceptable uncertainty could be an absolute measure of +/- uncertainty that does not depend on the predicted grade. The difference between absolute and relative uncertainty is illustrated in Figure 3. Consider the random variable X. The measure of +/- uncertainty could be some constant, Δx, applied to all volumes $V_i, i \in A$. The format for specifying acceptable uncertainty could then be *the true value will be within +/-* Δx *of the predicted value 19 times out of 20 (95% of the time) for at least 90% of the volumes* $V_i, i \in A$. This means that if there are more than 10% of the volumes where the true value does not have a 95% chance of being within +/- Δx, the specified level of acceptable uncertainty has not been met. Δx is a constant applied to all volumes making this an absolute format.

Other absolute measures of +/- uncertainty include the standard deviation and difference between quantiles. Invoking the Gaussian distribution may be a reasonable assumption for large production volumes. The proportion of values within one standard deviation for the Gaussian distribution is 68% and the proportion within two standard deviations is 95%. This is illustrated in Figure 4. Therefore, the standard deviation communicates both a +/- measure of uncertainty as well as the probability to be within the +/- measure of uncertainty. For example, a standard deviation of 1 g/t means that there is a 68% chance for the grade to be within +/- 1 g/t and a 95% chance for the grade to be within +/- 2 g/t. Acceptable uncertainty could be expressed as *the predicted value will have a standard deviation less than* σ *for at least 90% of the volumes* $V_i, i \in A$.

The assumption of a Gaussian distribution is not necessary when considering the difference between quantiles. Consider the 10th and 90th percentiles of a distribution. There is an 80% (90-10) chance for the grade to be between these percentiles. Assume that the difference between the 10th and 90th percentiles for a distribution is 1 g/t. This provides a +/- measure of uncertainty as well as the probability to be within the +/- measure. This level of uncertainty could be expressed as *the difference between the 10th and 90th percentiles for the predicted grade will be less than 1 g/t for at least 90% of the volumes* $V_i, i \in A$.

3.2. Relative Uncertainty Formats

Absolute measures can be modified to be relative measures by dividing by a measure of center such as the mean or median. Dividing the standard deviation by the mean provides the coefficient of variation. The +/- measure of uncertainty now depends on the value of the distribution center making it relative. Invoking the Gaussian distribution and given a coefficient of variation of 0.15, there is a 68% probability for the true value to be within 15% of the predicted value and a 95% probability for the true value to be within 30% of the predicted value. Therefore, the coefficient of variation communicates both a +/- measure of uncertainty as well as the probability to be within the +/- measure of uncertainty. An acceptable level of uncertainty could be expressed as *the predicted value will have a coefficient of variation less than 0.15 for at least 90% of the volumes* $V_i, i \in A$.

The difference between quantiles can also be standardized by dividing by a measure of center such as the median. This +/- measure of uncertainty is relative as it depends on the value of the distribution center. Consider again the 10th and 90th percentiles of a distribution. The difference between these percentiles divided by the median provides both a measure of +/- uncertainty as well as a probability to be within the +/- measure of uncertainty. For example, consider that the difference between the 90th and 10th percentiles divided by the median is 0.5. This means that there is an 80% probability for the true value to be within 25% of the predicted value. A numerical example is useful. Consider a distribution with median=10 g/t, 10th percentile=7.5 g/t, and 90th percentile=12.5 g/t. The difference between the 90th and 10th percentiles is 5 g/t; dividing this by the median gives 0.5. There is an 80% probability to be within +/- 2.5 g/t. 2.5 g/t is 25% of the predicted value of 10 g/t. Therefore, there is an 80% probability to be within 25% of the predicted value for a standardized difference between quantiles of 0.5. The acceptable level of uncertainty could be expressed as *the difference between the 90th and 10th percentiles divided by the median for the predicted grade will be less than 0.5 for at least 90% of the volumes* $V_i, i \in A$.

3.3. Misclassification Uncertainty Formats

Acceptable uncertainty could also be specified by the risk of misclassification. Classification is a common procedure in the earth sciences. Thresholds are used to classify a volume according to some property of that volume. There is a risk of misclassification, that it, incorrectly classifying a volume as being from one category when it truly belongs in another. When there are two possible classifications, there are two possible misclassification errors. For instance, a volume may be classified as either ore or waste, depending on the grade of the volume. A waste volume could be incorrectly classified as ore or an ore volume could be incorrectly classified as waste. The former error is called a false positive or Type I error; the latter a false negative or Type II error. The probability of making a Type I error is commonly represented by α while the probability of making a Type II error is commonly represented by β. To know whether a volume has been misclassified requires knowledge of the true classification which is inaccessible in practice.

One way of evaluating the risk of misclassification involves assuming a model of the truth. The model is sampled and the samples used to generate additional models. Each volume in the model can then be compared with its equivalent volume from the truth to determine the risk, or probability, of misclassification. Volumes whose grades are near a threshold have greater risk of misclassification than those volumes whose grades are far from a threshold.

An acceptable level of uncertainty could be specified with respect to one of the types of misclassification: *the probability of Type I (or II) error will be less than 10% for at least 90% of the volumes* $V_i, i \in A$. It could also be specified with respect to both types of misclassification: *the probability of misclassification will be less than 10% for at least 90% of the volumes* $V_i, i \in A$.

4. Complicating Factors

A number of factors complicate the specification of an acceptable level of uncertainty. One factor is the choice of area to certify. There may be no clear boundaries between areas with similar data spacings. Volumes at the fringe of an area may significantly reduce the proportion of volumes which meet the specified criteria leading to improper classification of the area. Another factor is the choice of acceptable tolerances. These are chosen based on varied technical, economic, and managerial considerations and must be customized for each deposit. For example, choosing relative uncertainty over absolute uncertainty causes low grade regions to be very uncertain. Furthermore, certain deposit types have less uncertainty but cannot be considered as all measured (Deutsch *et al.*, 2006). Another consideration is how to deal with uncertainty in the model. Geologic models only quantify the uncertainty we can imagine. There are additional aspects of uncertainty that cannot be accounted for in the modeling. Geological continuity is another complicating factor. There may be more continuity in one area than in another, but realizing this from limited sample data is difficult. There is also the issue of dealing with uncertainty in the data. This is of particular concern when acceptable uncertainty is specified with respect to risk of misclassification. Erroneous data can lead to locations which are thought to be certainly true to be truly false and vice versa. Another factor is the issue of scale. Krige's relation (Journel and Huijbregts, 1978; Deutsch, 2002) dictates that as the modeling scale increases, local uncertainty decreases. The choice of modeling scale must therefore be made with care.

5. Conclusions

There are numerous formats for expressing acceptable uncertainty. Some formats utilize a relative +/- measure of uncertainty; others use an absolute+/- measure. Probability of misclassification could also be used. The volume, +/- measure of uncertainty, required precision in the +/- measure of uncertainty as well as the proportion of volumes required to meet the criteria must be determined by a qualified person. There are a number of factors to consider when choosing a format for expressing acceptable uncertainty. The format to use depends on the data spacing, geological continuity, and reasonable subsets supported by the uncertainty within an acceptable limit. In many cases the probability to be within a relative percent of the mean should be the default. If there is a critical threshold related to environmental or regulatory specifications, then risk alpha or beta should be the default.

References

Canadian Securites Commissions (CSA), 2005. National Instrument 43-101: Standards of Disclosure for Mineral Projects.

Deutsch CV, 2002. Geostatistical Reservoir Modeling. Oxford University Press. 376p.

Deutsch CV Leuangthong O and Ortiz J, 2006. A case for geometric criteria in resources and reserves classification. Centre for Computational Geostatistics Annual Report Eight, University of Alberta.

Joint Ore Reserves Committee (JORC), 2004. Australasian Code for Reporting of Exploration Results, Mineral Resources and Ore Reserves (JORC Code), http://www.jorc.org/pdf/jorc2004web.pdf, accessed on June 21, 2010.

Journel AG and Huijbregts CJ, 1978. Mining Geostatistics. The Blackburn Press. 600p.

South African Mineral Resource Committee (SAMREC) within the South African Institute of Mining and Metallurgy, 2000. South African Code for Reporting of Mineral Resources and Mineral Reserves, http://www.geolsoc.org.uk/webdav/site/GSL/shared/pdfs/Fellowship/South%20Africa%20Code.pdf, accessed on June 21, 2010.

United States Securities and Exchange Commission (SEC), Industry Guide 7, http://www.sec.gov/about/forms/industryguides.pdf, accessed on June 21, 2010.

Figures

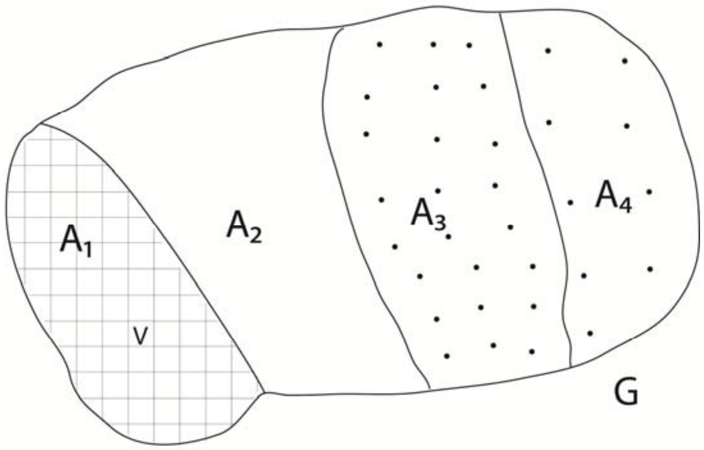

Figure 1: Illustration of the domains for which an acceptable level of uncertainty applies.

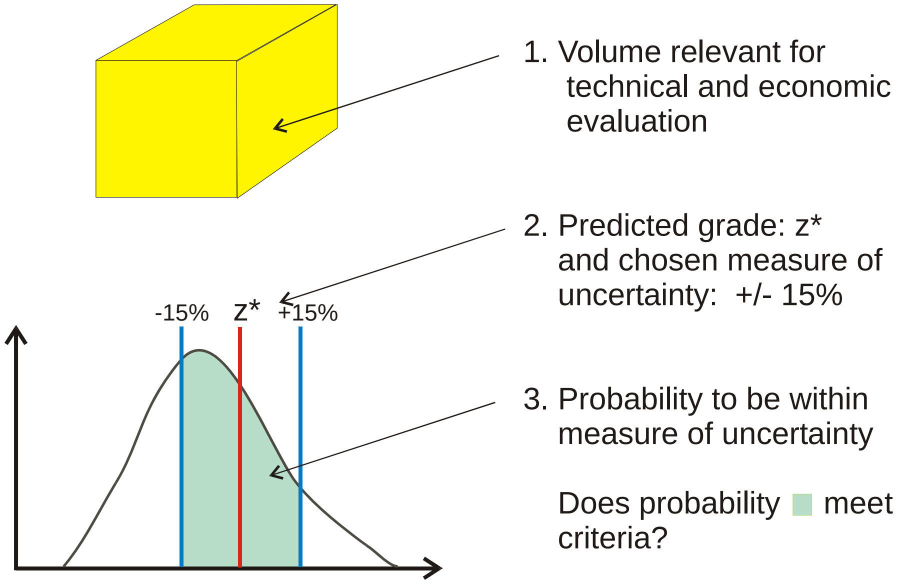

Figure 2: Illustration of the three parameters often used to specify uncertainty.

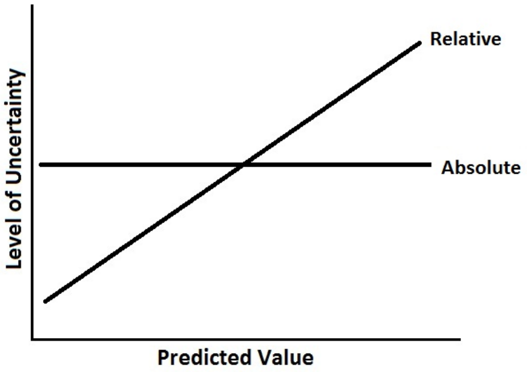

Figure 3: The level of allowable uncertainty for an absolute measure is constant while for a relative measure the level depends on the magnitude of the predicted value.

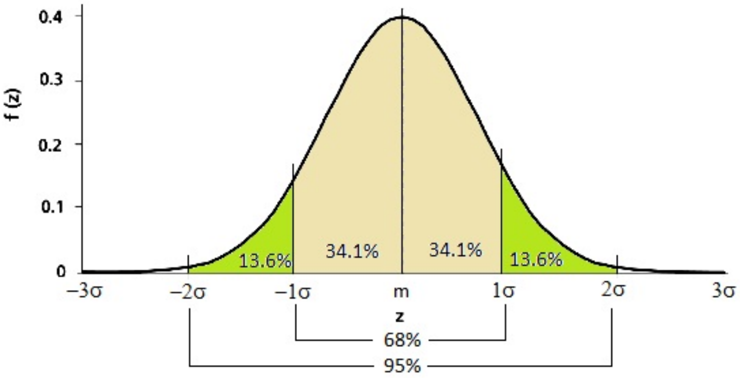

Figure 4: Illustration of probabilities to be within multiples of the standard deviation for the normal distribution.