

Effect of Size-Biased Sampling on Resource Predictions from the Three-Part Method for Quantitative Mineral Resource Assessment—A Case Study of the Gold Mines in the Timmins-Kirkland Lake Area of the Abitibi Greenstone Belt, Canada

Scientific Investigations Report 2018–5149

Effect of Size-Biased Sampling on Resource Predictions from the Three-Part Method for Quantitative Mineral Resource Assessment—A Case Study of the Gold Mines in the Timmins-Kirkland Lake Area of the Abitibi Greenstone Belt, Canada

By Karl J. Ellefsen

Scientific Investigations Report 2018–5149

**U.S. Department of the Interior
U.S. Geological Survey**

U.S. Department of the Interior
DAVID BERNHARDT, Acting Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2019

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <https://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <https://store.usgs.gov>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Ellefsen, K.J., 2019, Effect of size-biased sampling on resource predictions from the three-part method for quantitative mineral resource assessment—A case study of the gold mines in the Timmins-Kirkland Lake area of the Abitibi greenstone belt, Canada: U.S. Geological Scientific Investigations Report 2018–5149, 15 p., <https://doi.org/10.3133/sir20185149>.

ISSN 2328-0328 (online)

Acknowledgments

C.J. Hodgson, formerly of Queens University, answered questions about the data and suggested several publications with information about gold deposits. E. van Hees from the Ontario Geological Survey answered numerous questions about gold mines in the Timmins-Kirkland Lake area. Much of the initial work was performed during nights and weekends, taking time away from my family; consequently, my family deserves special thanks.

E.A. du Bray, P. Emsbo, W. Farmer, R.J. Goldfarb, K. Ryberg, B.S. Van Gosen, J. Yee, and M. Zientek reviewed the manuscript; their suggestions greatly improved it. M. Goldman helped construct figure 1. M.W. Bultman, M.E. Gettings, D.L. Leach, and W. Hamilton were among the first, if not the first, to identify problems with the three-part method for quantitative mineral resource assessment. Work during normal office hours was funded by the Mineral Resources Program of the U.S. Geological Survey.

Contents

Acknowledgments	iii
Abstract	1
Introduction.....	1
Background.....	2
Study Area.....	2
Dataset.....	3
Statistical Modeling.....	5
Three-Part Method	5
Evaluation of Predictions from the Three-Part Method	6
Discussion.....	9
Future Research.....	10
Software and Reproducibility	10
References Cited.....	11
Appendix 1. Data Compilation	13
Appendix 2. Sample Space	14
Appendix 3. Requirements for Datasets	15

Figures

1. Maps showing, the location of the Timmins-Kirkland Lake area within the Abitibi greenstone belt and, the locations of gold mines, major fault systems, and major mining camps within the Timmins-Kirkland Lake area, Canada	2
2. Graphs showing total produced ore, total produced gold and, gold grade as functions of the first year of production	3
3. Graphs showing total produced ore, total produced gold and, gold grade as functions of production order	4
4. Graphs showing results of the statistical modeling. Straight line fit to the total produced gold that was transformed with the common logarithm. Distribution showing the uncertainty in the slope of the straight line.....	5
5. Graph showing how the total produced gold is partitioned between existing and future mines in year 1915.....	6
6. Dot plots showing the distribution of the total produced gold for, the existing mines and, the future mines. The delineation between existing and future mines is shown in figure 5.....	7
7. Graphs showing, ratio of the means for the total produced gold from existing and future mines and, ratio of the medians for the total produced gold from existing and future mines	7
8. Graph showing the log-normal probability density function (that is fit to the total produced gold from 10 existing mines) and the known total produced gold for the 52 future mines.....	8
9. Graphs showing, the ratio of the means for the total produced gold from existing and future mines and, the ratios of the medians for the total produced gold from existing and future mines. The means and medians for the existing mines are calculated from the four probability density functions that are implemented in the software for the three-part method.....	9
10. Graphs showing results of the statistical modeling.....	10

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Mass	
gram (g)	0.03527	ounce, avoirdupois (oz)
kilogram (kg)	2.205	pound avoirdupois (lb)
metric ton (t)	1.102	ton, short [2,000 lb]
metric ton (t)	0.9842	ton, long [2,240 lb]

Effect of Size-Biased Sampling on Resource Predictions from the Three-Part Method for Quantitative Mineral Resource Assessment—A Case Study of the Gold Mines in the Timmins-Kirkland Lake Area of the Abitibi Greenstone Belt, Canada

By Karl J. Ellefsen

Abstract

The three-part method for quantitative mineral resource assessment is used by the U.S. Geological Survey to predict, within a specified assessment area, the number of undiscovered mineral deposits and the quantity of mineral resources in those undiscovered deposits. The effects of size-biased sampling on such predictions are evaluated in a case study that involves gold mines from the Timmins-Kirkland Lake area of the Abitibi greenstone belt, Canada. The gold mines are divided, based upon the time of the assessment, into two groups: existing mines and future mines. The total produced gold for the existing mines are used to predict, with the three-part method, the total produced gold for the future mines. Then the predictions are compared to the known, total produced gold for the future mines. For comparisons using the mean, the predictions are 1.6 to 12 times too high, depending upon the time of the assessment and the probability density function characterizing the total produced gold in the existing mines. For comparisons using the median, the predictions are 1.3 to 10 times too high, depending upon the time of the assessment. The reason for these excessively high predictions is that the three-part method is based on the assumption that the total produced gold from the existing mines is representative of the total produced gold in the future mines; this assumption is inappropriate because of size-biased sampling. There is reason to be concerned that size-biased sampling adversely affected the resource predictions of previous U.S. Geological Survey assessments that were conducted with the three-part method.

Introduction

A quantitative mineral resource assessment is a prediction, for a specified geographic area, of the number of undiscovered mineral deposits in that area and the amount of resources in those deposits. These assessments are important because, for example, land management agencies use these predictions, along with other information, to plan for appropriate land use. Quantitative mineral resource assessments have been conducted by the U.S. Geological Survey at least since 1986 (Drew and others, 1986; also see the publications listed at <https://minerals.usgs.gov/global/index.html#publications>);

almost all these assessments were performed using the three-part method, which is described in Singer (1993a) and Singer and Menzie (2010).

There is evidence that, in a specific geographic region, the amount of mineral resources in sequentially discovered deposits tends to decrease with time (Singer and Mosier 1981; Chung and others, 1992; Stanley, 1992; Long, 1995; Singer and Menzie, 2010, p. 98–101); that is, in a specific geographic region, mineral exploration companies initially tend to discover deposits with relatively large amounts of resources. With continued mineral exploration, the resource amounts of subsequently discovered deposits tend to decrease. In terms of statistical terminology, exploration may be represented by statistical sampling, and the tendency to discover specific deposits as a function of their resource amounts is a particular type of statistical sampling called “size-biased sampling” (Zimmerman, 2006). Such sampling also occurs, for example, in oil exploration (Kaufman and others, 1975; Schuenemeyer and Drew, 1983; and Long, 1988), manufacturing (Cox, 1969), forestry (Gove, 2003), and wildlife management (Patil and Rao, 1978).

Size-biased sampling is not taken into account by the three-part method, so it may adversely affect its resource predictions. Although several researchers have acknowledged this potential problem (Bultman and others, 1993; Harris and Rieber, 1993, p. 38, 144–146, 150, 304–309, 508; Bultman and Gettings, 1996), there are apparently no published investigations of it. Thus, these effects are investigated by conducting a case study involving gold mines in the Timmins-Kirkland Lake area of the Abitibi greenstone belt in Ontario, Canada. This region has been explored extensively for gold for more than a century, so there is a large amount of information about the gold mines. Indeed, the production history of the gold mines exhibits evidence of size-biased sampling (Stanley, 1992).

This report has five major sections. Section “Background” briefly reviews the geology of the greater Timmins-Kirkland Lake area, describes the data from the study area, and presents a statistical model for the gold data. Section “Three-Part Method” summarizes the three-part method for quantitative mineral resource assessments. Section “Evaluation of Predictions from the Three-Part Method” presents an evaluation of resource predictions and the reason for the inaccuracy of those predictions. Section “Discussion” explains the implications of these findings and other associated topics. Section “Future Research” presents two questions that could set a direction for future research on mineral resource assessments.

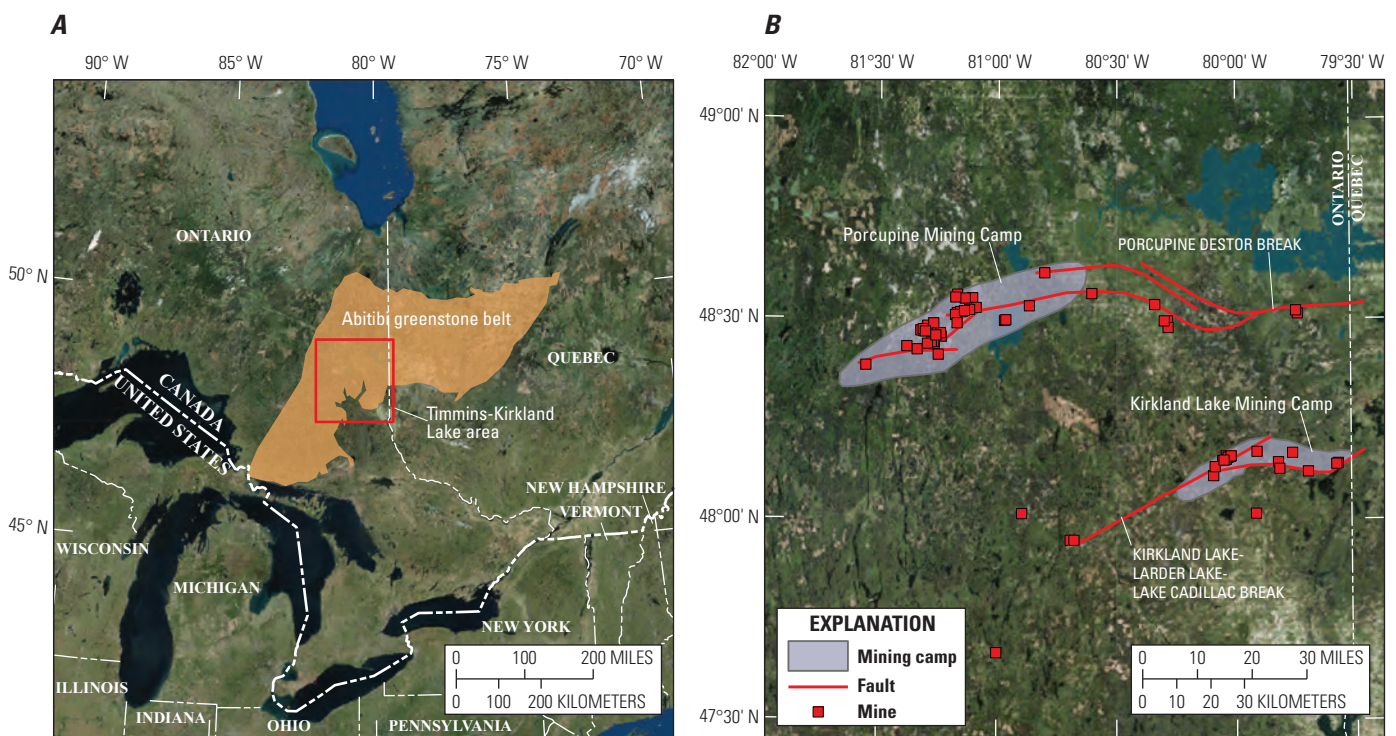
Background

Study Area

The Abitibi greenstone belt is in the Ontario and Quebec Provinces of Canada (fig. 1A). The southern part of this geologic region contains copper-zinc volcanogenic massive-sulfide deposits, orogenic gold deposits, and magmatic nickel-copper-platinum group deposits (Ayer and others, 2008). As of 2005, the total mineral production was valued at approximately 120 billion U.S. dollars (Thurston and others, 2008). Information about the geology of the gold-bearing, southern part of the Neoproterozoic Abitibi greenstone belt is summarized in Card (1990), Jackson and others (1994), Ayer and others (2002), and Monecke and others (2017).

This case study focuses on the Timmins-Kirkland Lake area, which is within the Neoproterozoic rocks of the Abitibi greenstone belt (fig. 1A). This area was chosen because it has been extensively explored and developed for more than a century; consequently, there is abundant information about the gold mines. This information has been systematically organized by the Ministry of Energy, Northern Development and Mines in the Province of Ontario, Canada, and the organized data are publicly available. The data, which are updated regularly, include location, mining and exploration history, geology, production statistics, reserve statistics, and resource statistics.

The gold-mineralized rocks within the Timmins-Kirkland Lake area are examples of a class of deposits that are called mesothermal gold deposits (Hodgson, 1993) or, more commonly, orogenic gold deposits (Groves and others, 1998; Goldfarb and others, 2005). In such deposits, the gold ores are adjacent to major, deep-crustal fault zones; the gold is in quartz veins and surrounding wallrocks. The wallrocks are often altered with carbonates, sulfides, and sericite. The country rock is regionally metamorphosed from low to medium grades. In the Timmins-Kirkland Lake area, the greenstone is cut by two major east-west fault systems (fig. 1B): the Porcupine-Destor break and the Kirkland Lake-Larder Lake-Lake Cadillac break. Most of the large gold deposits in the Timmins-Kirkland Lake area are along these two fault systems. The main cluster of these deposits along the Porcupine-Destor break is in the Porcupine mining camp, which is near the town of Timmins. It is the largest known Archean camp of orogenic gold deposits on Earth. One of main clusters of gold deposits along the Kirkland Lake-Larder Lake-Lake Cadillac break is the Kirkland Lake mining camp, which is near the town of Kirkland Lake. Additional information about the geology and deposits in this area is provided in Thomson (1948), Thomson and others (1950), Hodgson and MacGeehan (1982), Hodgson (1983), Kerrich and Watson (1984), Burrows and others (1993), Bateman and others (2005, 2008), Dube and others (2017), and Poulsen (2017).



Base from Environmental Research Systems Institute (ESRI) World Imagery, 2019, and U.S. Geological Survey digital data, 2015, 1:1,000,000. Albers Equal-Area Conic projection. Standard parallels 20° N. and 60° N. Central meridian 80°30' W. North American Datum 1983.

Figure 1. Maps showing, A, the location of the Timmins-Kirkland Lake area within the Abitibi greenstone belt and, B, the locations of gold mines, major fault systems, and major mining camps within the Timmins-Kirkland Lake area, Canada.

Dataset

The database from the Ministry of Energy, Northern Development and Mines was downloaded on June 18, 2018, so the database is current as of that date. The database was processed to extract the data that are pertinent to this case study, using the procedure described in appendix 1. The resulting dataset is in file “ProductionAndReserveData.xlsx” that accompanies this report. The dataset consists of information on 138 mines and developed prospects in the Timmins-Kirkland Lake area.

The analysis in this case study is based on the production statistics from just the mines, which number 62. For most mines, production occurred for several years. The total quantity of ore that was extracted during those years is called the “total produced ore,” and its units are metric tons (1,000 kilograms). The total quantity of gold that was extracted from the total produced ore is called the “total produced gold,” and its units are grams. The total produced ore and the total produced gold are the key production statistics. A derivative production statistic is the grade, which is defined as the total produced

gold divided by the total produced ore, and its units are grams per metric ton. Thus, the grade is a weighted average for the entire mine. The production statistics for a mine are referenced to the first year of production.

The mines are classified as either “past producer” or “current producer.” For a past producer, production ceased before June 18, 2018 (which is the date that the database was obtained), and the production statistics represent the total as of the year that production ceased. For a current producer, production is ongoing, and the production statistics represent the total as of approximately June 18, 2018. A mine classified as a past producer could be reopened for additional mining; a mine classified as a current producer almost certainly will produce more ore. Thus, the data may be interpreted as a snapshot of the total ore and gold production as of approximately June 18, 2018. This snapshot almost certainly will change after June 18, 2018.

The total produced ore, the total produced gold, and the grade are plotted as functions of the first year of production (fig. 2). Between 1910 and 1942, 46 mines started production. Two current producers are in this first interval. Between 1943 and 1981, only

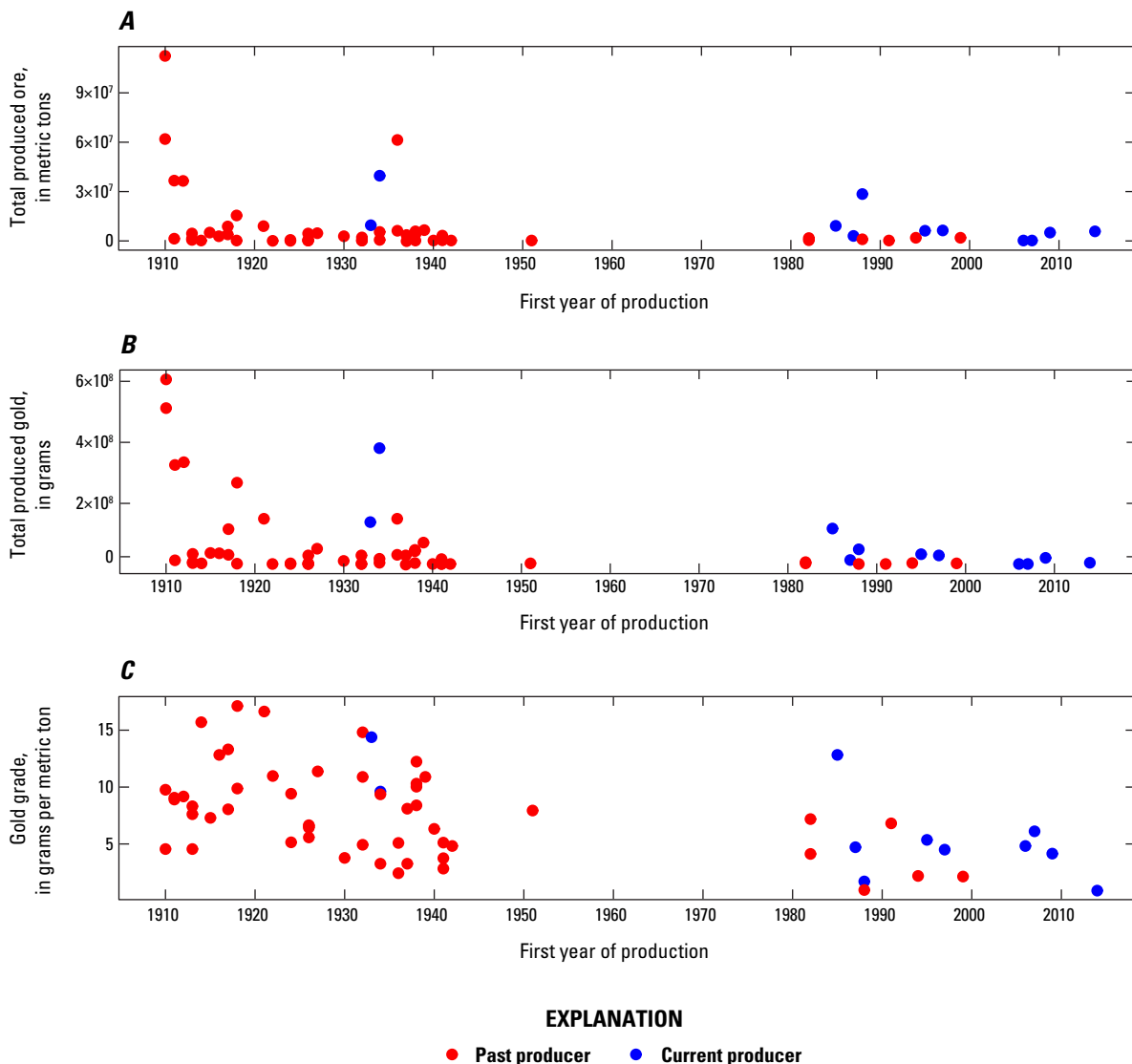


Figure 2. Graphs showing, *A*, total produced ore, *B*, total produced gold and, *C*, gold grade as functions of the first year of production.

4 Size-Biased Sampling on Resource Predictions from the Three-Part Method for Quantitative Mineral Resource Assessment

1 mine started production; this interval is associated with World War II and a period of relatively low gold prices (U.S. Geological Survey, 2014). Between 1982 and 2018, 15 mines started production. Nine current producers are in this third interval.

Compare the first and the third intervals (fig. 2). The first had 6 mines with total ore production greater than 2×10^7 metric tons, whereas the third had 1 mine. The first interval had 10 mines with total gold production greater than 1×10^8 grams, whereas the third interval had 1 mine. The first interval had 26 mines with gold grades greater than 8 grams per metric ton, whereas the third interval had 1 mine. These findings indicate that the total ore production, total gold production, and gold grade are declining.

Although the plots in figure 2 present the production statistics in the most straightforward way, there are some limitations with the plots. First, the appearance of the plot is strongly affected by economic conditions. An example is the second interval, 1943 to 1981. Second, some data points plot over other points; for example, the gold grade for a current producer in 1935 is obscured by another point. Third, the trends in the total produced ore and total produced gold are

difficult to observe because so many points are close to the horizontal axis.

To address these limitations, two changes are made to the plots. First, the total produced ore and total produced gold are plotted with a common logarithmic scale. Second, the first year of production is recast as production order (namely, the order in which the mines started production; (Chung and others, 1992)). If there are two or more mines having the same first year of production, then their production orders are assigned randomly. For example, the two mines in 1910 could be assigned orders of 1 and 2 or vice versa. Recasting the data as a function of production order diminishes the effects of economic conditions.

With these two changes, the total produced ore, the total produced gold, and the grade are replotted as functions of the production order (fig. 3). The salient feature in the replotted data is that the total produced ore, the total produced gold, and the grade have a great deal of variability. Nonetheless, all three quantities decrease as the production order increases. The decreases for total produced ore and gold are evidence of size-biased sampling.

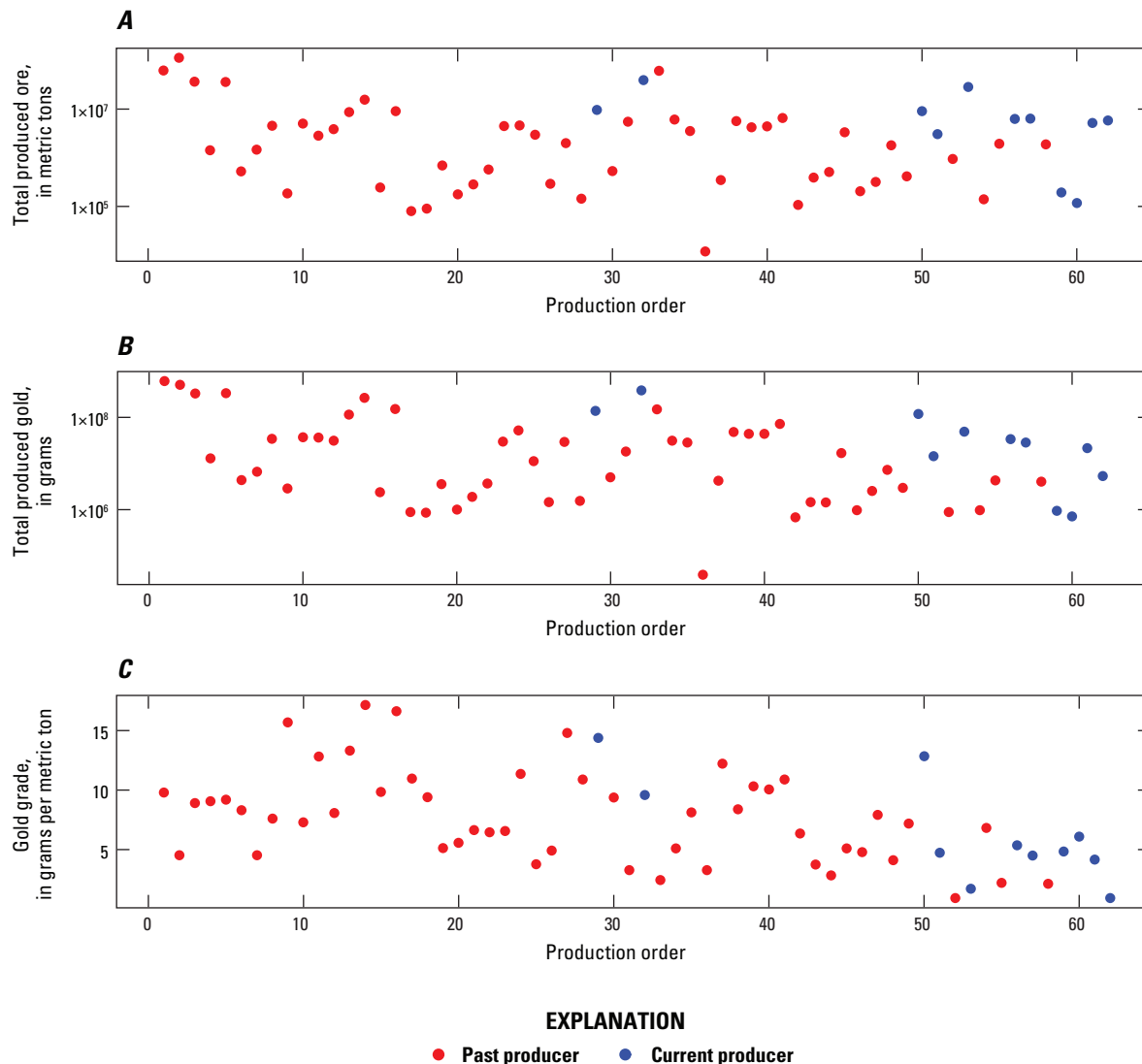


Figure 3. Graphs showing, A, total produced ore, B, total produced gold and, C, gold grade as functions of production order.

Statistical Modeling

Although the total produced ore, the total produced gold, and the gold grade decrease as the production order increases (fig. 3A, 3B, and 3C, respectively), there is significant uncertainty in these trends because of the high degree of variability in the data. To understand better how this variability affects the uncertainty in the trend, statistical modeling is performed. The statistical modeling focuses on just the total produced gold because only it is used for this case study.

The statistical modeling of the trend is performed using linear regression (DeGroot and Schervish, 2002, p. 609–636). The gold grams are transformed with the common logarithm. Although the trend may be modeled in many ways, the simplest way is just a straight line; consequently, a straight line is fit to transformed gold grams (fig. 4A). A straight line is characterized by two mathematical parameters, which are usually chosen to be the slope and the intercept with the vertical axis. The intercept is irrelevant to this analysis, so it is not discussed further. The uncertainty in the slope that is estimated by the linear regression is summarized by the distribution shown in figure 4B. The distribution, except for the far right tail, is associated with a negative slope; thus, there is a high degree of confidence that the slope is indeed negative. The implication is that, despite the high variability of the data, there is strong evidence that size-biased sampling is occurring.

Three-Part Method

The three-part method for quantitative mineral resource assessment (Singer, 1993a; Singer and Menzie, 2010, p. 10) obviously consists of three parts. In part one, the assessment geoscientists analyze geologic and related earth science data and then delineate the geographic regions, within the larger assessment area, that are likely to have mineral resources.

In part two, the number of undiscovered deposits within the regions is predicted in a probabilistic manner. To this end, the assessment geoscientists study the available earth science data from the delineated geographic regions. The data may include, for example, geologic maps, geochemical maps, and mining data. Then, based upon their experience, the assessment geoscientists guess a probability mass function for the number of undiscovered deposits—the probability mass function specifies the probability of 0 deposits, 1 deposit, 2 deposits, and so on. This probability mass function is the probabilistic prediction for the number of undiscovered deposits.

In part three, the resource quantity in each undiscovered deposit is predicted in a probabilistic manner. To this end, the assessment geoscientists compile historical mining data from discovered deposits that are the same type as those being assessed. These discovered deposits are assumed to be representative of the undiscovered deposits. (The historical mining data may be either ore quantity and average resource grade or resource quantity. Either type of data can be used for the resource assessment, and the assessment results are the same. In this case study, the resource quantity is used because it is simpler to explain and understand.) The resource quantities from many deposits are used to construct a probability density function, and it is the probabilistic prediction of the resource quantity in each undiscovered deposit.

If the probability mass function specifies that there is a 10-percent chance of two undiscovered deposits, then the resource quantity in each undiscovered deposit is characterized by the same probability density function. The important point is that, whatever the number of undiscovered deposits, the resource quantity in each undiscovered deposit is characterized by the same probability density function.

The key assumption in part three is that the discovered deposits are representative of the undiscovered deposits. The statement of this assumption is somewhat vague, so it requires further explanation (Ellefsen, 2017a). Both the discovered and undiscovered deposits constitute a finite mathematical

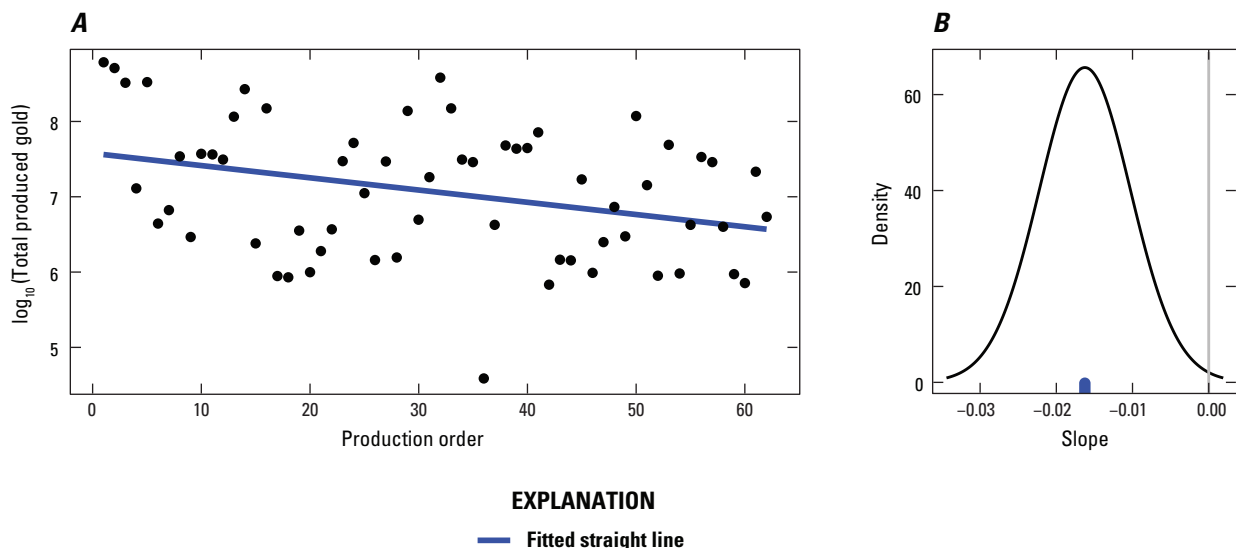


Figure 4. Graphs showing results of the statistical modeling. A, Straight line fit to the total produced gold that was transformed with the common logarithm. B, Distribution showing the uncertainty in the slope of the straight line.

set; this set is called a “sample space” and is described further in appendix 2. The discovered deposits are a simple random sample from the sample space; that is, the deposits have been sampled (discovered) without regard to any of their properties. The implication is that the properties of the discovered deposits are representative, on average, of all deposits in the sample space, including the undiscovered deposits. The important point is that the key assumption requires simple random sampling of the deposits.

Evaluation of Predictions from the Three-Part Method

The three-part method is formulated in terms of deposits, whereas the data from the Timmins-Kirkland Lake area pertain to mines. Although deposits and mines are different things, the three-part method may be applied to the mines. To this end, it is necessary to relate terms:

- A discovered deposit corresponds to an existing mine.
- An undiscovered deposit corresponds to a future mine.
- The sample space of discovered and undiscovered deposits corresponds to the sample space of existing and future mines.
- The resource quantity in a deposit corresponds to the total produced gold from a mine.

The first part in the three-part method is delineating the assessment area. For this case study, the assessment area is the geographic region shown in figure 1*B*. The gold deposits in this area are assumed to have similar features and thus group into one deposit type (namely, the orogenic gold deposit type). These gold deposits are the gold-rich quartz veins and surrounding wallrock. The existing and future mines in these deposits constitute the sample space.

Assume that the year is 1915, which is between production orders 10 and 11, and is represented by the vertical dashed line in figure 5. The data to the left of this line consist of the total produced gold from the existing mines. The data to the right of this line consist of the total produced gold from the future mines. These data on the future mines would be unknown for an actual assessment, but these data are known for this case study. This situation provides a unique opportunity to evaluate the predictions with the three-part method.

The second part in the three-part method is generating the probability mass function for the number of future mines. To keep the evaluation of the predictions as simple as possible, the probability mass function is chosen to have probability 1 for the exact number of future mines. For the example presented in figure 5, the probability mass function has probability 1 for 52 future mines. For this case study, the generation of this probability mass function is trivial, so it is not discussed again.

The third part in the three-part method is generating the probability density function for the total produced gold in the future mines. Before generating this function, the key assumption in the third part—the total produced gold from the existing mines is representative of the total produced gold from the future mines—is checked. For the example presented in figure 5, the total produced gold for the existing and future mines are presented as dot plots (fig. 6). In a dot plot, the horizontal axis is divided into intervals, and the total produced gold for a mine is represented by a dot within the appropriate interval. The dots collectively represent the distribution of the total produced gold.

The distribution for the future mines (fig. 6*B*) is shifted leftward with respect to the distribution for the existing mines (fig. 6*A*). This shift may be quantified with the ratio between the means of the two distributions. The ratios also are calculated for 11 existing mines, 12 existing mines, and so on. The last ratio is calculated for 52 existing mines because the 10 future mines are just enough to calculate a ratio with adequate precision. The ratios are plotted in figure 7*A*. Ratios are similarly calculated with the median and are plotted in figure 7*B*.

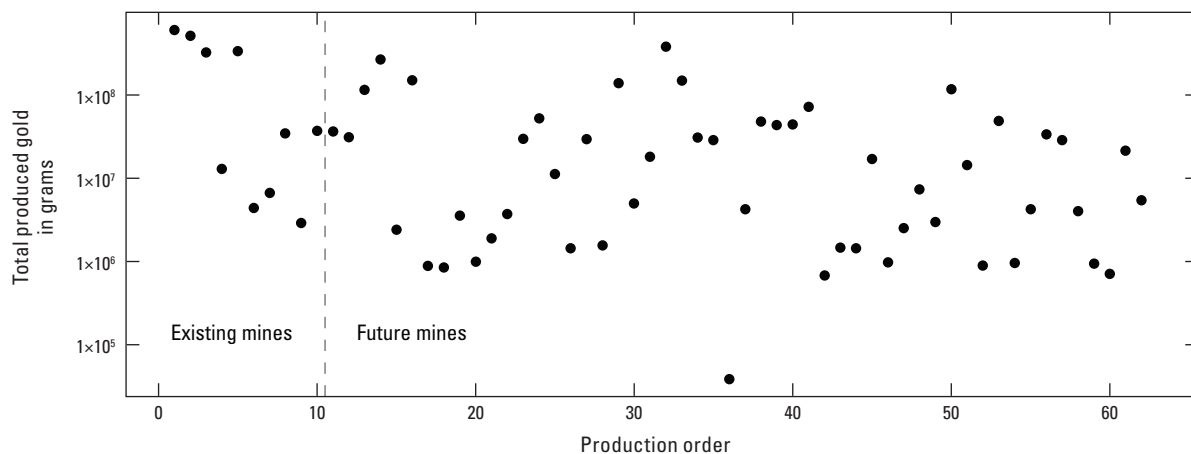


Figure 5. Graph showing how the total produced gold is partitioned between existing and future mines in year 1915.

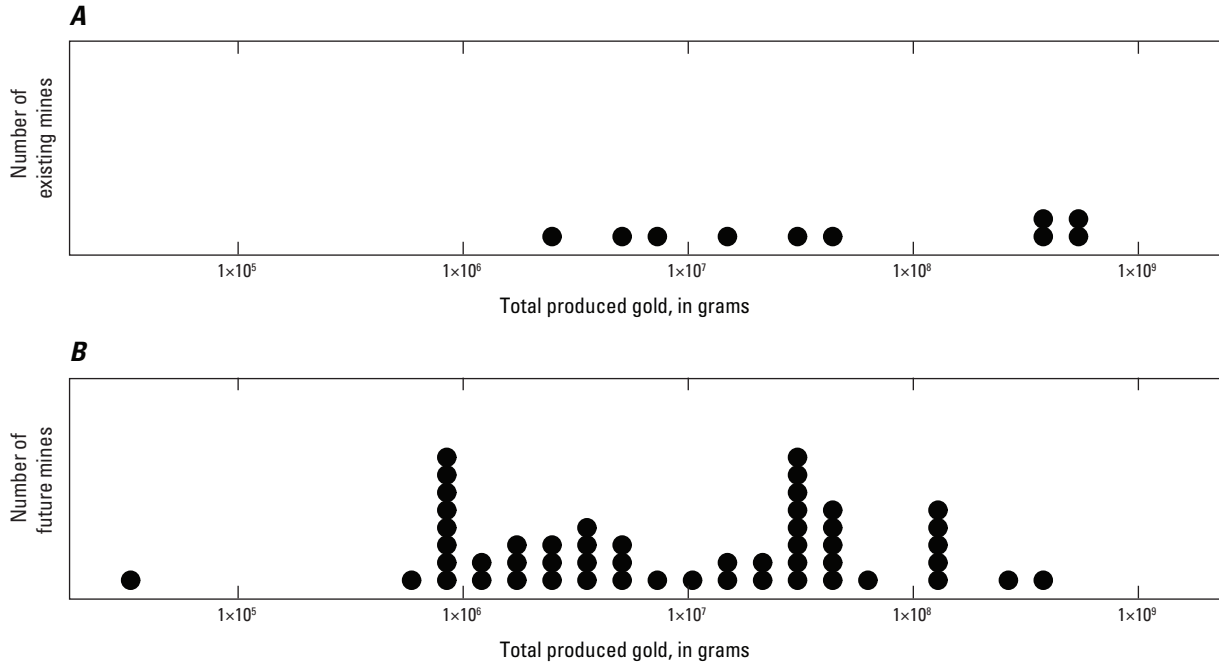
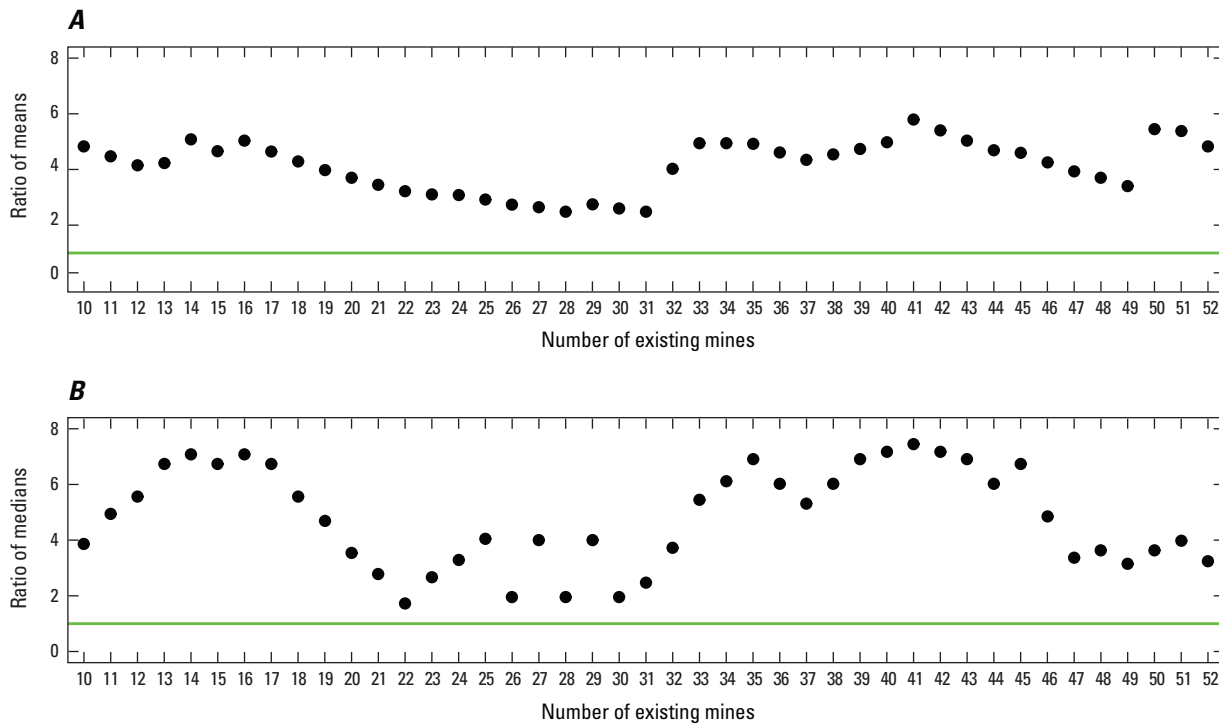


Figure 6. Dot plots showing the distribution of the total produced gold for, *A*, the existing mines and, *B*, the future mines. The delineation between existing and future mines is shown in figure 5.



EXPLANATION

— Ratio of 1

Figure 7. Graphs showing, *A*, ratio of the means for the total produced gold from existing and future mines and, *B*, ratio of the medians for the total produced gold from existing and future mines.

The ratios of the means range from 2.5 to 5.8 (fig. 7A), and the ratios of the medians range from 1.7 to 7.4 (fig. 7B). That the ratios are significantly greater than 1 indicates that the total produced gold from the existing mines is not representative of the total produced gold from the future mines—the key assumption in part three of the three-part method is inappropriate. Recall that this assumption requires simple random sampling of the mines (section “Three-Part Method”), but there is strong evidence of size-biased sampling (fig. 4). Thus, the inappropriateness of the assumption is caused by size-biased sampling.

Although the key assumption in generating the probability density function is inappropriate, the probability density function is developed nonetheless as part of the evaluation. In the current software implementation of the three-part method (Ellefsen, 2017b), there are four choices for the probability density function: log-normal, log-normal that is truncated at the lowest and highest values of the data, kernel density estimate, and kernel density estimate that is truncated at the lowest and highest values of the data. Consider the log-normal probability density function that is fit to the total produced gold for 10 existing mines (fig. 8). This log-normal probability density function looks like a normal probability density function because of the logarithmic scaling of the horizontal axis.

The probability density function (fig. 8) represents the predicted total produced gold in each of the 52 future mines that, in turn, are predicted with the probability mass function from part two. The range of this probability density function spans the known total produced gold of the 52 future mines—the total produced gold for any one of the 52 future mines could have come from this probability density function. However, there is a problem: The total produced gold for the 52 future mines, taken together, is shifted leftward with respect to the probability density function. The leftward shift may be quantified by the ratio between the mean of the probability density function and the mean of the total produced gold from the 52 existing mines. These ratios are calculated for the four different probability density functions that are implemented in the software. These

calculations are repeated for 11 existing mines, 12 existing mines, and so on. Again, the last ratio is calculated for 52 existing mines because the 10 future mines are just enough to calculate a ratio with adequate precision. The results are shown in figure 9A. This procedure is repeated using the ratio between the medians, and the results are shown in figure 9B.

For the ratios of the means (fig. 9A), the truncated probability density functions have much smaller ratios than the nontruncated probability density functions have. The reason is that the mean is strongly affected by the right tail of the probability density function. Discontinuities are caused by an especially high or low total produced gold; for example, the discontinuity between numbers 31 and 32 is caused by the very high value for production order 32 (fig. 5). The particular value for a ratio depends on the probability density function and the time of the assessment. Overall, the ratios range from 1.6 to 12 (fig. 9A). The interpretation of this finding is that the predicted total produced gold for the future mines would be 1.6 to 12 times too high, as measured by the mean.

The ratios of the medians, for a specific number of existing mines, are approximately equal (fig. 9B). The reason is that the medians for the four probability density functions are approximately equal. There are some discontinuities, but they are relatively small compared to those for the ratio of means (fig. 9A). The particular value for a ratio depends mostly on the time of the assessment. Overall, the ratios range from 1.3 to 10 (fig. 9B). The interpretation of this finding is that the predicted total produced gold for the future mines would be 1.3 to 10 times too high, as measured by the median.

If the probability density functions were representative of the total produced gold in the future mines, then the ratios of the means and medians would be approximately 1. The reason that the ratios usually are significantly greater than 1 is the inappropriateness of the key assumption for part three—the total produced gold from the existing mines is not representative of the total produced gold from the future mines. Again, this inappropriateness is caused by size-biased sampling (fig. 4).

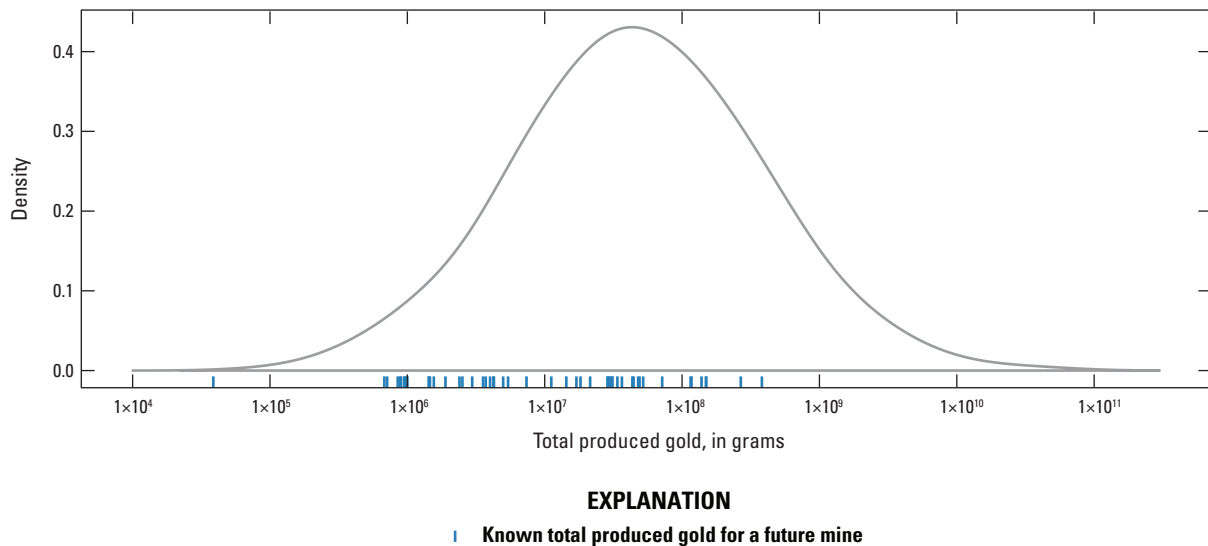


Figure 8. Graph showing the log-normal probability density function (that is fit to the total produced gold from 10 existing mines) and the known total produced gold for the 52 future mines.

Discussion

A common procedure in the three-part method is to aggregate data from sampling units that are proximate. “Sampling unit” is a generic term for mine, ore body, or mining camp (Singer, 1993b). Singer and Menzie (2010, p. 95–96) seem to present two different rationales for this procedure. First, if there is a mixture of different sampling units, then aggregating selected data will make the sampling units the same. Having the same sampling unit is crucial to accuracy of the three-part method. For this case study, the sampling unit always is a mine—there is no mixture of different sampling units. Second, the aggregated data will represent an ore body (namely, a deposit). The problem with this rationale is that the aggregated data would be an inaccurate characterization of the ore body: The aggregation is based on proximity, not on geology; the distance that defines proximate is arbitrary (Singer and Menzie, 2010, p. 96); and aggregated data from mines near the Earth’s surface would not characterize an ore body that extends well below the Earth’s surface. Therefore, based on either rationale, this procedure is inappropriate for this case study.

Another common procedure in the three-part method is to construct the probability density function using data from deposits of the same type but outside the assessment area (Cox and Singer, 1986; Singer and others, 1993). The principal problem with this

procedure is that the data from outside the assessment area usually differ from data inside the assessment area. This difference would decrease the accuracy of the prediction; consequently, this procedure is inappropriate for this case study.

To conduct a mineral resource assessment of the gold in the Timmins-Kirkland Lake area, it is important to recall that the goal is to predict the future gold resources. Obviously, these gold resources will come from mines in the Timmins-Kirkland Lake area; therefore, the best way to make a prediction is to analyze the total produced gold from existing mines in the Timmins-Kirkland Lake area. Aggregating data from proximate mines and including data from outside the Timmins-Kirkland Lake area will decrease the accuracy of the prediction.

Recall that the data from the Timmins-Kirkland Lake area are chosen to be just the total produced gold from the past and current producers (see the section “Background”). The reason for this choice is that it is the simplest of all possible choices. However, there are appropriate alternatives. Perhaps the best alternative involves adding the probable and proven reserves for the current producers to the total produced gold. The rationale for this alternative is that it would make the data for the current producers as close as possible to that of the past producers. A disadvantage of this alternative is that probable and proven reserves are available for only 7 of the 11 current producers.

Statistical modeling for the alternative dataset is performed as described in section “Statistical Modeling.” The results are

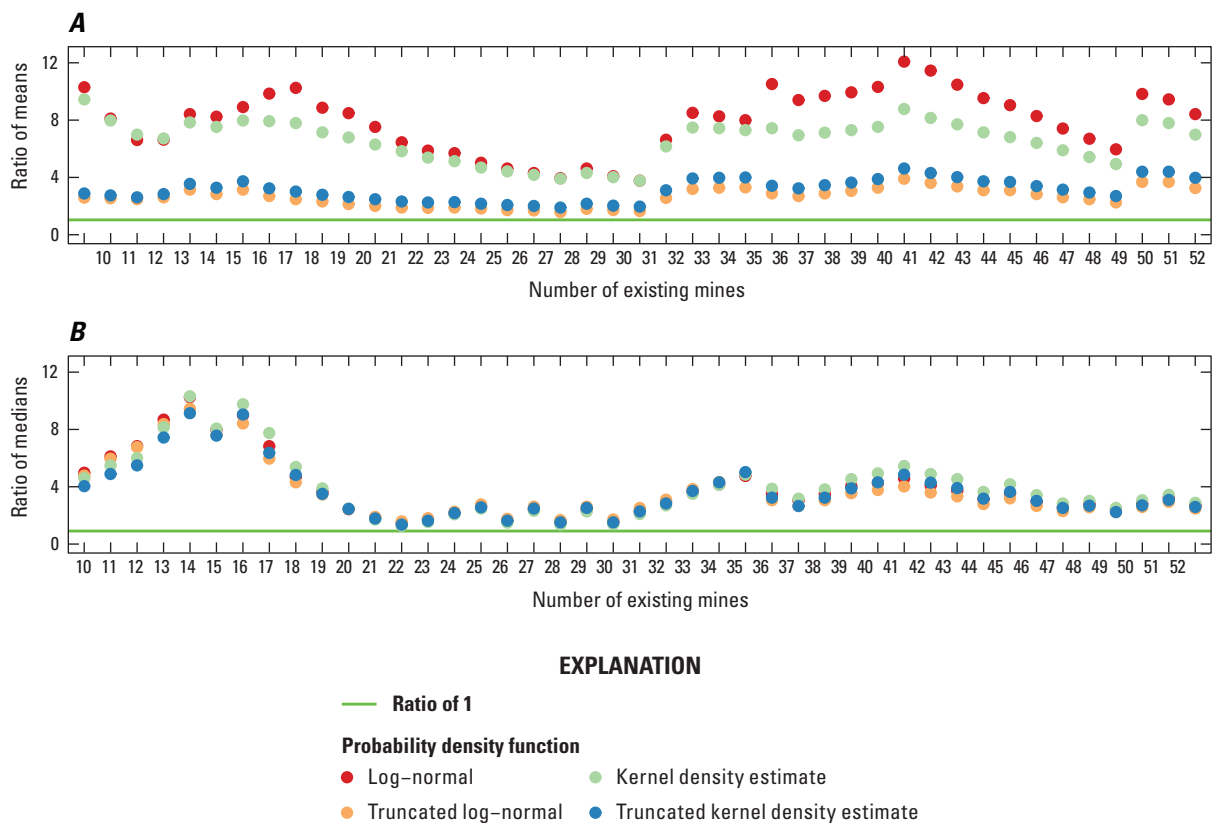


Figure 9. Graphs showing, *A*, the ratio of the means for the total produced gold from existing and future mines and, *B*, the ratios of the medians for the total produced gold from existing and future mines. The means and medians for the existing mines are calculated from the four probability density functions that are implemented in the software for the three-part method.

shown in figure 10A. The slope of the straight line that is fit to the data is negative. The distribution for the uncertainty in the slope is almost entirely along the negative part of the horizontal axis (fig. 10B), so there is a high degree of confidence that the slope is indeed negative. Thus, there is strong evidence of size-biased sampling in this alternative dataset, so the findings of the evaluation would be practically the same (see section “Evaluation of Predictions from the Three-Part Method”).

This case study pertains only to the gold mines in the Timmins-Kirkland Lake area; consequently, the findings of this case study should not be used to evaluate the accuracy of prior mineral resource assessments of other deposit types in other parts of the world. However, in four other regional datasets—mercury deposits in California (Chung and others, 1992); porphyry copper deposits in Argentina, Chile, and Peru (Long, 1995); porphyry copper deposits in Mexico and the western United States (Long, 1995); and sediment-hosted gold deposits in Nevada (Singer and Menzie, 2010, p. 98–101)—the quantity of mined resource declines with time. That is, these four datasets have evidence of size-biased sampling; thus, it is plausible that size-biased sampling is common or even ubiquitous. If so, then the effects of size-biased sampling must be incorporated into mineral resource predictions. Otherwise, the predictions will be too high. Because all previous mineral resource predictions that U.S. Geological Survey personnel made using the three-part method do not account for size-biased sampling, there is reason to be concerned that these predictions are too high.

Future Research

While conducting this case study, two questions pertinent to mineral resource prediction arose and should be considered in defining a direction for future research. To investigate these questions, several additional datasets that are like the dataset for this case study are needed. Requirements for these additional datasets are described in appendix 3.

The first research question is, “As mines are developed in an assessment area, how do their properties (for example, total produced resources and average grades) change?” Some aspects of this question have been investigated by Chung and others (1992), Stanley (1992), Long (1995), Singer and Menzie (2010, p. 98–101), and this case study. However, these investigations are limited to orogenic gold deposits, sediment-hosted gold deposits, mercury deposits, and porphyry copper deposits. Thorough investigations of other deposit types would provide a more comprehensive answer to the question.

The second research question is, “Can the properties of undiscovered mineral resources be predicted?” This question is pertinent because undiscovered mineral resources are important to the future world economy; however, significant challenges complicate prediction. For example, mineral resource datasets typically are small (that is, less than 50 significant deposits of any particular type), the datasets are incomplete (that is, ore tonnages and grades for many deposits are unavailable), and many deposit types have highly variable ore tonnage and resource grades. If it seems that properties of undiscovered mineral resource can be predicted, then a robust prediction method should be developed and thoroughly tested.

Software and Reproducibility

The data that are used for this case study are publicly available. These data were prepared as described in appendix 1, and the prepared data are stored in compressed file “ReportScripts.zip,” which accompanies this report. The calculations and the figures are generated with R-language scripts, which also are stored in compressed file “ReportScripts.zip.” Readers are encouraged to execute these scripts to check the calculations and figures in this report. Please report any errors.

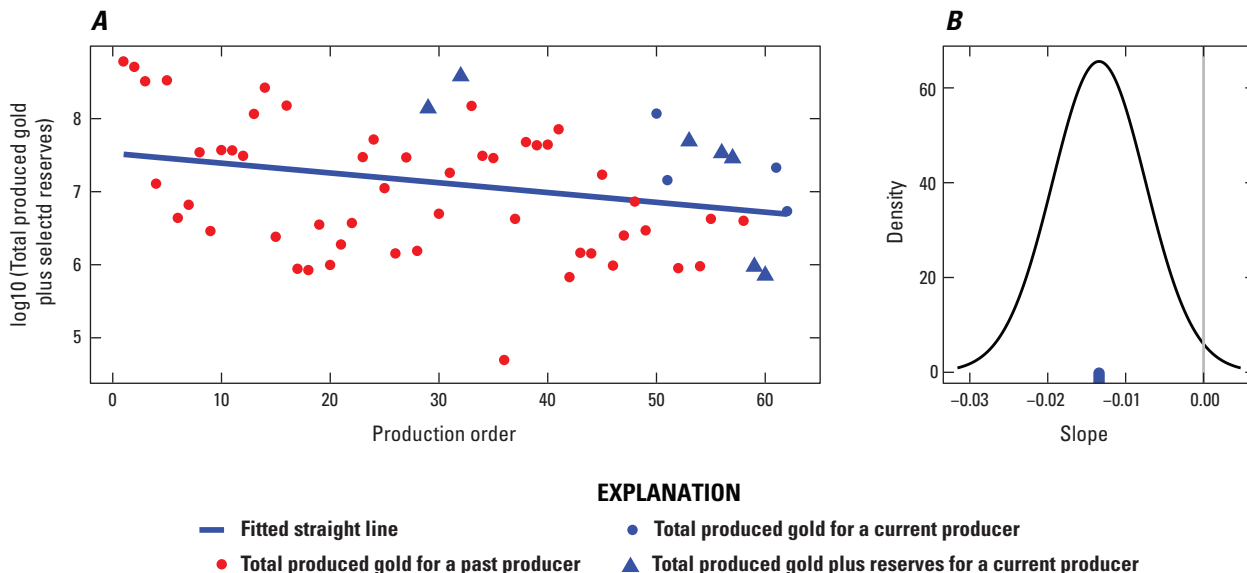


Figure 10. Graphs showing results of the statistical modeling. A, Straight line fit to the total produced gold plus selected reserves that were transformed with the common logarithm. B, Distribution showing the uncertainty in the slope of the straight line.

References Cited

- Ayer, J., Amelin, Y., Corfu, F., Kamo, S., Ketchum, J., Kwok, K., and Trowell, N., 2002, Evolution of the southern Abitibi greenstone belt based on U-Pb geochronology—Autochthonous volcanic construction followed by plutonism, regional deformation and sedimentation: *Precambrian Research*, v. 115, nos. 1–4, p. 63–95. [Also available at [https://doi.org/10.1016/S0301-9268\(02\)00006-2](https://doi.org/10.1016/S0301-9268(02)00006-2).]
- Ayer, J.A., Thurston, P.C., and Lafrance, B., 2008, A special issue devoted to base metal and gold metallogeny at regional, camp, and deposit scales in the Abitibi greenstone belt—Preface: *Economic Geology and the Bulletin of the Society of Economic Geologists*, v. 103, no. 6, p. 1091–1096. [Also available at <https://doi.org/10.2113/gsecongeo.103.6.1091>.]
- Bateman, R., Ayer, J.A., and Dubé, B., 2008, The Timmins-Porcupine gold camp, Ontario—Anatomy of an Archean greenstone belt and ontogeny of gold mineralization: *Economic Geology*, v. 103, no. 6, p. 1285–1308. [Also available at <https://doi.org/10.2113/gsecongeo.103.6.1285>.]
- Bateman, R., Ayer, J.A., Dubé, B., and Hamilton, M.A., 2005, The Timmins-Porcupine gold camp, northern Ontario—The anatomy of an Archaean greenstone belt and its gold mineralization—Discover Abitibi initiative: Ontario Geological Survey, Open File Report 6158, 90 p., accessed August 31, 2018, at <http://www.geologyontario.mndm.gov.on.ca/mndmfiles/pub/data/imaging/OFR6158/ofr6158.pdf>.
- Bultman, M.W., Force, E.R., Gettings, M.E., and Fisher, F.S., 1993, Comments on the “three-step” method for quantification of undiscovered mineral resources: U.S. Geological Survey, Open-File Report 9323, 59 p., accessed August 31, 2018, at <https://doi.org/10.3133/ofr9323>.
- Bultman, M.W., and Gettings, M.E., 1996, Quantitative mineral resource assessment of Coronado National Forest in du Bray, E.A., ed., *Mineral resource potential and geology of Coronado National Forest, southeastern Arizona and southwestern New Mexico*, U.S. Geological Survey Bulletin 2083 A–K, p. 183–202, accessed August 31, 2018, at <https://doi.org/10.3133/b2083AK>.
- Burrows, D.R., Spooner, E.T.C., Wood, P.C., and Jemielita, R.A., 1993, Structural controls on formation of the Hollinger-McIntyre Au quartz vein system in the Hollinger shear zone, Timmins, southern Abitibi greenstone belt, Ontario: *Economic Geology*, v. 88, no. 6, p. 1643–1663. [Also available at <https://doi.org/10.2113/gsecongeo.88.6.1643>.]
- Card, K.D., 1990, A review of the Superior Province of the Canadian Shield, a product of Archean accretion: *Precambrian Research*, v. 48, p. 99–156. [Also available at [https://doi.org/10.1016/0301-9268\(90\)90059-Y](https://doi.org/10.1016/0301-9268(90)90059-Y).]
- Chung, C.F., Singer, D.A., and Menzie, W.D., 1992, Predicting sizes of undiscovered deposits—An example using mercury deposits in California: *Economic Geology*, v. 87, no. 4, p. 1174–1179. [Also available at <https://doi.org/10.2113/gsecongeo.87.4.1174>.]
- Cox, D.P., and Singer, D.A., 1986, *Mineral deposit models*: U.S. Geological Survey, Bulletin 1693, 379 p., accessed August 31, 2018, at <https://doi.org/10.3133/b1693>.
- Cox, D.R., 1969, Some sampling problems in technology, in Johnson, N.L., and Smith, H., Jr., eds., *New developments in survey sampling*: New York, Wiley-Interscience, p. 506–527.
- DeGroot, M.H., and Schervish, M.J., 2002, *Probability and statistics*: Boston, Addison-Wesley, 816 p.
- Drew, L.J., Bliss, J.D., Bowen, R.W., Bridges, N.J., Cox, D.P., DeYoung, J.H., Houghton, J.C., Ludington, S.D., Menzie, W.D., Page, N.J., Root, D.H., and Singer, D.A., 1986, Quantitative estimation of undiscovered mineral resources—A case study of U.S. Forest Service wilderness tracts in the Pacific mountain system: *Economic Geology*, v. 81, no. 1, p. 80–88. [Also available at <https://doi.org/10.2113/gsecongeo.81.1.80>.]
- Dube, B., Mercier-Langevin, P., Ayer, J.A., Atkinson, B., and Monecke, T., 2017, Orogenic greenstone-hosted quartz-carbonate gold deposits of the Timmins-Porcupine camp, in Monecke, T., Mercier-Langevin, P., and Dube, B., eds., *Archean base and precious metal deposits, southern Abitibi greenstone belt, Canada*: Society of Economic Geologists, *Reviews in Economic Geology*, v. 19, p. 51–80.
- Ellefsen, K.J., 2017a, Probability calculations for three-part mineral resource assessments: U.S. Geological Survey *Techniques and Methods*, book 7, chap. C15, 14 p., accessed August 31, 2018, at <https://doi.org/10.3133/tm7C15>.
- Ellefsen, K.J., 2017b, User’s guide for MapMark4—An R package for the probability calculations in three-part mineral resource assessments: U.S. Geological Survey *Techniques and Methods*, book 7, chap. C14, 23 p., accessed August 31, 2018, <https://doi.org/10.3133/tm7C14>.
- Goldfarb, R.J., Baker, T., Dubé, B., Groves, D.I., Hart, C.J.R., and Gosselin, P., 2005, Distribution, character and genesis of gold deposits in metamorphic terranes, in Hedenquist, J.W., Thompson, J.F.H., Goldfarb, R.J., and Richards, J.P., eds., *One hundredth anniversary volume*: Littleton, Colo., Society of Economic Geologists, p. 407–450.
- Gove, J.H., 2003, Estimation and applications of size-biased distributions in forestry, in Amaro, A., Reed, D., and Soares, P., eds., *Modeling forest systems—Location unknown*: CAB International, p. 201–211, accessed August 31, 2018, at https://www.fs.fed.us/ne/newtown_square/publications/other_publishers/OCR/ne_2003gove01.pdf.
- Groves, D.I., Goldfarb, R.J., Gebre-Mariam, M., Hagemann, S.G., and Robert, F., 1998, Orogenic gold deposits—A proposed classification in the context of their crustal distribution and relationship to other gold deposit types: *Ore Geology Reviews*, v. 13, nos. 1–5, p. 7–27. [Also available at [https://doi.org/10.1016/S0169-1368\(97\)00012-7](https://doi.org/10.1016/S0169-1368(97)00012-7).]
- Harris, D.P., and Rieber, M., 1993, Evaluation of the United States Geological Survey’s three-step assessment methodology: U.S. Geological Survey Open-File Report 93–258-A, accessed August 31, 2018, at <https://doi.org/10.3133/ofr93258A>.

- Hodgson, C.J., 1983, The place of gold ore formation in the geologic evolution of the Abitibi greenstone belt, Ontario: *Transactions of the Institute of Mining and Metallurgy*, v. 95, sec. B, p. B183–B194.
- Hodgson, C.J., 1993, Mesothermal lode-gold deposits, *in* Kirkham, R.V., Sinclair, W.D., Thorpe, R.I., and Duke, J.M., eds., *Mineral deposit modeling: Geological Association of Canada, Special Paper 40*, p. 635–678.
- Hodgson, C.J., and MacGeehan, P.J., 1982, A review of the geological characteristics of gold-only deposits in the Superior Province of the Canadian Shield, *in* *Geology of the Canadian gold deposits—Proceedings of the Canadian Institute of Mining and Metallurgy Gold Symposium: Canadian Institute of Mining and Metallurgy, Special Volume*, v. SV 24, no. 1982, p. 211–229.
- Jackson, S.L., Fyon, J.A., and Corfu, F., 1994, Review of Archean supracrustal assemblages of the southern Abitibi greenstone belt in Ontario, Canada—Products of microplate interactions within a large-scale plate-tectonic setting: *Pre-cambrian Research*, v. 65, nos. 1–4, p. 183–205. [Also available at [https://doi.org/10.1016/0301-9268\(94\)90105-8](https://doi.org/10.1016/0301-9268(94)90105-8).]
- Kaufman, G.M., Balcer, Y., and Kruyt, D., 1975, A probabilistic model of oil and gas discovery, *in* Haun, J.D., ed., *Methods of estimating the volume of undiscovered oil and gas resources: American Association of Petroleum Geologists, Studies in Geology*, no. 1 p. 113–142.
- Kerrick, R., and Watson, G.P., 1984, The Macassa mine Archean lode gold deposit, Kirkland Lake, Ontario—Geology, patterns of alteration, and hydrothermal regimes: *Economic Geology*, v. 79, no. 5, p. 1104–1130. [Also available at <https://doi.org/10.2113/gsecongeo.79.5.1104>.]
- Long, K.R., 1988, Estimating the number and the sizes of undiscovered oil and gas pools: Tucson, Ariz., University of Arizona, Ph.D. dissertation, 160 p., accessed August 31, 2018, at <http://hdl.handle.net/10150/184516>.
- Long, K.R., 1995, Production and reserves of Cordilleran (Alaska to Chile) porphyry copper deposits, *in* Pierce, F.W., and Bolm, J.G., eds., *Porphyry copper deposits of the American Cordillera: Arizona Geological Society Digest 20*, p. 35–68.
- Monecke, T., Mercier-Langevin, P., Dube, B., and Friedman, M., 2017, Geology of the Abitibi greenstone belt, *in* Monecke, T., Mercier-Langevin, P., and Dube, B., eds., *Archean base and precious metal deposits, southern Abitibi greenstone belt, Canada: Society of Economic Geologists, Reviews in Economic Geology*, v. 19, p. 7–50.
- Patil, G.P., and Rao, C.R., 1978, Weighted distributions and size-biased sampling with applications to wildlife populations and human families: *Biometrics*, v. 34, no. 2, p. 179–189. [Also available at <https://doi.org/10.2307/2530008>.]
- Poulsen, K.H., 2017, The Larder Lake-Cadillac break and its gold deposits, *in* Monecke, T., Mercier-Langevin, P., and Dube, B., eds., *Archean base and precious metal deposits, southern Abitibi greenstone belt, Canada: Society of Economic Geologists, Reviews in Economic Geology*, v. 19, p. 133–168.
- Schuenemeyer, J.H., and Drew, L.J., 1983, A procedure to estimate the parent population of the size of oil and gas fields as revealed by a study of economic truncation: *Mathematical Geology*, v. 15, no. 1, p. 145–161. [Also available at <https://doi.org/10.1007/BF01030080>.]
- Singer, D.A., 1993a, Basic concepts in three-part quantitative assessments of undiscovered mineral resources: Nonrenewable resources, v. 2, no. 2, p. 69–81. [Also available at <https://doi.org/10.1007/BF02272804>.]
- Singer, D.A., 1993b, Development of grade and tonnage models for different deposit types, *in* Kirkham, R.V., Sinclair, W.D., Thorpe, R.I., and Duke, J.M., eds., *Mineral deposit modeling: Geological Association of Canada, Special Paper 40*, p. 21–30.
- Singer, D.A., and Menzie, W.D., 2010, *Quantitative mineral resource assessments—An integrated approach*: New York, Oxford University Press, 219 p.
- Singer, D.A., and Mosier, D.L., 1981, The relation between exploration economics and the characteristics of mineral deposits, *in* Ramsey, J.B., ed., *The economics of exploration for energy resources*: Greenwich, Conn., JAI Press, p. 313–326.
- Singer, D.A., Mosier, D.L., and Menzie, D.W., 1993, Digital grade and tonnage data for 50 types of mineral deposits: U.S. Geological Survey Open-File Report 930280, 107 p., accessed August 31, 2018, at <https://doi.org/10.3133/ofr93280>.
- Stanley, M.C., 1992, Statistical trends and discoverability modeling of gold deposits in the Abitibi greenstone belt, Ontario *in* Kim, Y.C., ed., *Proceedings of the 23rd International Symposium of Applications of Computers and Operations Research in the Mineral Industry*, April 7–11, 1992, Tucson, Ariz.: New York, Society for Mining, Metallurgy, and Exploration/American Institute of Mining, Metallurgical, and Petroleum Engineers, p. 11–17.
- Thomson, J.E., 1948, Regional structure of the Kirkland Lake-Larder Lake Area, *in* *Structural geology of Canadian ore deposits: Canadian Institute of Mining and Metallurgy, Geology Division*, p. 627–632.
- Thomson, J.E., Charlewood, G.H., Griffin, K., Hawley, J.E., Hopkins, H., MacIntosh, C.G., Orgizio, S.P., Perry, O.S., and Ward, W., 1950, *Geology of the main ore zone at Kirkland Lake: Ontario Department of Mines, Annual Report*, v. 57, pt. 5, p. 54–196.
- Thurston, P.C., Ayer, J.A., Goutier, J., and Hamilton, M.A., 2008, Depositional gaps in Abitibi greenstone belt stratigraphy—A key to exploration for syngenetic mineralization: *Economic Geology*, v. 103, no. 6, p. 1097–1134. [Also available at <https://doi.org/10.2113/gsecongeo.103.6.1097>.]
- U.S. Geological Survey, 2014, Gold statistics, *in* Kelly, T.D., and Matos, G.R., comps., *Historical statistics for mineral and material commodities in the United States: U.S. Geological Survey Data Series 140*, accessed August 2018 at <http://minerals.usgs.gov/minerals/pubs/historical-statistics/>.
- Zimmerman, D.L., 2006, Size-biased sampling, *in* *Encyclopedia of environmetrics*: John Wiley & Sons, Ltd, accessed July 2017 at <https://doi.org/10.1002/9780470057339.vas027.pub2>.

Appendix 1. Data Compilation

The Ministry of Energy, Northern Development and Mines for the Province of Ontario, Canada, maintains a database of mines, prospects, and occurrences within the province. This database is in the public domain (<https://www.mndm.gov.on.ca/en/mines-and-minerals/applications/ogsearth/mineral-deposits-mdi>) and was downloaded on June 18, 2018.

The database has information on many different mines, prospects, and occurrences throughout the province, so it is necessary to find and extract the information that is pertinent to this case study. To this end, the database was opened with computer program ArcGIS Pro, and the database was searched to find those mines and prospects that met the following criteria:

1. The mines and prospects are within the Timmins-Kirkland Lake area. This area is rectangular, and the coordinates of the rectangle are specified by the Universal Transverse Mercator (UTM) projection in zone 17. The easting coordinates for the rectangle are 416998 and 616792 meters, and the northing coordinates for the rectangle are 5250203 and 5439805 meters.
2. The mines and prospects are classified, within the database, as “developed prospect with reserves,” “past producing mine with reserves,” “past producing mine without reserves,” or “producing mine.” (These four classifications are defined in the report “Mineral Deposit Category Definitions,” which is in file “MDI Definitions and cut-off values.pdf” that accompanies the database.) The reason for this criterion is that these mines and prospects have production data, which are needed for this case study.
3. The primary commodity in the mines and prospects is gold.

That part of the database that met the three criteria comprises information on 138 developed prospects and mines. The information, for each developed prospect and mine, includes an internet link to a report that is published by the Ministry of Energy, Northern Development and Mines and is in the public domain. Each report, which is henceforth called the Mineral Data Inventory (MDI) data sheet, includes location, mining and exploration history, geology, production data, reserve data, and references.

The information that is needed for this case study is compiled from the part of the database that met the 3 criteria and the 138 MDI data sheets. The compiled information is in file “ProductionAndReserveData.xlsx” on spreadsheet “Resource Data.” There are 138 records for the 138 developed prospects and mines. Most columns in the spreadsheet are easy to understand, so only a few remarks are necessary. Column “MDI Identifier” lists the character-string identifiers that the Ministry of Energy, Northern Development and Mines uses to identify mineral deposits. Column “Comment” lists some problems that occurred during the compilation. (In file “ProductionAndReserveData.xlsx,” the other spreadsheets list ore and gold production data that are summarized on spreadsheet “Resource Data.” For example, spreadsheet “Cheminis” lists, for each year, the metric tons of ore and the grams of gold that were produced. The sums of these two quantities appear on spreadsheet “Resource Data.”)

For some developed prospects, ore was mined and processed to extract the gold. These production data are included in the compilation. A particularly difficult part of the compilation involved the reserve data. For some developed prospects and mines, reserves were estimated repeatedly; for the compilation, the most current estimated reserves are used. Sometimes, different types of reserves were estimated (for example, inferred reserves and proven reserves); for the compilation, each type is reported. Sometimes, different types were aggregated (for example, indicated and measured reserves were added together and reported as one reserve estimate); for the compilation, the aggregation is noted in the comment column of the spreadsheet. Finally, sometimes, reserves in different zones of the deposit were estimated; for the compilation, the reserves in different zones but of the same time were aggregated. The intention of these procedures was to make the compilation as consistent as possible.

Appendix 2. Sample Space

This appendix discusses the sample space that is used for this case study. The sample space is a finite mathematical set in which the elements are the gold mines in the Timmins-Kirkland Lake area. There are several criteria for the gold mines in the sample space. A comprehensive discussion of all criteria is beyond the scope of this report, but a brief discussion of a few criteria is appropriate. First, the ore in the different mines must be created by the same geologic process. Second, the gold mines must be in the same geologic region. This criterion increases the chances that the geologic processes that created the ore are as similar as possible. Third, the gold mines must be at about the same depth below land surface. To understand why this criterion is important, consider mines at the ground surface and mines at great depth. It is likely that the deep mines are more expensive to develop than the shallow mines are; hence, the quantity of ore and gold grades would have to be higher for the deep mines to be profitable, if all other factors are the same. Thus, there would be two different groups of mines, which would greatly complicate the resource predictions. Fourth, the equipment that is used to mine and process the ore must be roughly similar in technological sophistication. To understand why this criterion is important, consider the consequences of a new technology to profitably mine low-grade ore. Mines developed with this new technology likely would have low grades and perhaps even high grades, whereas mines developed with an older technology would have just high grades. Thus, there would be two different groups of mines, which would greatly complicate the resource predictions.

The sample space itself must satisfy three criteria. First, the sample space must be exhaustive; that is, the sample space must comprise all existing and future gold mines. Second, the gold mines in the sample space must be mutually exclusive; that is, there cannot be two or more elements in the sample space for the same gold mine. Third, the sample space must be at the appropriate level of granularity; that is, elements in the sample space must have information that is needed for prediction but not extraneous information. The gold mines in the sample space for this case study satisfy these three criteria.

The abstract concept of the sample space aids understanding of size-biased sampling and simple-random sampling. In addition, the abstract concept of the sample space sets limits on what can be predicted—only gold mines in the sample space can be predicted; gold mines outside the sample space cannot.

The sample space described in this appendix differs from the sample space that is the basis of the three-part method. The latter is defined as the Cartesian product of the set of non-negative integers, which pertains to the number of undiscovered deposits, and the set of positive real numbers, which pertains to the resource quantities in those deposits. The principal problem with this formulation is that the sample space is unrelated to sampling (namely, the process of mineral exploration).

Appendix 3. Requirements for Datasets

This appendix describes the requirements of a dataset that would be used for future research on mineral resource prediction.

The first requirement is that a dataset must constitute a sample space (appendix 2); thus, a dataset must satisfy all criteria of a sample space. Second, a dataset must pertain to an assessment area that is well explored and developed so that there is extensive information about the deposits. Third, a dataset must include even those mines for which some information (for example, the resource grades or the quantities of ore tonnages) is missing. Fourth, a dataset must include the times that production started in the mines and the locations of the mines.

It is desirable that a dataset consists of many mines because, with a large dataset, the deleterious effects of high variability can be mitigated. It also is desirable that the datasets pertain to different mineral resources, in different geologic settings, and in different parts of the world.

