

# A Statistical Study of Survey Errors and Closure Adjustments

or

## How Accurate are our Cave Surveys?

*Copyright 2000, 2009 by Robert Thrun*

Occasionally, cave surveyors will be asked "How accurate is the survey?" This is not an easy question to answer well. It is difficult to obtain information on accuracy and it is difficult to express the information. We really don't have a good quantitative way of expressing survey accuracy. Usually the person who asks such a simple question wants a simple answer, such as "One or two percent." We don't bother to state whether the one or two percent applies to the raw survey measurements or to the distances between points on maps. In some ways this is about as good an answer as we can give. This paper will show what the errors are in various surveys, and will discuss some methods of answering the question about the accuracy.

The British Cave Research Association has a system of survey grades based on accuracy of measurements [1]. It has been proposed that American cavers adopt a similar system [2]. I've gone on record as opposing the BCRA Grading system for several reasons [3]. Will White [4], speaking of survey grades, said "A description of survey technique, instruments used, and closure errors is much more satisfying." This is true, but describing the techniques does not say how accurate the survey is. Ray Cole [5] assigned accuracy numbers to sections of Organ Cave based on the adjustments in the surveys leading to each section.

In a pioneering study on cave survey errors, Denis Warburton [6] compared 28 loop closure errors in 11 cave surveys with what would be expected based on his estimate of the measurement errors. Irwin and Stenner [7] followed on Warburton's work and presented the theory and the curves in somewhat more detail. My cave

survey data reduction program, CMAP, produces error ratios based on Warburton's methods, but not his error estimate curves.

An explanation of the theory is in order. Any measurement, other than counting, has an error. When a measurement is the sum of several other measurements, the overall error distribution soon resembles a normal or Gaussian distribution, the familiar bell-shaped curve. The normal distribution, and most other probability distributions, are characterized by a mean and a standard deviation. If a variable  $x$  has a normal distribution, then it can be said to have a mean  $m$ , a standard deviation  $\sigma$ , and probability density  $Z$ .

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Since we are dealing with the probability of random errors, the mean will be zero.

The normal distribution is one-dimensional. The most probable one-dimensional error is zero. In two or three dimensions, the situation is different. For the two dimensional error to be near zero, both the X-error and Y-error must be near zero. For the error to be near  $r$ , the X-error and Y-error can be anywhere near a circle of radius  $r$ . For a three-dimensional situation, an error near  $r$  can lie anywhere near a spherical shell of radius  $r$ . For random errors, the arithmetic average will always be zero. We are more interested in the absolute values of the errors. In one, two, and three dimensions, if the errors are the same in all directions, they will have the following distributions:

$$\sqrt{2/\pi} e^{-x^2/2\sigma^2} \quad 1\text{-D}$$

$$\sqrt{2\pi} x e^{-x^2/2\sigma^2} \quad 2\text{-D}$$

$$\sqrt{8\pi} x^2 e^{-x^2/2\sigma^2} \quad 3\text{-D}$$

The two and three-dimensional distributions are specific cases of a chi-squared distribution. The two-dimensional distribution is called a Rayleigh distribution, and the three-dimensional distribution is called a Maxwell distribution.

These distributions and the cumulative probability, the probability that the error will be less than a specific value, are shown in Figures 1 and 2. I am omitting figure captions in the interest of saving space. The figure numbers are in the upper right corner of the graphs. Many writers fail to realize that the closure errors in a cave survey are not one-dimensional. They will say that there is a 68% probability that the closure error will be less than one standard deviation, etc. Other writers will discuss “probable error”. All errors are probable in that they have some probability of occurring. It is more precise to speak of the “most probable error”, which is simply the maximum of the probability distribution. We may also be interested in the median or mean value of the distribution.

	Maximum	Median	Mean
1-D	0.000	0.674	0.798
2-D	1.000	1.177	1.253
3-D	1.414	1.538	1.596

The squares of the standard deviations, called *variance*, add linearly. If we have  $n$  survey shots of length  $\ell$  adding up to one traverse of length  $L$ , then

$$\sigma_L = \sigma_\ell \sqrt{n}$$

Since the standard deviation is proportional to the probability of an error, we may wish to

speak of percent error, using standard deviation as an approximation of the expected error:

$$\begin{aligned} \frac{100\sigma_L}{L} &= \frac{100\sigma_\ell}{\ell} \frac{1}{\sqrt{n}} \\ &= \frac{100\sigma_\ell}{\ell} \sqrt{\frac{\ell}{L}} \end{aligned}$$

For a given survey shot length, the percent error should go down with  $1/\sqrt{n}$  or  $1/\sqrt{L}$ . Short loops should have large percentage errors and very long loops should have low percentage errors. Many surveyors brag about very long loops with very low percentage error, but this is exactly what they should expect in the absence of blunders or systematic errors. When I first did least-squares adjustments of cave surveys, I was disconcerted by the large percentage adjustments that were made to short traverses between junctions in the survey network. This led me to report the ratio of adjustment to expected error. At the time I put this feature in my program, I made estimates of the measurement accuracies for a good survey. The program, in its usual distribution, still uses these estimates even though experience now shows them to be optimistic

Most analyses of cave survey errors make some simplifying assumptions about the errors. In the preceding analysis of error buildup I assumed that all the survey shots were the same length. Otherwise, I would have an infinite combination of survey shot lengths. The compass variance is assumed to be constant, but it may depend on the steepness of the survey shot. The tape error is also assumed to be fixed, although tape sag or stretch may make long lengths less accurate. When there is a station position error in the analyses, it is assumed to be fixed, although there are some techniques that can minimize or eliminate position error. Conditions vary from shot to shot and affect accuracy. Some surveyors are more accurate than others. I go with the majority and use fixed errors because I don't have enough solid information to do otherwise.

A cave survey shot has a length error, a compass error, a clinometer error, and a station position error. The length, compass, and clinometer errors are at right angles to each other. They determine an error ellipse in two dimensions and an ellipsoid in three dimensions. It is possible to add the error ellipses or ellipsoids of successive survey shots of a loop to form an overall ellipse or ellipsoid for the closure error. I presume that the error distributions would resemble the chi-squared distributions, but with elliptical shells instead of spherical shells.

The way that the standard deviation of the combined random errors accumulates depends on the kind of errors. If the errors are strictly proportional to length, then the shorter the survey shots, the better. If there is a fixed error on each survey shot, then the longer the survey shots the better. Schwinge [8] showed that there is an optimum survey shot length that minimizes the error volume. This occurs when the length error is equal to the angle error. If a given distance is to be covered by a given number of survey shots, we have the least expected error when all the survey shots are the same length. Obviously we can't choose our survey shot lengths solely to minimize random errors, but we should avoid very short or very long shots. On long shots we should try to measure angles accurately, and on short shots we should try to reduce length and position errors.

Systematic errors, where the error is the same amount in the same direction on every shot, build up linearly with the number of survey shots or the length of the traverse. Some systematic errors might not show up as closure errors. For instance, an error in the magnetic declination will simply rotate the entire survey.

A simple loop returns to the starting point. A least-squares adjustment of a survey has traverses between junctions. It is not obvious whether or not the amount of adjustment in a traverse between junctions in a network is the same as the amount of adjust in a traverse that returns to its starting point. At one time I considered doing random walks thru survey net-

works and comparing the closure errors to the error ellipsoids. I estimate that a survey network with  $N$  junctions has between  $2^N$  and  $3^N$  closed loops. An exhaustive search of the larger networks would be impossible, so the best we can do is a random sampling of loops. I never programmed the random walks, but I encourage anyone who wants to improve on this paper to do so.

For this study I used the quantities already returned by CMAP. It reports the amount of adjustment of each traverse, that amount as a percentage of the traverse length, and the ratio of the adjustment divided by an estimate of the most probable error. In writing CMAP, I made the simplifying assumption in the closure adjustment procedure that the errors are the same in all directions. This simplifies the math, reduces the amount of memory needed by arrays by a factor of 9, and reduces the solution time by a factor of 27. Experiments with alternative  $1/n$  and  $1/L$  weightings shows the final positions of the survey stations to be remarkably insensitive to the weighting scheme used. I suspect this is due to the fact that the adjustments correspond closely to the errors.

The estimates of measurement errors I used in CMAP are reasonable, but arbitrary. Other cave survey programs use other arbitrary estimates. The only accuracy specification that has a widespread acceptance is the BCRA system of survey grades. These too were arrived at arbitrarily, but they are accepted, at least in some countries. Whenever someone labels a map as a "Grade 5 Survey", he is making a claim that the survey methods meet the BCRA Grade 5 accuracy specification. I have seen many British maps so labeled, but I have never heard of any effort to determine if the survey actually met that specification. Many compass, clinometer, and tape surveys are simply called Grade 5. This paper will be the first attempt to see if surveys actually meet the Grade 5 requirements.

The specification of a Grade 5 survey is given by Ellis [1]:

Grade 5 A magnetic survey. Horizontal and vertical angles **accurate** to  $\pm 1$  degree; distances **accurate** to  $\pm 10$  cm; station position error less than  $\pm 10$  cm.

The BCRA specification emphasizes accuracy:

3. The term accuracy, used in the definitions, means the nearness of a result to the **true** value; it must not be confused with precision which is the nearness of a number of repeat results to each other, irrespective of their accuracy.

5. It is essential for instruments to be **properly** calibrated to attain grade 5 — details of calibration are given in "Surveying Caves".

The term calibration is often confused with graduation. Graduation refers to the markings on the instrument. For instance, a compass dial might have graduations at 1-degree intervals. Calibration means establishing the differences between a reading and a true reading. The text in "Surveying Caves" suggests taking surface readings to get true north. This procedure corrects for both magnetic declination and differences between compasses. A compass may also have an eccentricity error, caused by the pivot not being centered relative to the dial markings.

The language of the BCRA specifications is that of the machine shop, not that of statistics. When a dimension is given with plus or minus limits, it means that anything between the limits is acceptable, and anything outside the limits is unacceptable. This is not the same as specifying a standard deviation where 32% of the readings may be outside the 1-sigma limits. The authors of the Surverx program interpret the BCRA specification as the 3-sigma limits, which would put 99.7% of the readings within these limits. I interpret the BCRA specification as a uniform distribution. Bryan Ellis, who helped develop the BCRA specifications, agreed with me [9].

I use the following standard deviations for the BCRA errors in horizontal and vertical angles, lengths, and station positions (with conversion from centimeters to feet):

$$\sigma_{angle} = \ell \sin 1^\circ / \sqrt{3}$$

$$\sigma_{length} = 10 \text{ cm} / \sqrt{3}$$

$$\sigma_{position} = 10 \text{ cm} / \sqrt{5}$$

The angle and length error limits form a rectangular prism. The position error limit is a sphere. Ellis [1] says the position error applies to both the instrument and target locations. I apply it only once per survey shot because I think that is his intent.

In CMAP the variance for each survey shot is the station position variance plus the maximum of the length variance and the angle variance. For a shot of the optimum length, the variances are the same in all directions and the CMAP approximation is correct. For shots less than the optimum length, this overestimates the angle errors. For shots longer than the optimum length, this overestimates the length error. For the BCRA specification, the optimum length is 18.8 feet (5.73 meters) in two dimensions, and 26.6 feet (8.10 meters) in three dimensions.

For many of my studies the first cave I try is the Friars Hole System (West Virginia). It is a very large system with large loops and small loops, and surveys of varying quality. A plot of the error ratio versus traverse length is shown in Figure 3. The error ratio of a traverse is the amount of adjustment divided by the most probable error. I experiment with different ways of plotting. The use of linear scales, rather than logarithmic scales, bunches the data near the axes. I did not use percentage error because, as I mentioned before, the short traverses have large percentage errors. CMAP lists the number of shots in each traverse. When I used this, there were too many traverses bunched at each number of shots, particularly 1 and 2-shot traverses.

The plot of error ratio versus traverse length for Friars Hole produces an asymmetrical cloud of points. The center of the cloud appears to be near an error ratio of 2.0. There are some very bad traverses. There are fewer good long traverses.

I always like to verify my computer results with a case that has a known answer. I could not think of a way to generate an artificial grid with a realistic distribution of survey shot lengths and traverse lengths. I hit on the idea of adding artificial errors to the adjusted cave survey. I made a special version of CMAP that did a complete adjustment, then generated random errors using the BCRA Grade 5 specification for the limits, added them to the adjusted survey, and wrote the survey out. I then ran the survey back thru CMAP. It's nice having all the source code so things like this can be done. The adjusted survey with added errors is the worst a survey of that pattern can be and still qualify as a Grade 5 survey. The survey with artificial errors produced a cloud of points that was centered around an error ratio of 0.5 or 0.6, rather than 1.0.

Since the cloud centered around 0.5 or 0.6 was unexpected, I checked all my programming and math. There is an implicit assumption that the error in a traverse is the same as its closure adjustment. This is true for simple loops, but it is not true for traverses between the adjusted locations of junctions in a network. I generated a series of surveys that were regular polygons of various sizes, ranging from triangles with 1.5-foot sides to polygons with 80 sides of 120 feet each. I got a cloud centered around an error ratio of 1.0. The absolute adjustment went up with range, and the percentage error went down with range, with approximately  $\sqrt{L}$  or  $1/\sqrt{L}$  as theory predicts. The reason is that the standard deviation different for different length shots. The clouds are made up of many shot lengths with different lengths predominating at the ends and the middle. Finally, the error ratios are shown for polygons that had been adjusted and given BCRA errors by CMAP. This shows that

CMAP adds errors to the adjusted survey in the same way that the standalone program generated errors. I did not copy code from one to the other, except for the random number generator, which was taken from a good math library.

If there is a traverse with a large error (blunder) in it in the middle of a good survey network and we do a least-squares closure adjustment on the network, part of the error will remain in the bad traverse, and part of the error will be absorbed by the neighbors of the bad traverse. On the basis of adjustments, the bad traverse will look better than it really is, and the good neighbors will look worse. The bad traverses and their neighbors show up on the cloud plots as outliers with large error ratios.

The poorer long traverses may be due to two reasons. Systematic errors, such as differences between compasses, will build up and become more noticeable on long traverses. Long traverses are probably more like simple loops than they are like connections in a network.

We now have a way of assessing the overall accuracy of a cave survey, but it is cumbersome. We assume a standard for measurement accuracies, do a closure adjustment using weights based on that standard, add errors based on the standard to the adjusted survey, and do the adjustment again. We get two plots with clouds of points to compare. The boundaries and centers of the clouds are vague. We don't find out what each of the assumed standard deviations should have been. We only get a rough idea how the entire set was. Even with all the problems, we now have a way of measuring the accuracy of a cave survey, which we did not have before.

To see what sort of accuracies are achieved in actual cave surveys, I gathered as many cave surveys as I could. In my caving area, Virginia and West Virginia, everybody uses the same instruments. Some surveyors are more careful than others. More and more are taking foresights and backsights on every survey shot. I tried to get surveys done with other instruments and methods. I was able to get some older sur-

veys and surveys from other parts of the country. I was not able to convert all the surveys I received to CMAP format. I apologize to those who sent me surveys I did not use. Some caves have too few loops. Others presented format conversion difficulties. I may be able to add them to this paper before I get it published. I tend to categorize the surveys as old or new, Brunton or Suuntos, careful or sloppy, and by geographical area. Rather than confuse you with these categories, I will simply present the surveys alphabetically.

### 2009 Addendum

I was not able to finish my text in 2000. I did not have any conclusions in the text of my paper. If I were rewriting this paper, I would eliminate a lot of the caves and plots. I am simply presenting the plots as I did in 2000. Many of the caves have had additional surveying done since 2000. I have not updated my data.

At the end of the plots I present plots based on adjustment and percentage adjustment. These plots are only to show that these quantities are not suitable for evaluating the accuracy of a survey.

I have some conclusions and observations.

No survey meets 1976 BCRA Grade 5 standards.

Some surveys are so close that there may be a Grade 5 survey somewhere.

There is no distinct dividing point for survey accuracy.

The Hamilton survey, one of the best, did not use calibrated compasses. Much of the errors are due to differences between the directions of magnetic north used by adjacent surveys.

The old survey of New River Cave, done with a Brunton Pocket Transit, was more accurate than the new survey done with Suuntos and backsights.

### References

1. Ellis, Bryan, *Surveying Caves*, British Cave Research Association, 1976, page 2
2. McFarlane, Donald A., "Cave Surveys - A Call for Order", letter, *NSS News*, Vol. 43, No. 7, July 1985, page 223
3. Thrun, Robert, "Survey Grading", *Compass & Tape*, Vol. 6, No. 1, Summer 1988, pages 10-16
4. White, William B., "The Preparation of Geographical Cave Reports", *NSS News*, Vol. 24, No. 5, May 1966, pages 85-92
5. Cole, Raymond Jr., "Survey Accuracy Data", Appendix F of *Caves of the Organ Cave Plateau*, Paul J. Stevens, editor, West Virginia Speleological Survey Bulletin 9, 1988, pages 195-196
6. Warburton, Denis, "The Accuracy of a Cave Survey", *Wessex Cave Club Journal*, No. 89, pages 166-181, 1963, reprinted in *Compass & Tape*, Vol. 8, No. 3, Winter 1990-1991, pages 13-19
7. Irwin, D.J., and R.D. Stenner, "Accuracy and Closure of Traverses in Cave Surveying", *Transactions of the British Cave Research Association*, Vol. 2, No. 4, pages 151-165, December 1975
8. Schwinge, Heinz T., "The Accuracy of Cave Survey", *Bulletin of the National Speleological Society*, Vol. 24, Part 1, January 1962, pages 40-47
9. Ellis, Bryan M., Letter, *Compass Points*, Issue 16, June 1997, pages 4-6



































