Averaging Cave Survey Measurements

David McKenzie, Austin TX
Adapted from Compass & Tape, Vol. 4, Nos. 3 & 4, Spring 1987

I. Introduction

The concept of variance is crucial to the theory of survey network adjustment – the resolution, or averaging, of conflicting measurements. In particular, when the *least squares* method is used to adjust cave survey data, assumed variances are used to assign *weights,* or relative degrees of confidence, to our recorded measurements. If these variances don't reflect what we believe to be the true statistical behavior of our measurement errors, then at least some information in the data is being thrown away.

This may or may not be an acceptable loss. An adjustment method based on simple assumptions about variance is certainly not "wrong", and to some extent it's unavoidable – especially in cave surveying. It does, however, ignore information that could potentially improve our location estimates. The question is: how much improvement? An operation that averages unbiased measurements, even when less than optimally weighted, can still take advantage of redundancy (surveyed loops) while possibly being easier to implement and describe. So how accurate do our weight assignments need to be from a practical standpoint? To answer this question quantitatively we'll need to derive, at least for comparison purposes, a model for variance that's as realistic as we can make it.

II. The Role of Variance

To visualize the problem of cave survey adjustment, consider a set of random vectors – perhaps a set of line segments "vibrating" in such a way that the length and orientation of any segment at a future time can be characterized by a probability function. Our goal might be to estimate from a sample, or snapshot, the means of those vectors when the means are known to form a consistent network. We know which endpoints attach to which, but we also observe that the sampled vectors are inconsistent. This means that any attempt to connect their endpoints produces gaps. How close a fit to expect will depend on a property of the probability function known as *variance* – a measure of the random fluctuation of an observed vector about its mean. If the vectors are derived from survey measurements, for example, the assumed variances might be based on the estimated standard deviations of errors, assumed random, of reading the instruments and positioning the instruments at marked stations.

More formally, the variance of a random variable is its expected squared deviation from the mean. The usual case is that the variable in question takes on a single numeric value; however, our approach here will be to deal directly with 3-dimensional displacements, or vectors. If we regard the error in a vector observation as a column vector whose components are random variables with zero means, then error variance is defined as the expected value (mean) of the error vector times its transpose. Variance, in this case, is a symmetric nonnegative definite matrix, or *covariance matrix*. The least squares method of estimation is most effective when

vector observations are scaled, or *weighted*, by the inverses of their corresponding covariance matrices.

We need to take the concept of variance beyond just the weighting of observations, however. If our estimation method is a function of the data alone, then our estimates, obtained repeatedly with new snapshots, will also be random vectors with statistical properties. As will be explained in Part IV, effective data screening requires that we calculate the variances of our estimated means when the error variances of the observed vectors are known (or rather assumed). Also, we should note that the elements of an error covariance matrix are the coefficients of a quadratic equation describing ellipsoidal surfaces of constant error probability (assuming normal distribution of errors). Such surfaces can be used to define *confidence regions* for an estimated location. Variances, therefore, are not only important for data screening and averaging; they are also necessary for assessing survey accuracy.

III. NTV Model for Error Variance

Suppose that a three-dimensional random vector is viewed with respect to the coordinate frame that aligns with a particular sample, or observation, of that vector. In other words, if the vector connecting two points were derived from an observed length (*l*), azimuth (*a*), and inclination (*i*), we would rotate our coordinate frame so that the vector would coincide with the x-axis and could be written

$$\overline{\mathbf{x}} = (\overline{x} \quad \overline{y} \quad \overline{z})^T = (l \quad 0 \quad 0)^T . \tag{1.1}$$

(Normally, we will follow the convention that the east, north, and up directions coincide with the x-, y-, and z-axis directions, respectively. This is an example of a *right-handed* coordinate system: thumb, forefinger, and middle finger representing the x-, y-, and z-axes.)

In terms of the vector's mean or "true value", $\tilde{\mathbf{x}}$, which is unknown, we would then have

$$\begin{aligned} \overline{\mathbf{x}} &= \tilde{\mathbf{x}} + \mathbf{e}_{\mathbf{x}} , \\ (l \quad 0 \quad 0)^T &= \tilde{\mathbf{x}} + (e_x \quad e_y \quad e_z)^T , \end{aligned} \tag{1.2}$$

where vector $\mathbf{e}_{\mathbf{x}} = (e_x \quad e_y \quad e_z)^T$ is considered a random variable, the unknown error in the new reference frame. The motivation in choosing this rotated frame is that it will simplify our derivation of the covariance matrix associated with a vector observation – in this case, the variance of vector $\mathbf{e}_{\mathbf{x}}$, which is

$$\mathbf{V}_{\mathbf{x}} = E(\mathbf{e}_{\mathbf{x}} \mathbf{e}_{\mathbf{x}}^T) , \tag{1.3}$$

where *E* is the mean or "expected value" operator. Aligning the frame with the observed vector will simplify the math and also make the operation somewhat intuitive. Once we have the vector's covariance matrix in this frame, we can easily apply a linear transformation to convert it

to an equivalent covariance matrix in the original x-y-z reference frame. I'll introduce this transformation later.

The covariance matrix elements should, of course, depend on the error properties of actual measurements. In cave surveying the quantities normally measured, shot lengths, azimuths, and inclinations, resemble the coordinates of a spherical coordinate system. (We'll consider later the influence of station positioning errors.) Therefore, to approximate the covariance matrix of $\mathbf{e_x}$ in (1.2), we will express this vector in terms of a spherical coordinate system $(r\text{-}a\text{-}q)$ such that

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} \bar{r}\cos\bar{q}\sin\bar{a} \\ \bar{r}\cos\bar{q}\cos\bar{a} \\ \bar{r}\sin\bar{q} \end{bmatrix}. \tag{1.4}$$

From (1.1), the observed vector's spherical coordinates would be $\bar{r} = l,\ \bar{a} = p/2, \text{and } \bar{q} = 0$, with angles expressed in radians. We then have

$$\begin{aligned} \mathbf{e_x} = \bar{\mathbf{x}} - \tilde{\mathbf{x}} &= \begin{bmatrix} \bar{r}\cos\bar{q}\sin\bar{a} \\ \bar{r}\cos\bar{q}\cos\bar{a} \\ \bar{r}\sin\bar{q} \end{bmatrix} - \begin{bmatrix} (\bar{r}-e_r)\cos(\bar{q}-e_q)\sin(\bar{a}-e_a) \\ (\bar{r}-e_r)\cos(\bar{q}-e_q)\cos(\bar{a}-e_a) \\ (\bar{r}-e_r)\sin(\bar{q}-e_q) \end{bmatrix} = \\ & \begin{bmatrix} l-(l-e_r)\cos e_q \sin(p/2-e_a) \\ -(l-e_r)\cos e_q \cos(p/2-e_a) \\ (l-e_r)\sin e_q \end{bmatrix} = \begin{bmatrix} l-(l-e_r)\cos e_q \cos e_a \\ (l-e_r)\cos e_q \sin e_a \\ (l-e_r)\sin e_q \end{bmatrix}. \end{aligned} \tag{1.5}$$

Although the components of $\mathbf{e_x}$ are nonlinear functions of the spherical coordinate errors, e_r, e_a, and e_q, a simple approximation emerges when the errors are sufficiently small:

$$\mathbf{e_x} = \begin{bmatrix} e_x \\ e_y \\ e_z \end{bmatrix} = \begin{bmatrix} e_r \\ -le_a \\ le_q \end{bmatrix} + \mathbf{r} \tag{1.6}$$

The vector $\mathbf{r}$ on the right is the remainder of a Taylor series expansion about the zero vector, with the components of $\mathbf{r}$ involving only higher powers of the spherical coordinate errors.

Now let's express $\mathbf{e_x}$ in terms of the actual measurement errors e_l, e_a, and e_i. Clearly, rotating the original coordinate frame to align with the observed vector had no effect on the sizes of the length and inclination errors; hence $e_l = e_r$, and $e_i = e_q$. This isn't true, however, for the azimuth error. Consider, for example, two adjacent meridians of the globe. Since they converge at the poles, their angular separation when viewed from the earth's center becomes smaller as the line of sight moves from the equator to a pole. This decreasing angle corresponds to spherical

coordinate error e_a in our rotated frame, whereas e_a, the error in the *horizontal* azimuth, corresponds to the angular separation at the equator. In fact, for small e_a we have

$$e_a = \tan^{-1}(\cos i \tan e_a) \approx e_a \cos i \,. \tag{1.7}$$

Finally, ignoring the remainders in (1.6) and (1.7) gives us the following approximation:

$$\mathbf{e_x} \approx (e_l \quad -l_h e_a \quad l\, e_i)^T \,, \tag{1.8}$$

where $l_h = l\cos i$ is the observed vector's horizontal length. Thus, when the measurement errors are small enough, the error in the derived vector observation approximates a random *linear function* of those measurement errors. Also, if the measurement errors are independent with zero means, then so are the components of $\mathbf{e_x}$, although only in an approximate sense.

The NTV model for error variance assumes that the expected measurement errors in tape, compass, and inclination measurements are small enough, say $|e_l| \le 0.02\, l$, $|e_a| \le 3^\circ$, $|e_i| \le 3^\circ$, such that their contribution to the covariance matrix of $\mathbf{e_x}$, in the *observed vector's coordinate frame,* is reasonably approximated by a diagonal matrix:

$$\mathbf{V}_{NTV} = E(\mathbf{e_x}\mathbf{e_x}^T) \approx diag(v_N, v_T, v_V) \,. \tag{1.9}$$

Therefore, the effect of measurement errors alone is a set of vector error components that are oriented *normal* (parallel), *transverse*, and *vertical* (NTV) to the vector's line of sight. In (1.9) we are assuming that these error components are statistically independent (uncorrelated) with individual variances v_N, v_T, and v_V, respectively. The approximation (1.8) suggests we use

$$\mathbf{V}_{NTV} \approx diag(s_l^2,\ l_h^2 s_a^2,\ l^2 s_i^2) \,, \tag{1.10}$$

where s_l, s_a, and s_i are the assumed measurement standard deviations. Note that the variance associated with the azimuth measurement, $l_h^2 s_a^2$, will decrease with increasing shot inclination. This follows from the assumption that the compass standard deviation, s_a, is independent of shot inclination. An azimuth error of given size, regardless of its "likelihood", will have a lesser effect on estimated location for inclined shots than for horizontal shots. In cave surveying it's true, however, that reading a magnetic compass is generally less reliable the more a shot is inclined. One possible way to account for this greater uncertainty in tilted compass readings is to add another parameter, say s_c, to our NTV variance model such that

$$\mathbf{V}_{NTV} \approx diag(s_l^2,\ l_h^2 s_a^2 + l_h^2 s_c^2 \sin^2 i,\ l^2 s_i^2) \,, \tag{1.11}$$

where i is the measured inclination.

Due to its simple geometric interpretation, the NTV model can easily be revised in other ways as well. For example, if the azimuth or inclination is the average of a pair of foresight-backsight measurements, v_T and/or v_V can be halved to account for this. The key idea is that the NTV variances should describe three mutually independent sources of error whose directions of influence are normal (parallel), transverse (left-right), and vertical (up-down) to the line of sight coinciding with the observed vector.

If a measured vector in a cave survey were fortuitously aligned with the x-axis of our global frame of reference, matrix $\mathbf{V}_{NTV}$ of (1.11) might approximate the vector's covariance matrix with respect to that frame (again ignoring station positioning errors for the moment). In this special case, the off-diagonal elements of the covariance matrix would be zero. Now, suppose that the vector were not so favorably aligned, but still had the same error properties. Then there would exist a linear transformation, a rotation matrix $\mathbf{R}_\mathbf{x}$, that transforms the original observation, $\mathbf{x}$, to a new observation, $\bar{\mathbf{x}} = \mathbf{R}_\mathbf{x}\mathbf{x} = (l \quad 0 \quad 0)^T$, with a diagonal error covariance matrix:

$$\begin{aligned} &\bar{\mathbf{x}} = \mathbf{R}_\mathbf{x}\mathbf{x} = \mathbf{R}_\mathbf{x}\tilde{\mathbf{x}} + \mathbf{R}_\mathbf{x}\mathbf{e}_\mathbf{x}, \\ &\mathbf{V}_{NTV} = E((\mathbf{R}_\mathbf{x}\mathbf{e}_\mathbf{x})(\mathbf{R}_\mathbf{x}\mathbf{e}_\mathbf{x})^T) = \mathbf{R}_\mathbf{x}E(\mathbf{e}_\mathbf{x}\mathbf{e}_\mathbf{x}^T)\mathbf{R}_\mathbf{x}^T = \mathbf{R}_\mathbf{x}\mathbf{V}_\mathbf{x}\mathbf{R}_\mathbf{x}^T \approx diag(v_N, v_T, v_V). \end{aligned} \tag{1.12}$$

Unless $\mathbf{x}$ is purely vertical $(i = \pm p/2)$, we can compute $\mathbf{R}_\mathbf{x}$ in terms of the vector's observed components $(x, y, \text{ and } z)$:

$$\mathbf{R}_\mathbf{x} = \begin{bmatrix} x/l & y/l & z/l \\ -y/l_h & x/l_h & 0 \\ -xz/(l\,l_h) & -yz/(l\,l_h) & l_h/l \end{bmatrix}, \tag{1.13}$$

where $l_h = \sqrt{x^2 + y^2}$. Alternatively, we can compute $\mathbf{R}_\mathbf{x}$ in terms of just the angle measurements (a and i):

$$\mathbf{R}_\mathbf{x} = \begin{bmatrix} \sin a \cos i & \cos a \cos i & \sin i \\ -\cos a & \sin a & 0 \\ -\sin a \sin i & -\cos a \sin i & \cos i \end{bmatrix}. \tag{1.14}$$

Note that $\mathbf{R}_\mathbf{x}$ is *orthogonal*: $\mathbf{R}_\mathbf{x}\mathbf{R}_\mathbf{x}^T = \mathbf{R}_\mathbf{x}^T\mathbf{R}_\mathbf{x} = \mathbf{I}$. Therefore, for our arbitrarily oriented observation $\mathbf{x}$, we obtain from (1.12) its error variance due to measurement errors alone:

$$\mathbf{V}_\mathbf{x} = \mathbf{R}_\mathbf{x}^T\mathbf{R}_\mathbf{x}\mathbf{V}_\mathbf{x}\mathbf{R}_\mathbf{x}^T\mathbf{R}_\mathbf{x} = \mathbf{R}_\mathbf{x}^T\mathbf{V}_{NTV}\mathbf{R}_\mathbf{x} \approx \mathbf{R}_\mathbf{x}^T diag(v_N, v_T, v_V)\mathbf{R}_\mathbf{x}. \tag{1.15}$$

Except for special orientations of $\mathbf{x}$, $\mathbf{V}_\mathbf{x}$ is a covariance matrix featuring nonzero off-diagonal elements.

Finally, in cave surveying there is another important source of error, *station positioning*, that has little to do with instrument readings or shot orientation. Since it is an independent source, we can use a separate covariance matrix to describe it. We then add this matrix to the one describing the measurement errors:

$$\mathbf{V}_{\mathbf{x}} \approx \mathbf{R}_{\mathbf{x}}^{T} diag(s_{l}^{2},\ l_h^2 s_a^2 + l_h^2 s_c^2 \sin^2 i,\ l^2 s_i^2)\,\mathbf{R}_{\mathbf{x}} + diag(s_h^2,\ s_h^2,\ s_z^2)\,. \tag{1.16}$$

Our final model for a vector's variance, therefore, involves six parameters, s_l, s_a, s_c, s_i, s_h, and s_z, that in practice must be either assumed *a priori* or estimated. The only case we haven't yet dealt with is a pure vertical shot $(i = \pm p/2)$, where azimuth and inclination measurements are not taken at all. Here, an appropriate diagonal matrix can be assumed for $\mathbf{V}_{\mathbf{x}}$, an obvious candidate being

$$\mathbf{V}_{\mathbf{x}} \approx diag(s_h^2, s_h^2, s_z^2 + s_l^2). \tag{1.17}$$

A cave survey typically consists of many long strings of observed vectors, called *traverses*. The variance of a traverse (a vector sum) is obtained by simply summing up the covariance matrices of the component vectors. As shown in Part II, the traverse variances play an integral role in the least squares solution of a cave survey network.

IV. Accuracy Assessment and Data Screening

There are well-established methods for deriving confidence regions for location estimates in linear models. *A priori* estimates of precision are based on the network variances alone, which depend on our assumed observation variances, $\mathbf{V}_{\mathbf{x}}$, and how the network is connected together. How well the observations fit the model is reflected in the *residual sums of squares*,

$$S_e = \sum_{\mathbf{x}} (\mathbf{x} - \hat{\mathbf{x}})^T\, \mathbf{V}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mathbf{x}})\,, \tag{1.18}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the observed and estimated vectors, respectively. S_e is the statistic minimized in a least squares solution. It, along with the number of loop closures, or independent cycles (N_c), can be used to revise the network variances and to produce *a posteriori* confidence regions. Usually, a variance is revised by simply multiplying by the *unit variance estimate* (*UVE*), which is (for our 3-dimensional data set)

$$UVE = \frac{S_e}{3N_c}\,. \tag{1.19}$$

Now, if $\mathbf{e}$ is an error vector with covariance matrix $\mathbf{V}$,

$$\mathbf{e}^T\, \mathbf{V}^{-1} \mathbf{e} = c \tag{1.20}$$

is the equation of an ellipsoid in the components of **e**. For example, the constant c = 7.81 gives us a 95-percent confidence region for a three-dimensional, normally distributed error vector with mean zero and variance **V**. To obtain c for other confidence levels we can consult tables for the chi-square distribution ($c^2_{df=3}$). This assumes, of course, that **V** is the true variance of **e**.

In surveying, **e** might represent the error in the estimated displacement, $\hat{\mathbf{x}}$, between two specified stations. There may not exist a direct measurement, **x**, between them. In any case, the *a priori* confidence region would be based on $\bar{\mathbf{V}}_{\mathbf{x}}$, the computed network variance for this displacement (see Part II). $\bar{\mathbf{V}}_{\mathbf{x}}$ is a function of the model alone and doesn't reflect the observed internal consistency of our data set. Therefore, if the loop count N_c is sufficiently large, we might prefer an *a posteriori* confidence region:

$$\mathbf{e}^T \bar{\mathbf{V}}_e^{-1} \mathbf{e} = c\,, \tag{1.21}$$

where $\bar{\mathbf{V}}_e = UVE \times \bar{\mathbf{V}}_{\mathbf{x}}$ is the network variance scaled by the unit variance estimate.

A powerful technique for detecting blunders in observation data depends on knowing how much each observation's separate deletion would reduce the residual sum of squares, S_e. The availability of network variances, $\bar{\mathbf{V}}_{\mathbf{x}}$, allow these quantities to be efficiently computed and compared with what would be expected if there were no blunders. Specifically, if an observed vector **x** with assumed error variance $\mathbf{V}_{\mathbf{x}}$ were deleted from a network with a final estimated displacement $\hat{\mathbf{x}}$ and network variance $\bar{\mathbf{V}}_{\mathbf{x}}$, then the total sum of squares, S_e, would decrease by

$$dS_e = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{V}_{\mathbf{x}} - \bar{\mathbf{V}}_{\mathbf{x}})^{-1} (\mathbf{x} - \hat{\mathbf{x}}) \tag{1.22}$$

The vector's *F-ratio* can then be computed as

$$F = \frac{(dS_e / 3)}{((S_e - dS_e) \;/\; (3(N_c - 1))}\,. \tag{1.23}$$

Given the assumptions of our model (including normality), this statistic has a *central F-distribution* with parameters 3 and $3(N_c - 1)$. The expected value of F is $N_c/(N_c - 2)$ when $N_c > 2$. Therefore, observations with unusually large F-ratios (much greater than one) can be considered suspect in a screening procedure. Note that the denominator in the above expression for F would be the new UVE after the vector's deletion.

Finally, the adjustment that would have been applied to the observed vector if it had been assigned infinite variance (i.e., given zero weight) is easily obtained as

$$\mathbf{x}_e = -\mathbf{V}_{\mathbf{x}} (\mathbf{V}_{\mathbf{x}} - \bar{\mathbf{V}}_{\mathbf{x}})^{-1} (\mathbf{x} - \hat{\mathbf{x}})\,. \tag{1.24}$$

In the cave surveying program WALLS, $\mathbf{x}_e$ is called the observed vector's *best correction*. It's not unusual to find that $\mathbf{x}_e$ matches almost perfectly a mistake discovered later in transcribed field notes.

The computed network variance, $\bar{\mathbf{V}}_\mathbf{x}$, has still another potential use. If A and B are distinct networks of vectors with two nodes in common, the final solution for A, when considered as part of the combined network, could be obtained by representing B as a single vector "observation". We would first obtain the x-y-z components and error variance of this observation by processing B separately to get an estimate of the displacement between the two attachment nodes, along with the corresponding network variance, $\bar{\mathbf{V}}_\mathbf{B}$. The latter would now be considered an observation variance. The final result, after combining this new observation with A, would be exactly the same as if both networks had been solved simultaneously. If a program were designed handle covariance matrices corresponding to multiple attachment nodes, then networks of practically unlimited size could be solved in stages.

V. Simplified Variance Models

A three-dimensional survey consisting of spherical coordinate measurements could be modeled in the above fashion: as a network of random vectors with 3 x 3 covariance matrices. However, with cave surveys in particular, it may be more practical to adjust the model slightly to reduce program complexity and computation costs, especially if the assigned variances would be crude estimates of the "true" variances anyway. One approach that has been used before in adjustment software is to reshape the error ellipsoids so that their planes of symmetry align with a common coordinate system. The variances then become diagonal matrices and the network effectively breaks into three one-dimensional networks. Unfortunately, this simplification produces least squares estimates that depend somewhat on the chosen reference frame.

A more appealing, though rougher, approximation of variance is to assume additionally that each error ellipsoid has a circular horizontal cross section $(s_x^2 = s_y^2)$, in which case a simultaneous solution for all coordinates is possible while storing each variance as two numbers: a horizontal component variance and a vertical component variance. For example, giving each component a variance proportional to vector length produces the historically popular "compass rule" adjustment. This model involves no other parameters and is easy to program efficiently.

With a simplified variance model we might still enjoy the benefits of using the least squares theoretical framework. These include: 1) exposing gross errors so they can be thrown out or corrected, 2) averaging the remaining data to obtain better location estimates, and 3) obtaining confidence regions for those estimates. We must ask ourselves if the discarded information has any practical significance, or if it is worth the cost of extraction. Other influences, such as uncaught mistakes, systematic errors, or just our lack of knowledge about the situation, could easily erase the advantages of a variance model featuring more parameters – parameters which themselves must be estimated.

Are the location estimates resulting from the length-proportional variance approximation significantly less accurate than those obtained with full covariance matrix model? In my opinion,

this is rarely if ever the case in cave surveying. As a software designer, however, I believe that cave surveyors, as software users, should ideally have the *option* to weight their field observations in a theoretically near optimal way. Data that's so often gathered with extreme difficulty deserves our utmost respect.