Map Accuracy Assessment Using Line Intersect Sampling

A. K. Skidmore
School of Geography, University of New South Wales, Kensington, NSW 2033, Australia
B. J. Turner
Department of Forestry, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, Australia

ABSTRACT: One of the most important questions that a user of a map produced from a geographic information system (GIS) can ask is: "How accurate is the map?" Line intersect sampling is used to estimate the length of cover class boundaries on a map that coincide with the true boundaries of the cover classes on the ground. A ratio of coincident boundary to total boundary is proposed as one measure of map accuracy.

INTRODUCTION

TWO MAJOR TYPES OF MAP ERROR have been identified, attribute error and locational error (Chrisman, 1987; Veregin, 1989; Chrisman, 1989). Attribute error (also called thematic or descriptor error) is the error when a thematic attribute or class name is incorrect, but the boundaries are correct. Locational error, which has been variously termed cartographic and positional error, is the error in the position of features such as points, lines, and grids. In reality, both types of error occur together, making them difficult to separate (Chrisman, 1989). Attribute and locational error should be viewed as end points along a continuum, and together they contribute to the overall error (Chrisman, 1989).

The conventional method for assessing attribute accuracy is to take a number of point samples, and check whether the cover class predicted on the map being tested is the same as the actual cover class on the ground, usually through the use of an error or confusion matrix (e.g., Bishop et al., 1975; Story and Congalton, 1986). Error matrices may be constructed from raster or vector maps, though error matrices are used extensively in remote sensing. Locational error is sometimes expressed as root mean square (RMS) error, which is the standard error of the difference between the true and mapped positions for a number of ground control points (ASPRS, 1988). Both these methods require extensive checking of numerous point samples, making their use extremely time consuming.

An alternative technique for assessing locational error, which should be more efficient for field checking, is proposed here for vector maps, though the technique may be generalized to raster maps. The technique is based on line intersect sampling. Line intersect sampling has been applied to a number of forestry related problems including the assessment of logging waste (Warren and Olsen, 1964), the total length of roads and waterways (Matérn, 1964), and forest fuel sampling (Van Wagner, 1968). Hildebrandt (1975), reported in De Vries (1986), suggests that line intersect sampling may be used for estimating the total length of borders between different ecosystems. In this case, line intersect sampling is used to estimate the total boundary length around homogeneous areas of forest cover type maps.

For a cover type map generated from aerial photographs or satellite remotely sensed imagery, the boundary between cover types on the generated map will either correctly or incorrectly follow the true boundary. It is proposed that the ratio of correctly located boundaries to total boundary length is a measure of map accuracy. This ratio is called the "boundary error" and may be calculated using line intersect sampling theory. In addition, a hypothesis that an estimate of boundary length obtained by line sampling equals the actual boundary length is tested.

LINE INTERSECT SAMPLING THEORY

INTRODUCTION

It is assumed that a map showing cover classes (for example forest type classes or agricultural crop types) has been derived by conventional aerial photograph interpretation or satellite image analysis, and geometrically rectified to a cartographic projection. The cover classes have a fixed, unordered orientation.

Transect lines are randomly placed onto the generated map, and a count is made of the number of times that the lines intersect with the cover class boundaries. From this count, boundary length per unit area of map may be estimated.

ERROR BANDS

Experience with forest cover class maps has shown that boundaries between forest types are located on the map with some inaccuracy (Skidmore, 1989). When mapping natural forests, resource managers frequently delineate groups of trees which are relatively homogeneous with respect to structural or floristic characteristics. Such groupings are often called forest types (Baur, 1965). As the environmental conditions change (for example, elevation, parent material, climate), the ability of a species to survive will be increased or decreased and, hence, the composition of the forest will change (Whittaker, 1967). Sometimes this change will be quite abrupt, but often it will be gradual and there may be uncertainty as to where the boundary between the two types should be placed.

Inaccurate boundaries may also be due to map lines having a finite thickness and, hence, an inherent "width" that varies with the scale of the map. Similarly, a remotely sensed image will have boundary pixels that are a mixture of cover types, with the frequency of mixed pixels increasing as the pixel size becomes larger relative to the size of the ground features. Map boundary errors may also occur due to the photogrammetric and cartographic techniques used. Different interpreters will delineate boundaries slightly differently on aerial photographs in response to their qualitative interpretation of pattern, tone, texture, shape, and association (Wilson *et al.*, 1960). Transcribing forest type boundaries from aerial photographs to a cartographically correct map base will also involve operator error. If map lines are digitized for input into a geographical information system (GIS), further operator error may be introduced (Burrough, 1986; Goodchild, 1978). A boundary may also be erroneous due to a cartographer smoothing a curve in order to

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 58, No. 10, October 1992, pp. 1453–1457.

0099-1112/92/5810–1453$03.00/0
©1992 American Society for Photogrammetry
and Remote Sensing

improve the aesthetics of a map. As Chrisman (1984) pointed out, "slick graphic presentation ... may obscure variations in our knowledge."

Chrisman (1989) emphasized that cartographers are sensitive about the term "error," because in cartography the term implies a blunder. However, he notes that a blunder is usually identified and corrected, while the positional and thematic errors being considered here are more akin to statistical errors which occur due to random fluctuations during the process of information extraction and data processing.

Other errors may be associated with the physical medium used to "store" the map (e.g., paper stretch and distortion) and computer rounding of real numbers (Burrough, 1986).

Thus, there is inherent uncertainty surrounding the physical depiction of map boundaries. Blakemore (1984) used the concept of epsilon distance developed by Perkal (1966) to describe the uncertainty associated with a map line. Blakemore (1984) suggested a point can have five states in the vicinity of a polygon line, i.e., definitely in, possibly in, ambiguous (i.e., on-the-line), possibly out, and definitely out (Figure 1a). However, instead of the five states, only three states were used in this study, i.e., definitely in, ambiguous (i.e., on-the-line), and definitely out. The on-the-line class was a merged class containing the possibly in, on-the-line, and possibly out class (Figure 1b). This is reasonable as the possibly in, on-the-line, and possibly out categories are all ambiguous categories, with uncertainty as to the actual name which should be used to represent these categories.

Calculation of the Boundary Length and Variance

The boundary length of polygons per unit area of the map can be estimated using the following formula (De Vries, 1986):

$$\hat{X} = \frac{\pi \cdot m}{2L} \tag{1}$$

where

$\hat{X}$ = estimated boundary length of polygons per unit area,
L = total length of the transect line(s), and
m = the number of intersections of the transect line(s) with the polygon boundaries.

The estimated variance of the boundary length of polygons per unit area of the map is

$$\hat{\text{var}}\ \hat{X} = \left(\frac{\pi}{2L}\right)^2 m \tag{2}$$

Number of Samples

The estimated number of line samples ($\hat{n}$) required to meet a prespecified precision can be calculated (from De Vries, 1986) as

$$\hat{n} = \left(\frac{t_{n-1}^{0.975} \cdot \hat{CV}(\hat{X})}{E}\right)^2 \tag{3}$$

where the coefficient of variation is

$$\hat{CV}(\hat{X}) = \left(\frac{\sqrt{\hat{\text{var}}\ \hat{X}}}{\hat{X}}\right) \tag{4}$$

and E is the prespecified allowable error fraction. The number of sample lines (n) that satisfies Equation 3 is found iteratively such that (De Vries, 1986)

$$\frac{n}{(t_{n-1}^{0.975})^2} = \left(\frac{\hat{CV}(\hat{X})}{E}\right)^2 \tag{5}$$

METHODS

A map of cover types generated from aerial photograph interpretation or image processing of remotely sensed data may be geometrically rectified to a cartographically correct projection. A simplified example of such generated boundaries is shown in Figure 2 for a hypothetical forest type map. (Note that the map has been reduced, so that the measurements recorded below cannot be directly measured from the map.) The error band is represented as a thick line of constant width, and describes the uncertainty associated with a line on a map.

The error band represents the error introduced by a number of factors, such as geometric rectification, map compilation, and ecotone width. The total error variance may be calculated from the weighted arithmetic average of the variances for the contributing error sources (Neter and Wasserman, 1974), where the error is the distance measured from the true boundary to the incorrect point. A random sampling design may be used to calculate the variance contributed by the error sources, with the random samples being placed along the true boundary.

For the purpose of example, let the error variance be 169 m^2 (Figure 2), or a standard deviation of 13 m. According to the Empirical Rule (Mendenhall and Reinmuth, 1978), it is possible to conclude that that the epsilon distance would be ± 26 m

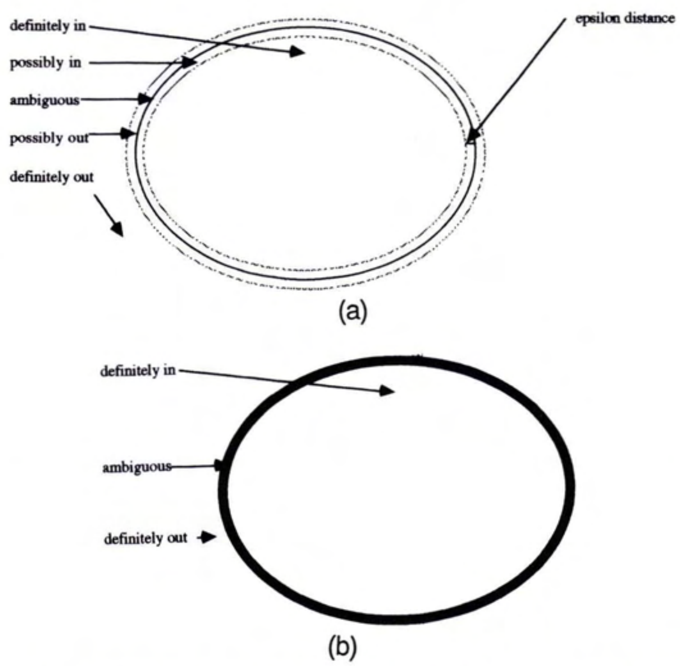

Fig. 1. (a) Error zones as defined by Blakemore (1984). (b) Error zones used in this study.

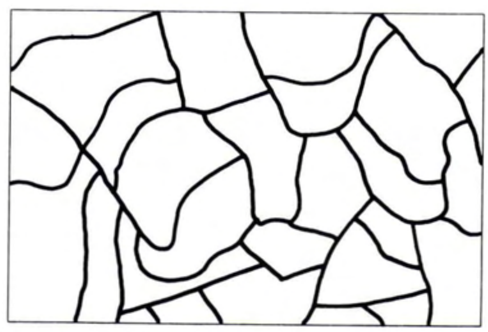

Fig. 2. Forest type map.

about the true boundary line, with a 95 percent confidence level. Thus, if the generated line is within ± 26 m of the true line, we are 95 percent confident that the line is correctly mapped. An alternative approach would be to specify the epsilon distance as the maximum error that will be tolerated for a given scale (e.g., no point should be more than 12.5 m from its true point at a scale of 1:25,000.

A map showing the true forest type boundaries overlaid onto the generated map is detailed as Figure 3.

Line Sampling

It is necessary first to estimate the number of transect lines and the length of transect lines required to estimate parameters at prespecified 95 percent confidence intervals.

It can be shown (see Appendix for details) that

$$C\hat{V}(\hat{X}) = \frac{1}{\sqrt{m}} \tag{6}$$

Now let $E = 0.1$ and $n = 9$. Therefore,

$$t_8^{0.975} = 2.306$$

and, from Equation 5,

$$C\hat{V}(\hat{X}) = 0.13$$

Substituting into Equation 6, $m = 59$. Thus, one needs to obtain approximately 6.5 intersections per transect, over the nine transects, to ensure the estimate of boundary length per unit area is within 95 percent confidence intervals.

A pilot transect line of length 12.5 cm was chosen as this length appeared on average to intersect approximately six times with the cover type polygon boundaries on Figure 3. The transect lines were located on the map (Figure 3) by randomly generating the eastings and northings of the starting point, and randomly generating the compass direction of the transect. Nine transect lines ($n = 9$, $L = 112.5$) were randomly located on the map in this manner (Figure 3).

The points of intersection of the true map boundary lines with the transect lines were identified. If the true map boundary falls within the error bands of the generated map at the intersection with the transect, then the true and generated map lines coincide (point [1] on Figure 3). If the true map line falls outside the error bands, the true and generated map boundaries do not coincide (point [2] on Figure 3).

The total number of intersections (m) of the line transect with the map boundaries was used to calculate the total length of

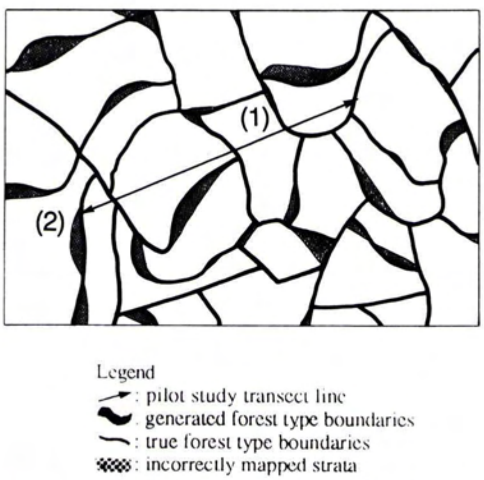

Fig. 3. True location of forest types overlaid onto the generated map.

boundaries per unit area of map. The number of intersections where the true map boundary lay within the generated error bands was also counted and notated as 'r'. For the nine transect lines, the number of intersections with the generated map boundaries was 43 ($m = 43$). There were 28 points at which the true map boundary and generated error bounds also coincided ($r = 28$).

The number of intersections ($m = 43$) is less than the required number ($m = 59$) to ensure the estimate of boundary length per unit area is within 95 percent confidence intervals. This occurred because only five intersections per transect line were being recorded, rather than the required 6.5 intersections per transect, due to the estimated transect line length (12.5 cm) being too short. It was decided to add another two transect lines (again of constant length 12.5 cm) to increase the sample to $n = 11$ and $L = 137.5$ cm. From Equation 5, the coefficient of variation becomes 0.149, and from Equation 6, $m = 45$. The required intersections per transect line is then 4.1.

For the 11 transect lines, the number of intersections with the generated map boundaries was 52 ($m = 52$). There were 34 points at which the true map boundary and generated error bounds also coincided ($r = 34$).

From these counts, the estimated total density (length per unit area) of generated boundary lines (X_m), and the estimated density of boundary lines where the generated map boundary coincided with the true map boundaries (X_r) were calculated. The variance of X_r and X_m was also estimated.

$$X_m = 0.594 \text{ cm cm}^{-2}$$
$$X_r = 0.388 \text{ cm cm}^{-2}$$

A measure of map accuracy is the ratio of X_r to X_m, i.e.,

$$X_t = \{X_r \div X_m \times 100\} = 65.4 \text{ percent}$$

This ratio is interpreted below.

Actual Boundary Lengths and Areas

The length of boundaries on the generated map (Figure 2) and the length of coincident boundaries between the generated map and the true map (Figure 3) were directly measured using a digital planimeter.

The results were as follows:

Area of map = 260.0 cm²
Length of true map boundary = 160.0 cm
Therefore, $X_{m'} = 0.615$ cm cm^{-2}
Length where maps are coincident = 102.6
Therefore, $X_{r'} = 0.394$ cm cm^{-2}.
Thus, $x_{t'} = X_{r'} \div X_{m'} = 64.1$ percent.

On inspection, the actual densities compare very favorably with the sampling results. A research hypothesis was tested: whether there is a significant difference between the true and estimated density of boundaries, i.e.,

Ho: $X_m = X_{m'}$
Ha: $X_m \neq X_{m'}$

and

Ho: $X_r = X_{r'}$
Ha: $X_r \neq X_{r'}$

The confidence interval test was used, i.e., reject Ho and conclude there is a significant difference between the true and estimated length of boundaries if $X_{m'}$ and $X_{r'}$ occur outside the confidence intervals defined by

$$X_m \pm t_{\alpha/2} \frac{S_m}{\sqrt{n}}$$

where

$$S_m = \frac{\pi}{2L}\sqrt{m} = \frac{\pi}{2 \times 137.5}\sqrt{52} = 0.082$$

and

$$X_r \pm t_{\alpha/2}\frac{S_r}{\sqrt{n}}$$

where

$$S_r = \frac{\pi}{2L}\sqrt{r} = \frac{\pi}{2 \times 137.5}\sqrt{34} = 0.067$$

Thus, at $\alpha' = 0.05$, $n = 11$:

$$X_m = 0.594 \pm 2.228\ (0.082 \div 3.317) = \underline{0.594 \pm 0.055}$$

and

$$X_r = 0.388 \pm 2.228\ (0.067 \div 3.317) = \underline{0.388 \pm 0.045}.$$

As $X_{m'} = 0.615$ and $X_{r'} = 0.394$, we conclude that the estimated boundary density is not significantly different from the true boundary density at $\alpha' = 0.05$.

Two strata were recognized on Figure 3, that is, "correctly mapped" and "incorrectly mapped." The area of both strata were measured with a digital planimeter. The total area of the map was 260.00 cm², and the area that was incorrectly mapped was 19.17 cm². The areal map accuracy was therefore 92.7 percent.

DISCUSSION

The true boundary densities were not significantly different from the boundary densities estimated using line intersect sampling, with $\alpha' = 0.05$. The estimated boundary accuracy (65.4 percent) was very close to the true boundary error (64.1 percent).

Mapping accuracy expressed as boundary accuracy is a conservative measure of map accuracy, because it measures coincident boundary length, and not the area of the map that is correctly classified. In other words, the boundary accuracy is a lower map accuracy estimate compared with an area accuracy statement, and would therefore give an increased probability that the map would not have an acceptable mapping accuracy.

The technique has been evaluated in the field, using the following method. The start of each transect was identified using aerial photographs (note that perhaps a global positioning satellite receiver may be used). A chain and compass was used to measure the transect length and direction; the position of the "true" type change, as well as the position of the uncertain ecotone areas, were noted along the transect. In the office, the epsilon distance may be calculated, based on the ecotone distances and processes such as geometric rectification and map compilation. The transects are plotted with the position of the "true" type change indicated (Figure 4a). The forest type lines generated by the aerial photograph interpreter are overlaid onto the transects (Figure 4b). Finally, the epsilon error bands may be plotted around the forest type boundaries generated by the interpreter (Figure 4c). A count is made of the total number of intersections of the line transect with the (true) type changes recorded along the transects (m), as well as the number of intersections where the generated map boundary lay within the error bands (r).

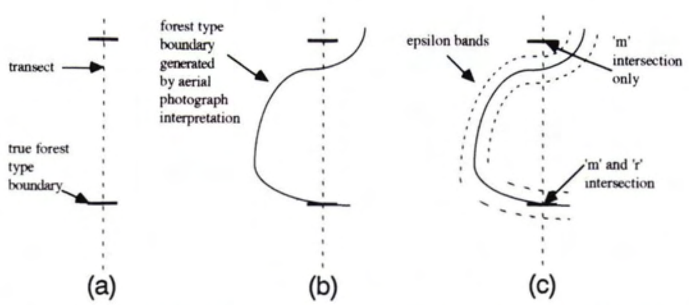

FIG. 4. Field implementation of the line intersect sampling technique.

In the design of the sampling strategy, there is a trade off between having a few relatively long transects, or many relatively shorter transects. If a small number of transects are chosen, then the total number of intersections (m) required is higher than if many transects are sampled (for example, replacing n = 9 with n = 11 for the example cited in this paper results in "m" decreasing from 59 to 45). The final decision on the balance between "n" and "m" will be determined by factors such as the access, ease of locating transect starting points, and the suitability of the area for surveying long transects by compass and chain.

Uncertainty in the boundary position occurs at the transition zone between forest cover types. If there are erroneous forest types occurring as patches within otherwise homogeneous polygons, then the line intersect sampling method would reflect a lower boundary accuracy.

The advantage of the line intersect sampling technique is that it is fast, cheap, and simple compared with sampling numerous points. A major disadvantage with the line intersect method compared with point sampling is that it does not assess the accuracy of area estimates. For example, a forester may wish to calculate the accuracy of the area of forest types or timber volumes from a map.

Line transect sampling offers the possibility of yielding a higher precision estimate of map accuracy for a given cost, as it is usually cheaper and simpler to traverse lines using a compass bearing and chain during ground truth operations, than to identify numerous point sample locations in the field. The technique would be particularly suitable where the edges between different forest types, soil types, or ecosystems are of interest. Examples of potential applications include wildlife habitat modeling and checking the accuracy of aerial photograph interpretation. However, it should again be emphasized that line intersect sampling is but one method of estimating map accuracy, and the appropriate statistic (or statistics) should be chosen for the accuracy assessment task being considered.

ACKNOWLEDGMENT

Insightful comments from three anonymous referees were used to correct and revise the manuscript.

REFERENCES

ASPRS, 1988. ASPRS Interim accuracy standards for large-scale maps. *Photogrammetric Engineering & Remote Sensing* 54(7):1079-1081.

Baur, G.N., 1965. *Forest types in New South Wales*. Forestry Commission of N.S.W., Research Note No. 17, Forestry Commission of N.S.W., Sydney.

Bishop, Y.M., S.E. Fienburg, and P.W. Holland, 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Massachusetts.

Blakemore, M., 1984. Generalization and error in spatial databases. *Cartographica* 21(2&3):131-139.

Burrough, P.A., 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Oxford University Press, Oxford).

Chrisman, N.R., 1984. The role of quality information in the long-term functioning of a geographic information system. *Cartographica* 21(2&3):79-87.

———, 1987. The accuracy of map overlays: a reassessment. *Landscape and Urban Planning* 14:427-439.

———, 1989. Modeling error in overlaid categorical maps. *The Accuracy of Spatial Databases*, Chapter 2 (M. Goodchild and S. Gopal, editors). Taylor and Francis, London.

De Vries, P.G., 1986. *Sampling Theory for Forest Inventory*. Springer-Verlag, Berlin.

Goodchild, M.F., 1978. Statistical aspects of the polygon overlay problem. *Harvard Papers on Geographic Information Systems*, Volume 6 (G. Dutton, editor). Addison-Wesley, Reading, Massachusetts.

Hildebrandt, G., 1975. Dir verwendung von luftbild-linienstichproben zur ermittlung der länge linienförmiger geländeobjekte. *Allgemeine Forst Zeitschrift* 30(11):29-31.

Matérn, B., 1964. A method of estimating the total length of roads by means of a line survey. *Studia Forestalia Suecica* 18:68-70.

Mendenhall, W., and J.E. Reinmuth, 1978. *Statistics for Management and Economics*. Wadsworth, Belmont, California.

Neter J., and W. Wasserman, 1974. *Applied Linear Statistical Models*. Irwin, Illinois.

Perkal, J., 1966. On the length of empirical curves. Discussion paper 10, Michigan Inter-University Community of Mathematical Geographers. Cited by Blakemore, M., 1984. Generalization and error in spatial databases. *Cartographica* 21(2&3):131-139.

Skidmore, A.K., 1989. An expert system classifies eucalypt forest types using Landsat Thematic Mapper data and a digital terrain model. *Photogrammetric Engineering & Remote Sensing* 55(10):1449-1464.

Story, M., and R.G. Congalton, 1986. Accuracy assessment: A user's perspective. *Photogrammetric Engineering & Remote Sensing* 52(3):397-399.

Van Wagner, C.E., 1968. The line intersect method in forest fuel sampling. *Forest Science* 14(1):20-26.

Veregin, H., 1989. Error modeling for the map overlay operation. *The Accuracy of Spatial Databases*, Chapter 1 (M. Goodchild and S. Gopal, editors). Taylor and Francis, London.

Warren, W.G., and P.F. Olsen, 1964. A line intersect technique for assessing logging waste. *Forest Science* 10(3):267-276.

Whittaker, R.H., 1967. Gradient analysis of vegetation. *Biological Review* 42:207-264.

Wilson, R.C., J. Carow, R.A. Chapman, G.A. Choate, J.R. Dilworth, D. Burger, R.C. Heller, S.T.B. Losee, R.G. Miller, K.E. Moessner, E.J. Rogers, E.G. Stoeckeler, and H.E. Young, 1960. Photo interpretation in forestry. *Manual of Photographic Interpretation*, Chapter 7, pp. 457-520 (R.N. Colwell, editor). American Society of Photogrammetry, Washington D.C.

(Received 23 April 1990; revised and accepted 22 January 1992)

APPENDIX

Derivation of the formula linking coefficient of variation with "m" is as follows:

$$\hat{X} = \frac{\Pi m}{2L}$$

$$\mathrm{var}\,\hat{X} = \left(\frac{\Pi}{2L}\right)^2 m = \left(\frac{\Pi m}{2L}\right)^2 \frac{1}{m} = \frac{\hat{X}^2}{m}$$

$$\hat{CV}(\hat{X}) = \frac{\sqrt{\mathrm{var}\,\hat{X}}}{\hat{X}} = \frac{\sqrt{\dfrac{\hat{X}^2}{m}}}{\hat{X}} = \frac{1}{\sqrt{m}}$$

LIST OF "LOST" CERTIFIED PHOTOGRAMMETRISTS

We no longer have valid addresses for the following Certified Photogrammetrists. If you know the whereabouts of any of the persons on this list, please contact ASPRS headquarters so we can update their records and keep them informed of all the changes in the Certification Program. Thank you.

Robert Ball
Milosh Benesh
Ivan Bentley
Binisain Bhawani
Dewayne Blackburn
Eugene Caudell
William Clements
Ronald L. Deakin
Robert Denny
Russell Doolittle
Jimmie Felton
Leo Ferran
Charles Foster
Stanley Frederick
Robert Fuoco
Franek Gajdeczka
George Glaser
Jose Gotay
William Grehn, Jr.
David Gustafson
Jack Guth
Elwood Haynes
F.A. Hildebrand, Jr.
James Hogan
Clyde Hubbard, Jr.
Daniel Hughes
William Janssen
Harold Johnson
Lawrence Johnson
Marc Leupin
B. Timothy McGovern
Francisco Milande
Marinus Moojen
M. David L. Morgan
William A. Moriarty
Wesley Norris
Gene A. Pearl
Joe Pinello
Larry Reed
Sherman Rosen
Carl Schafer
Lane Schultz
Charles Sheaffer
Leo Strack
Gelacio Sumagaysay
Wallie Swain
Keith Syrett
William Thomasset
Conrad Toledo
Raymond Town
Robert Tracy
Peter Warneck
Lawrence Watson
Frank J. Wobber