# ACCURATE POLYGON CENTROID COMPUTATION USING ARC/INFO GIS[a]

**Journal of Computing in Civil Engineering 9(1):88. January 1995**

## Discussion by Francis E. Greulich[3]

In their technical note, Warwick and Haness, express a need for a more accurate procedure for the computation of polygon centroids. For those who need rapid, computationally easy, and accurate estimates of polygon centroids, there is a numerical solution procedure that they should consider. Geographers (Blair and Biss 1967), Geologists (Hall 1976) and Forest Engineers (Greulich 1991) have independently discovered and are using the following very general formula that could be of use to other professionals:

$$A = \frac{1}{2}\sum_{i=1}^{N}(x_i y_{i+1} - x_{i+1} y_i) \tag{4}$$

$$x_0 = \frac{1}{6A}\sum_{i=1}^{N}(x_{i+1} + x_i)(x_i y_{i+1} - x_{i+1} y_i) \tag{5}$$

$$y_0 = \frac{1}{6A}\sum_{i=1}^{N}(y_{i+1} + y_i)(x_i y_{i+1} - x_{i+1} y_i) \tag{6}$$

where $A$ = area of the polygonal region as calculated by the coordinate area formula and $(x_0, y_0)$ are the coordinates of the centroid. The vertices, $[(x_i, y_i)\ i = 1, N]$, of the polygonal region are sequentially numbered counterclockwise from an arbitrary starting vertex and $(x_{N+1}, y_{N+1}) \equiv (x_1, y_1)$. Very general polygonal regions may be evaluated using this formula including disconnected polygons and those with holes. For additional details, interested readers are referred to subsequent literature in the previously cited fields.

One advantage of the preceding formula is that for a polygonal region it is exact. It yields results of an accuracy that procedures based on finite subdivision of the region can, in general, only approach in the limit as their cell size goes to zero. However, whereas the method of Warrick and Haness is immediately available, this reader is unaware of any attempt to adapt the formula-based approach to ARC/INFO applications.

## APPENDIX. REFERENCES


Blair, D. J., and Biss, T. H. (1967). "The measurement of shape in geography: an appraisal of methods and techniques." *Bull. of Quantitative Data for Geographers*. 11, Department of Geography, Nottingham Univ., Nottingham, U.K.

Greulich, F. E. (1991). "Optimal landing location on flat, uniform terrain." *Can. J. For. Res.*, 21(5), 573–584.

Hall, J. K. (1976). "Algorithms and programs for the rapid computation of area and center of mass." *Comp. and Geosci.*, 1, 203–205.


## Closure by J. J. Warwick,[4] Member, ASCE, and S. J. Haness[5]

The writers would like to thank Greulich for the discussion. First, while version 5.0.1 of ARC/INFO (*ARC/INFO* 1988) has no command that will accurately compute the centroid location for a polygon, for ARC/INFO Version 6.1 (*ARC/INFO* 1992), the developer, Environmental

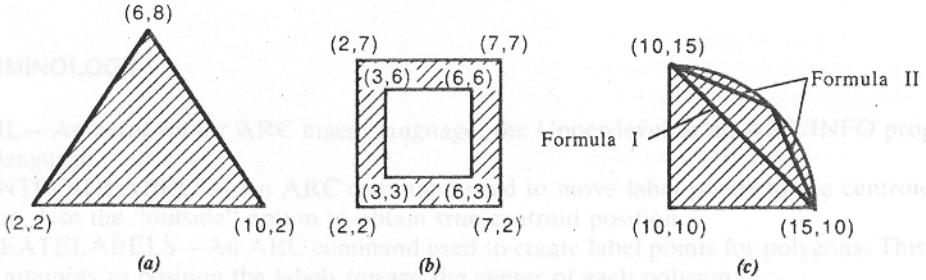


**FIG. 4. Selected Shapes and Associated Vertex Locations**

[a]July, 1993, Vol. 7, No. 3, by J. J. Warwick and S. J. Haness (Technical Note 2196).
[3]P.E., Prof., Forest Engrg., Univ. of Washington, Seattle, WA 98195.
[4]Dir., Grad. Program in Hydro./Hydrogeology, Univ. of Nevada–Reno, Reno, NV 89512.
[5]Agency for Toxic Substances and Disease Registry, 1600 Clifton Rd., E32, Atlanta, GA 30333.

TABLE 1. Comparison of Centroid Computations

| Shape (1) | Method (2) | $A$ ($m^2$) (3) | $x_0$ (m) (4) | $y_0$ (m) (5) |
|---|---|---|---|---|
| Triangle | Exact | 24.00 | 6.00 | 4.00 |
| Triangle | Formula | 24.00 | 6.00 | 4.00 |
| Triangle | ARC/INFO | 24.00 | 6.00 | 4.00 |
| Square doughnut | Exact | 16.00 | 4.50 | 4.50 |
| Square doughnut | Formula | 17.00 | 4.50 | 4.50 |
| Square doughnut | ARC/INFO | 16.00 | 4.50 | 4.50 |
| Quarter circle | Exact | 19.64 | 12.12 | 12.12 |
| Quarter circle | Formula I | 12.50 | 11.66 | 11.66 |
| Quarter circle | Formula II | 17.50 | 12.00 | 12.00 |
| Quarter circle | Formula III | 18.84 | 12.08 | 12.08 |
| Quarter circle | ARC/INFO | 19.64 | 12.12 | 12.12 |

Systems Research Institute (ESRI), has added the CENTROIDLABELS command, which should be executed following CREATELABELS to move the created label point to the centroid of a polygon. The "outside" option should be specified with the CENTROIDLABELS command to find the true polygon centroid. Beyond this update, the writers wish to gratefully acknowledge the discussion, as it will certainly assist others in accurately defining the location of polygon centroids. To further assist interested readers we have conducted a small comparative study using simple geometric shapes for which centroids are mathematically defined. The formulas, as presented in the discussion, from Blair and Biss (1967), Hall (1976), and Greulich (1991) are used along with the new CENTROIDLABELS command found within ARC/INFO version 6.1. The writers' selection of ARC/INFO should not be misconstrued as an endorsement of the ESRI product. ARC/INFO selection was based solely upon its growing popularity within both the public and private sectors.

In (4)–(6), where $A$ is the area of the polygonal region ($m^2$); $N$ is the number of vertices; $x_i$ is the $x$-coordinate of a vertex (m); $y_i$ is the $y$-coordinate of a vertex (m); $x_0$ is the $x$-coordinate of centroid (m); and $y_0$ is the $y$-coordinate of centroid (m), the vertices of the polygonal region must be sequentially numbered counterclockwise from an arbitrary starting vertex.

Three simple geometric shapes were chosen to evaluate the accuracy of the formula method [(4)–(6)] versus using the CENTROIDLABEL command. The selected shapes, along with assigned vertex coordinates ($x_i$, $y_i$), are given in Fig. 4. Table 1 summarizes the comparison for all three shapes tested. All methods give the same result for the triangle. The formula method correctly computes the centroid of the square donut shape but incorrectly computes the area ($A$). Finally, application to the quarter circle depends upon the number of vertices ($N$) used in the formula method. ARC/INFO using 91 vertices (one per degree) to define a quarter circle, which may explain its superior performance over the formula method, which does converge to the correct solution as $N$ increases. The approach shown in Table 1 as "formula I" uses only one line segment to approximate the arc of a quarter circle (also see Fig. 4). Two equal-sized line segments are used to define the arc for approach "formula II" shown in Table 1, while three equal sized line segments are used in the approach called "formula III."

Individuals who wish to apply the formula method within ARC/INFO will need to write their own AML to perform vertex identification, numbering, and formula computations. The algorithm used with ARC/INFO is not noted in the user's manual (*ARC/INFO* 1992) and is therefore generally not available for inspection since the coding is proprietary. ESRI was contacted to explain their method of computing polygon centroids but declined to respond in a timely fashion.

## APPENDIX I. ARC/INFO TERMINOLOGY

AML—An acronym for ARC macro language, the Upper-level or an ARC/INFO programming language.

CENTROIDLABELS—An ARC command used to move label points to the centroid of a polygon. Use the "outside" option to obtain true centroid position.

CREATELABELS—An ARC command used to create label points for polygons. This command attempts to position the labels toward the center of each polygon.

## APPENDIX II. REFERENCE

*ARC/INFO* Version 6.1. (1992). Environmental Systems Research Institute, Redlands, Calif.