THE CENTROID? WHERE WOULD YOU LIKE IT TO BE?

R E Deakin [1]
Dr R.I. Grenfell [2]
and
S.C. Bird [3]

[1] Lecturer, Department of Geospatial Science
Email: rod.deakin@rmit.edu.au

[2] Business Development Manager, Geospatial Science Initiative
Email: ron.grenfell@rmit.edu.au

[3] Student, Department of Geospatial Science
Email: stbird@cs.rmit.edu.au

RMIT University
GPO Box 2476V MELBOURNE VIC 3001
Tel: +61 3 9925 2213 / Fax: +61 3 9663 2517

ABSTRACT

The concept of a centroid is useful for many spatial applications, and the determination of the centroid of a plane polygon is standard functionality in most Geographic Information System (GIS) software. A common reason for determining a centroid is to create a convenient point of reference for a polygon, often for positioning a textual label. For such applications, the rigour with which the centroid is determined is not critical, because, in the positioning of a label for example, the main criteria is that it be within the polygon and reasonably central for easy interpretation.

However, there may be applications where the determination of a centroid has, at the very least, an impact on civic pride and quite possibly financial repercussions. We refer here to an administrative or natural region where a nominated centroid has a certain curiosity value with the potential to become a tourist attraction. Such centroids provide economic benefit to those in a sub-region, usually in close proximity to the centroid.

Various interpretations of a centroid exist and this paper explores these and the methods of calculation. Variation in position resulting from different interpretations is examined in the context of the centroid of the Australian State of Victoria, and GIS software are evaluated to determine the efficacy of their centroid functions.

INTRODUCTION

In physics, it is often useful to consider the mass of a body as concentred at a point called the centre of mass (or the centre of gravity). For a body of homogeneous mass, this point coincides with its geometric centre or *centroid*. Thus, we have the commonly accepted meaning that the centroid is equivalent to the centre of mass. For example, the centre of mass of a homogeneous sphere (a geometric solid) is coincident with its centre.

In geospatial science, we often deal with relationships between points on the surface of the earth, but generally, these points are represented as projections onto a plane (a map) and areas of interest are defined by polygons. The centroid of a polygon, in this case, does not have a tangible connection with the centre of mass of the object, since it is merely a series of lines (or points) on paper or a computer image. Therefore, we often equate the centroid to the geometric centre of the polygon, which in the case of complicated polygons is often impossible to determine. Thus we resort to mathematical formulae and Cartesian coordinates to calculate centroids.

An age-old joke has an accountant being asked for an opinion, the reply to which is, "what would you like it to be". Geospatial professionals asked to determine the centre of a complex polygon might reply in a similar vein. In this paper, in the context of geospatial science, we shall demonstrate that there are several plausible definitions of a centroid, most leading to relatively simple means (or averages) of coordinates, but some requiring more advanced methods of computation. Some examples will be given to demonstrate methods of calculation and highlight cases where restrictions need to be placed on certain centroid definitions.

In addition, we compare centroids computed by different methods with those of several GIS software packages in general use. This comparison should shed some light on the computation methods used by these software products. Finally, some results of the calculation of the centroid(s) of Victoria, one of the States of the Commonwealth of Australia, and its capital Melbourne are presented.

DEFINITIONS OF CENTROIDS

The OpenGIS Specification for Feature Geometry (OGC 1999) defines a centroid object without greatly assisting in its calculation for individual polygons:

> The operation "centroid" shall return the mathematical centroid for this GM_Object. The result is not guaranteed to be on the object. For heterogeneous collections of primitives, the centroid only takes into account those of the largest dimension. For example, when calculating the centroid of surfaces, an average is taken weighted by area. Since curves have no area they do not contribute to the average.

In this section, a number of centroids are defined and named; the names are only relevant to this paper. In some cases computational formulae are given. Additional information relevant to the computation of certain centroids is given in Section 3.

The Moment Centroid

The New Shorter Oxford English Dictionary (SOED 1993) defines the centroid as: "A point defined in relation to a given figure in a manner analogous to the centre of mass of a corresponding body." Using this definition, and regarding the body as a plane area *A* of uniformly thin material, its centroid is

$$\bar{x} = \frac{M_y}{A} \qquad \text{and} \qquad \bar{y} = \frac{M_x}{A} \tag{1}$$

and M_x and M_y are (first) moments with respect to the *x*- and *y*-axes respectively (Ayres 1968). [The moment M_L of a plane area with respect to a line *L* is the product of the area and the perpendicular distance of its centroid from the line.] The centroid computed using this method has a physical characteristic that is intuitively reassuring. That is, if we cut out a shape from uniformly thin material (say thin cardboard) and suspend it freely on a string connected to its centroid, the shape will lie horizontal in the earth's gravity field. In this paper, we will call this centroid a *Moment Centroid.*

The Area Centroid

In a similar vein, we may divide the uniformly thin shape of area *A* into two equal portions A_1 and A_2 about a dividing line $B - B'$. If the shape were symmetrical about this line, it would balance if it were placed on a knife-edge along $B - B'$. If we choose another line $C - C'$, which divides the area equally, then the intersection of the dividing lines (or balance lines) defines a point we call the *Area Centroid*. This centroid also has an intuitively appealing simplicity but unfortunately, for a general polygon of area *A*, different pairs of balance lines intersect at different points! That is, the method does not yield a unique point unless the direction of the dividing lines is defined. In addition, unless the figure is symmetrical about both balance lines, this centroid will not coincide with the moment centroid.

The Volume Centroid

The area centroid could be regarded as the plane analogue of a *Volume Centroid* defined as follows. Consider, as a scale model of the earth, a spherical shell with its interior filled with homogenous mass. On the surface of the sphere, a region of interest (say Australia) appears as a uniformly thin surface layer. Any great circle plane, which divides the region into equal areas, will also divide the volume of the earth equally. Two great circle planes, both of which divide the region into equal areas, will themselves intersect along a diameter of the sphere, which cuts the surface at a point. We call this point the Volume Centroid.

Is this point some sort of centre of gravity? Consider the spherical shell lying at rest, empty of all matter, with the region of interest (Australia) in contact with a frictionless level surface. The shell is in equilibrium and the direction of gravity through the *equilibrium point* (or contact point) will pass through the centre of the shell. As will an infinite number of planes, one of which will contain the great circle dividing the region in two. The Volume Centroid will lie somewhere along this great circle but not necessarily at the equilibrium point; they will only coincide when the region is symmetric about the point.

In Australia, a point called the geographical centre of Australia and named the *Lambert Centre* (in honour of the former Director of National Mapping, Dr. Bruce P Lambert) has been calculated by the Queensland Department of Geographic Information. The principle of computation seems to follow our definition for the Volume Centroid (DGI 1988, ISA 1994).

Average Centroids

In the *Geodetic Glossary* of the US National Geodetic Survey (NGS 1986), a centroid is defined as: "The point whose coordinates are the average values of the coordinates of all points of the figure."

This concept of the centroid of a figure as a point having *average* values of the coordinates (of the points defining the figure) encompasses three types of averages; the *mean*, the *median* and the *mode* (Reichmann 1961). All three are measures of central tendency.

The first type of average, the mean, can be subdivided into *arithmetic mean, root mean square mean, harmonic mean* and *geometric mean,* all of which can be defined by using two equations from Apostol (1967). The *pth-power mean* M_p of a set of real numbers $x_1, x_2, \ldots, x_n$ is

$$M_p = \left(\frac{x_1^p + x_2^p + \cdots + x_n^p}{n} \right)^{1/p} \tag{2}$$

where the number M_1 is the *arithmetic mean,* the number M_2 is the *root mean square* and M_{-1} is the *harmonic mean.* The *geometric mean G* of a set of real numbers $x_1, x_2, \ldots, x_n$ is

$$G = (x_1 x_2 \cdots x_n)^{1/n} \tag{3}$$

The second type of average, the median, is the central value, in terms of magnitude, of a set of *n* values ordered from smallest to largest. If *n* is odd, the median is the middle value and if *n* is even, it is the arithmetic mean of the middle pair of values. In both cases, there will be *k* values less than or equal to the median and *k* values greater than or equal to the median.

The third type of average, the mode, is the value that occurs most frequently in a set of numbers. It is used in statistics to indicate the value of a substantial part of a data set. As a measure of a central point of a geometric figure, the mode has little or no practical use. It will not be discussed further.

In addition, we can define a centroid as the arithmetic mean of the maximum and minimum values in a set of real numbers. This point, as we show below, is the centre of a rectangle that encloses the figure (the Minimum Bounding Rectangle).

Six different "average" centroids are therefore defined below.

The Arithmetic Mean Centroid

Using (2) with *p* = 1, we define the *Arithmetic Mean Centroid* $(\bar{x}, \bar{y})$ as

$$\bar{x} = \frac{\sum_{k=1}^{n} x_k}{n} \qquad \text{and} \qquad \bar{y} = \frac{\sum_{k=1}^{n} y_k}{n} \tag{4}$$

The Root Mean Square Centroid

Using (2) with *p* = 2, we define the *Root Mean Square Centroid* $(\bar{x}, \bar{y})$ as

$$\bar{x} = \sqrt{\frac{\sum_{k=1}^{n} x_k^2}{n}} \qquad \text{and} \qquad \bar{y} = \sqrt{\frac{\sum_{k=1}^{n} y_k^2}{n}} \tag{5}$$

The Harmonic Mean Centroid

Using (2) with $p = -1$, we define the *Harmonic Mean Centroid* $(\bar{x}, \bar{y})$ as

$$\frac{1}{\bar{x}} = \frac{\sum_{k=1}^{n} \frac{1}{x_k}}{n} \qquad \text{and} \qquad \frac{1}{\bar{y}} = \frac{\sum_{k=1}^{n} \frac{1}{y_k}}{n} \tag{6}$$

The Geometric Mean Centroid

Using (3) and taking natural logarithms to overcome numerical problems encountered with long products, we define the *Geometric Mean Centroid* $(\bar{x}, \bar{y})$ as

$$\ln \bar{x} = \frac{\sum_{k=1}^{n} \ln x_k}{n} \qquad \text{and} \qquad \ln \bar{y} = \frac{\sum_{k=1}^{n} \ln y_k}{n} \tag{7}$$

The Median Centroid

If the x and y coordinates of the n points defining the figure, are each ordered, from smallest to largest, into two arrays $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_n \end{bmatrix}$ then the *Median Centroid* $(\bar{x}, \bar{y})$ is

$$\text{for } n \text{ odd:} \qquad \bar{x} = x_k \qquad \text{and} \qquad \bar{y} = y_k \qquad \text{where } k = \frac{n+1}{2} \tag{8a}$$

and

$$\text{for } n \text{ even:} \qquad \bar{x} = \frac{x_k + x_{k+1}}{2} \qquad \text{and} \qquad \bar{y} = \frac{y_k + y_{k+1}}{2} \qquad \text{where } k = \frac{n}{2} \tag{8b}$$

The Minimum Bounding Rectangle Centroid

The *Minimum Bounding Rectangle Centroid* $(\bar{x}, \bar{y})$ is defined as

$$\bar{x} = \frac{x_{MIN} + x_{MAX}}{2} \qquad \text{and} \qquad \bar{y} = \frac{y_{MIN} + y_{MAX}}{2} \tag{9}$$

where x_{MIN}, y_{MIN} and x_{MAX}, y_{MAX} are the minimum and maximum values of the x and y coordinates respectively. This centroid will lie at the centre of a rectangle, whose sides are parallel with the coordinate axes, and which completely encloses the figure. The dimensions of the rectangle are $\text{width} = x_{MAX} - x_{MIN}$ and $\text{height} = y_{MAX} - y_{MIN}$.

The Minimum Distance Centroid

We define a *Minimum Distance Centroid* $(\bar{x}, \bar{y})$, as the point where the sum of the distances d_k from the centroid to every point defining the polygon is a minimum. That is, the minimum value of the function

$$f(\bar{x}, \bar{y}) = \sum_{k=1}^{n} d_k = \sum_{k=1}^{n} \sqrt{(x_k - \bar{x})^2 + (y_k - \bar{y})^2} \tag{10}$$

Minimum Distance centroids have a connection with spatial analysis where a *gravity function* may be used to model relationships between points or regions. Such functions, often arbitrarily defined, usually have an inverse relationship with distance (or distance squared); thus minimising distances maximises the gravity function. The earliest reference to such a gravity function (as a concept in studies of human interaction) is attributed to H.C.Carey, who reportedly stated in the early 1800's; "The greater the number collected in a given space, the greater is the attractive force that is there exerted … Gravitation is here, as everywhere, in the *direct* ratio of the mass, and the *inverse* one of distance." (Carrothers 1956, p.94).

The Negative Buffer Centroid

A polygon of n sides has an internal polygon of m sides ($m \leq n$) where each side is parallel to, but offset by a constant distance from, its complimentary side in the original polygon. This process of reduction of size (and shape) is known (in GIS parlance) as *negative buffering*. Positive buffering creates zones of constant width around polygons (Prescott 1995). By repeated negative buffering, a polygon of many sides can be reduced to a simpler polygon of far fewer sides. In many cases, negative buffering leads to a triangular polygon with an easily derived geometric centre. We define a *Negative Buffer Centroid* as the geometric centre of the polygon having the least number of sides resulting from repeated negative buffering of an original polygon.

The Circle Centroid

Closely allied to the Negative Buffer centroid, we define a *Circle Centroid* as the centre of the inscribed circle of a polygon. The Circle centroid can be obtained by using negative buffering to reduce the original polygon to a simpler internal polygon of (generally) far fewer sides. The centre of the inscribed circle of this simpler shape will also be the centre of the inscribed circle of the original polygon. In most cases, negative buffering will reduce an n-sided polygon to a triangle whose inscribed circle can be easily calculated. It should be noted that the Circle centroid (the centre of the inscribed circle) is not, in general, coincident with the Negative Buffer centroid.

CALCULATION OF CENTROIDS

To demonstrate some of the methods used to compute centroids, and to show the variation between different centroids, the simple polygon in Figure 1 will be used. The coordinates (metres) of the polygon corners are given in Table 1 and the area of the polygon is 2643.17 m^2.

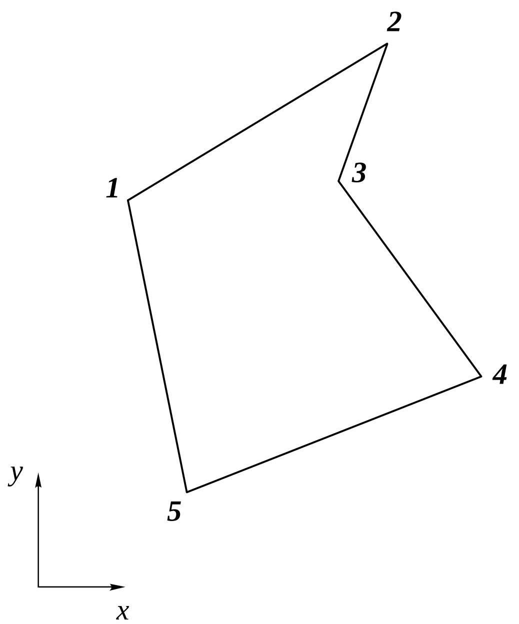

Figure 1

Point	x	y
1	103.450	287.760
2	151.860	315.990
3	141.030	289.410
4	167.830	255.920
5	114.130	235.840

Table 1

Since the calculation of some centroids requires areas of polygons, the following algorithm will be useful, noting that it yields positive areas proceeding clockwise around polygons and negative areas anticlockwise.

$$2(Area) = \sum_{k=1}^{n} \{x_k (y_{k-1} - y_{k+1})\} \tag{11}$$

The Moment Centroid of Figure 1

The Moment centroid is calculated from (1) where the moments M_y and M_x are found by regarding the polygon of area *A* as being composed of a network of triangles of area A_k each having a centroid $(\bar{x}_k, \bar{y}_k)$. We can then employ the rule for calculating moments of composite areas: the moment of a composite figure with respect to a line is the sum of the moments of the individual areas with respect to the line. With each triangle having moments $M_{y_k} = A_k \bar{x}_k$ and $M_{x_k} = A_k \bar{y}_k$ the centroid of the composite figure is given by

$$\bar{x} = \frac{M_y}{A} = \frac{M_{y_1} + M_{y_2} + \cdots + M_{y_n}}{A} = \frac{\sum_{k=1}^{n} A_k \bar{x}_k}{A} \quad \text{and similarly,} \quad \bar{y} = \frac{\sum_{k=1}^{n} A_k \bar{y}_k}{A} \tag{12}$$

Figure 2 shows the component triangles in the polygon and Table 2 shows their centroid coordinates and areas. Note that the centroid of a triangle is located at the average values of the *x* and y coordinates of the triangle vertices.

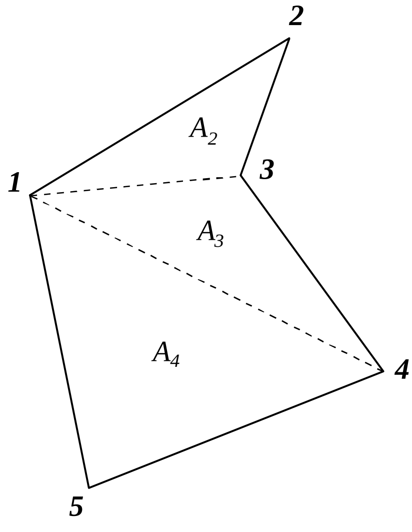

Figure 2

Triangle	$\bar{x}_k$	$\bar{y}_k$	Area A_k
A_2	132.1133	297.7200	490.5035 m^2
A_3	137.4367	277.6967	651.3871 m^2
A_4	128.4700	259.8400	1501.2792 m^2

Table 2

The Moment Centroid of the composite figure is

$$\bar{x} = \frac{\sum_{k=1}^{3} A_k \bar{x}_k}{A} = 131.356\,\text{m}$$

$$\bar{y} = \frac{\sum_{k=1}^{3} A_k \bar{y}_k}{A} = 271.270\,\text{m}$$

The values in Table 2 (component areas A_k and centroids $\bar{x}_k$, $\bar{y}_k$) have been computed from the following algorithms

$$A_k = \tfrac{1}{2}\left\{\Delta x_k \sum_{i=1}^{k} \Delta y_i - \Delta y_k \sum_{i=1}^{k} \Delta x_i\right\} \tag{13}$$

$$\bar{x}_k = x_1 + \tfrac{1}{3}\left\{2\sum_{i=1}^{k} \Delta x_i - \Delta x_k\right\} \quad \text{and} \quad \bar{y}_k = y_1 + \tfrac{1}{3}\left\{2\sum_{i=1}^{k} \Delta y_i - \Delta y_k\right\} \tag{14}$$

where Δx and Δy are coordinate differences of each side of the polygon and it should be noted that for a polygon having *n* sides, $A_1 = A_n = 0$. The algorithms (13) and (14) can be used to compute the area and centroid location of any polygon. The authors have not been able to find any references to these formulae, although it is unlikely that they are original; a derivation is given in Appendix A.

The Area Centroid of Figure 1

Area centroids of Figure 1, determined by the intersection of balance lines do not give a unique point. This is demonstrated by the two cases shown below. In the first case, Figure 3 shows the balance lines *B-B′* and *C-C′* in the cardinal directions. In the second case, Figure 4 shows another pair of balance lines *B-B′* and *C-C′*, still perpendicular to each other, but no longer in cardinal directions. The centroids for both cases differ by an appreciable amount.

Area Centroid: Case 1 (Cardinal directions)

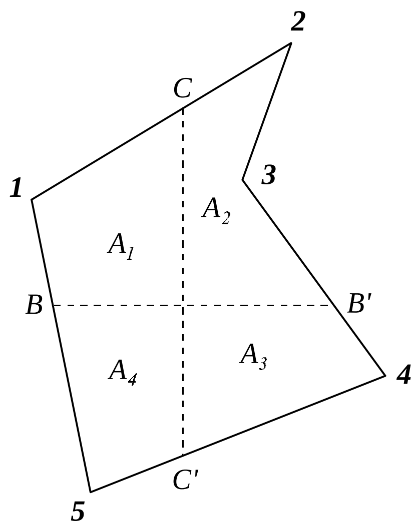

Figure 3

Point	x	y	Area
B	107.163	269.710	$A_1 = 685.28\,\mathrm{m}^2$
B'	156.795	269.710	$A_2 = 636.33\,\mathrm{m}^2$
C	131.006	303.829	$A_3 = 685.26\,\mathrm{m}^2$
C'	131.006	242.156	$A_4 = 636.32\,\mathrm{m}^2$

Table 3

The Area Centroid of the figure is

$\bar{x} = 131.006\,\mathrm{m}$

$\bar{y} = 269.710\,\mathrm{m}$

Area Centroid: Case 2 (Non-Cardinal directions)

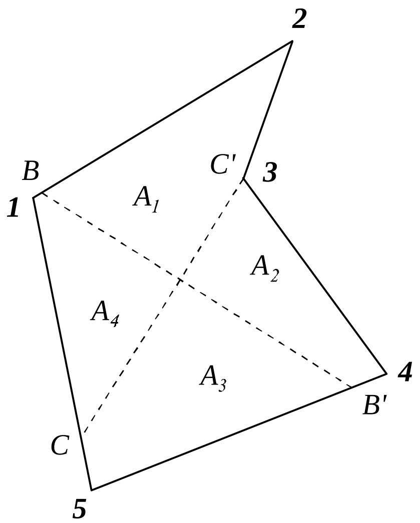

Figure 4

Point	x	y	Area
B	104.336	288.277	$A_1 = 793.47\,\mathrm{m}^2$
B'	160.327	253.114	$A_2 = 528.12\,\mathrm{m}^2$
C	112.467	243.926	$A_3 = 793.47\,\mathrm{m}^2$
C'	141.030	289.410	$A_4 = 528.13\,\mathrm{m}^2$

Table 4

The Area Centroid of the figure is

$\bar{x} = 131.142\,\mathrm{m}$

$\bar{y} = 272.071\,\mathrm{m}$

The "Average" Centroids of Figure 1

These centroids are calculated using equations (4) to (9) and tabulated below.

Average Centroid	$\bar{x}$	$\bar{y}$
Arithmetic Mean	135.660	276.984
Root Mean Square	137.728	278.399
Harmonic Mean	131.363	274.082
Geometric Mean	133.521	275.542
Median	141.030	287.760
Minimum Bounding Rectangle	135.640	275.915

Table 5. Average Centroids of Figure 1

It should be noted that all of these centroids, except the Minimum Bounding Rectangle, are functions of the number of points that make up the figure, not necessarily its shape. For example, in Figure 5 three points have been added, which divide the line *5* to *1* into four equal parts. The shape of the figure has not altered but the centroids will change substantially.

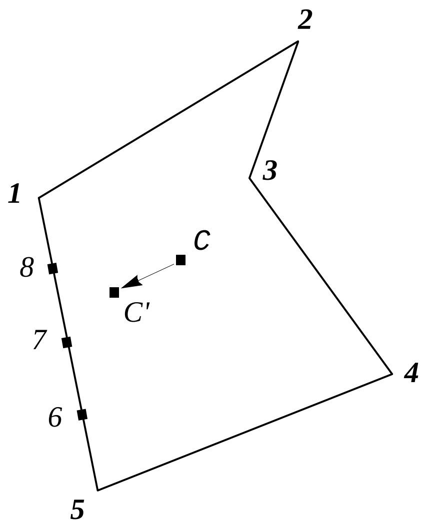

Figure 5

In Figure 5, *C* is the Arithmetic Mean Centroid of points *1* to *5* and *C'* is the Arithmetic Mean Centroid using the additional three points. This movement from *C* to *C'*, caused by the additional points along the line *1*-*5,* is representative of changes in location of the Root Mean Square, Harmonic Mean and the Geometric Mean Centroids.

The Minimum Distance Centroid of Figure 1

The Minimum Distance Centroid of Figure 1 is obtained by minimising (10). This cannot be done by simple arithmetic but instead requires sophisticated function minimisation techniques. Microsoft's Excel ™ solver has been used to obtain the minimum value of function (10) as

$$\bar{x} = 141.019\,\text{m}$$
$$\bar{y} = 289.397\,\text{m}$$

The solution that the Excel solver obtains may be described in the following way. Imagine a fine mesh grid placed over the figure with every grid intersection a computation point with coordinates

x_j, y_k. From each computation point the distance to every point in the figure is computed and summed, yielding a single value, say $z_{j,k}$. As *j* and *k* vary from 1 to *n* a grid of *z* values will be obtained and a three dimensional plot would reveal a surface with a low point. The *x, y* coordinates of this low point will be the Minimum Distance Centroid of the figure.

The Negative Buffer Centroid and Circle Centroid of Figure 1

Figure 6 shows the original five-sided polygon of Figure 1 and two internal polygons, *ABCDE* of five sides and *PQR* of three. The internal polygons have been created by negative buffering; ie zones of constant width inside a polygon define a smaller polygon. *ABCDE* is defined by a 10 metre wide zone within the original figure and *PQR* is defined by a 10-metre buffer zone within *ABCDE.* The Negative Buffer Centroid in this case is the geometric centre (and centroid) of the triangle *PQR.*

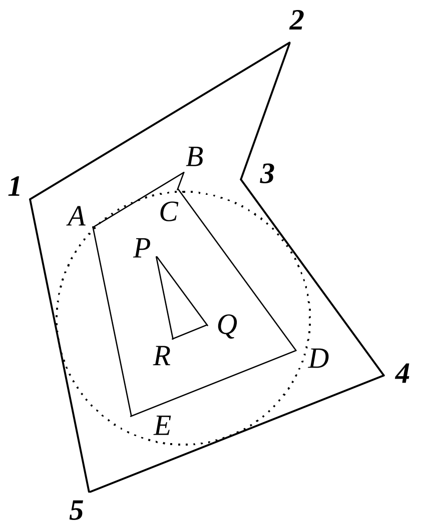

Figure 6

Point	x	y	Point	x	y
A	114.692	282.740	*P*	126.337	275.761
B	131.510	292.547	*Q*	134.962	264.982
C	129.554	287.746	*R*	129.012	262.757
D	151.396	260.451			
E	121.571	249.299			

Table 6. Coordinates of polygons *ABCDE* and *PQR*

The Negative Buffer Centroid of the figure is

$\overline{x} = 130.104\,\text{m}$

$\overline{y} = 267.833\,\text{m}$

The Circle Centroid is the centre of the inscribed circle of triangle *PQR.* The radius *r* of the inscribed circle is given by $r = \sqrt{s(s-a)(s-b)(s-c)} \big/ s$ where *a, b* and *c* are the sides of the triangle and $s = (a+b+c)/2 =$ is the semi-perimeter (Spiegal 1986). Now, since sides *PQ, QR* and *RP* are parallel to sides *3-4, 4-5* and *5-1* respectively, then $r + 20$ will be the radius of the largest circle that can be drawn inside the original polygon. This inscribed circle is shown in Figure 6, its radius is 22.492 metres and its centre is located at $x_c = 130.867$ m and $y_c = 266.111$ m. Note that the Circle Centroid and the Negative Buffer Centroid are not the same point.

COMMENTARY ON CENTROID DEFINITIONS

The Arithmetic Mean, Root Mean Square, Harmonic Mean, Geometric Mean and Median Centroids are all legitimate measures of spatial central tendency. While they have the benefit of being easily calculated, we have shown that they are a function of the vertices of a polygon but are not sensitive to the order of the vertices and therefore the shape of the polygon. They are, therefore considered to be useful descriptors of point data sets but not of polygons. Similarly, while the Minimum Bounding Rectangle Centroid is easy to calculate and is representative of the four extreme vertices of the polygon, it is not sensitive to the entire shape.

The Area Centroid suffers from not providing a unique result for a given polygon unless the balance lines are constrained to particular directions, such as for the Minimum Bounding Rectangle where the cardinal directions are assumed.

The Negative Buffer and Circle Centroids are not sensitive to large poorly conditioned portions of a polygon (best described as spikes). The computation of these centroids is difficult to automate, particularly to deal with highly irregular polygons.

The Moment and Minimum Distance Centroids both provide a logical and intuitively appealing result, but the latter requires sophisticated function minimisation software for calculation. This computational drawback, and deficiencies mentioned above of the other centroids, leads the authors to prefer the Moment Centroid as the best measure of the centre of a complex polygon.

GIS CENTROID FUNCTIONS

Determinations of a polygon centroid is a standard requirement in Geographic Information System (GIS) software, primarily for providing a representative point within a polygon. This point provides a focus for the positioning of attribute labels. Because, in general, a centroid cannot be guaranteed to fall within the polygon to which it relates, the representative point is often a *paracentroid*, this satisfies this criterion at the expense of centrality. The definition of centroid and method of calculation adopted by a particular GIS product is rarely documented for the user.

Three commonly used GIS packages were tested with the simple polygon in Figure 1 to determine the likely definition of centroid which they adopt. These packages are *MapInfo* Professional V5.5 (MapInfo Corporation), *ARCView* V3.1 (ESRI) and *ARCInfo* V7.2.1 (ESRI).

The *MapInfo centroidX* and *centroidY* functions produce results consistent with the Minimum Bounding Rectangle (MBR) definition of centroid.

According to the *ARCView* help documentation, the *ReturnCenter* method also returns the MBR Centroid, unless it falls outside the polygon, in which case it is moved in the x direction the minimum amount to place it within the polygon. In practice, *ReturnCenter* was found to return the MBR Centroid regardless of whether it fell within the polygon or not.

ARCInfo uses the *createlabels* command to create a paracentroid within a polygon. The documentation clearly states that this point will not necessarily be at the centroid (without defining centroid). Once paracentroids (label points) have been created, they can be moved to the centroid using the *centroidlabels* command. Although the definition of centroid adopted is not documented, the point produced by *centroidlabels* is consistent with the Moment Centroid.

CENTROIDS OF VICTORIA

To calculate the various centroids of Victoria a digital outline of the state was derived from a coordinate data set known as VIC500-2000, which is part of the L500 library of Natural Resources and Environment (NRE) Corporate Geospatial Data Library. The VIC500-2000 data set are VICMAP_TM coordinates which are related to a grid superimposed over a Transverse Mercator projection of latitudes and longitudes on the Australian Geodetic Datum 1966 (AGD66). The projection has a central meridian of 145° 00' 00" with a scale factor of unity. The north and east axes of the grid are parallel to the central meridian and equator respectively and the origin of coordinates is 500,000 metres west and 10,000,000 metres south of the intersection of the central meridian and the equator. The GIS software product *ARCInfo* was used to transform the VIC500-2000 coordinates to AGD66 latitudes and longitudes (ϕ, λ). The route analysis function of *ARCInfo* was then used to generate a subset of 1648 consecutive points at 2 km intervals around the

Victorian borders and coastline. The ϕ, λ coordinates of the 1648 points were then transformed to east and north (*E*,*N*) coordinates (kilometres) related to an Equal Area Cylindrical projection of a sphere of radius *R* using the following formulae

$$\begin{aligned} X &= R(\lambda - \lambda_0) \\ Y &= R\sin\phi \end{aligned} \tag{15}$$

where $R = 6372.800$ km, λ_0 = 145° 00' 00" and

$$\begin{aligned} E &= X + 5{,}000 \text{ km} \\ N &= Y + 5{,}000 \text{ km} \end{aligned} \tag{16}$$

An Equal Area projection was chosen in this study because two of the centroids require the computation of areas of plane figures and this projection preserves area scale. That is, an element of area on the spherical Earth *dA* is projected as an element of area on the map *da* and the ratio *da/dA* = 1.

The positional accuracy of the original data (VICMAP_TM) is stated in the NRE documentation as ±0.5 km and the derived digital outline (1648 points) does not include any offshore islands and is a relatively coarse approximation of Victoria's boundaries. No tests were conducted on the accuracy of the *ARCInfo* transformation formulae (VICMAP_TM to ellipsoid) and the transformation to *E*,*N* coordinates assumes the points are located on a sphere of radius equal to the mean radius of curvature of the AGD ellipsoid appropriate for the latitude of Victoria. All centroid values for Victoria, computed from the derived digital outline, are rounded to the nearest 0.5 km (*E*,*N*). These values are then regarded as exact for the purposes of transformation to latitudes and longitudes, using inverse relationships obtained from (15 and (16), which are then rounded to the nearest 15 seconds of arc. These results are shown in Table 9.

Centroid	*E*	*N*	**Latitude**	**Longitude**
Moment	4921.0	1180.0	-36° 49' 45"	144° 17' 30"
Area (Cardinal directions)	4874.5	1168.5	-36° 57' 30"	143° 52' 15"
Arithmetic Mean	4972.0	1170.0	-36° 56' 30"	144° 45' 00"
Root Mean Square	4981.0	1178.0	-36° 51' 00"	144° 49' 45"
Harmonic Mean	4954.0	1154.5	-37° 07' 00"	144° 35' 15"
Geometric Mean	4963.0	1162.0	-37° 02' 00"	144° 40' 00"
Median	4975.5	1128.5	-37° 24' 30"	144° 46' 45"
Minimum Bounding Rectangle	5051.5	1207.5	-36° 31' 15"	145° 27' 45"
Minimum Distance	4815.5	1122.0	-37° 29' 00"	143° 20' 30"
Negative Buffer	4711.5	1205.5	-36° 32' 30"	142° 24' 30"
Circle (*r* = 160.1 km)	4711.5	1206.5	-36° 31' 00"	142° 24' 30"

Table 9 Coordinates of Centroids

Figure 7 shows an Equal Area Cylindrical projection of Victoria and parts of New South Wales and South Australia, and three centroids (i) the Negative Buffer Centroid, (ii) the Minimum Bounding Rectangle (MBR) Centroid and (iii) the Moment Centroid. The Negative Buffer Centroid is at the centroid of a triangle, the result of successive negative buffering of the digital outline of Victoria. The Circle Centroid (the centre of the inscribed circle of this triangle) is also the centre of the largest circle (radius 160.1 km) that can be inscribed within this projection of Victoria. The Circle Centroid and Negative Buffer Centroid differ by only 1 km but the size and location of the inscribed circle is dependent on the type of projection. The MBR Centroid (the simplest to compute) and the Negative Buffer Centroid (the most difficult to compute) appear to be evenly balanced about the Moment Centroid, located at Mandurang, approximately 6 km south-east of Bendigo, a major regional city.

Figure 8 shows an enlargement of the region near Bendigo with the major roads and regional towns. The Moment Centroid is shown south of Bendigo with the other centroids (Area, Minimum Distance and the Average centroids), dispersed across the region. The map covers an area of approximately 180 km by 90 km.

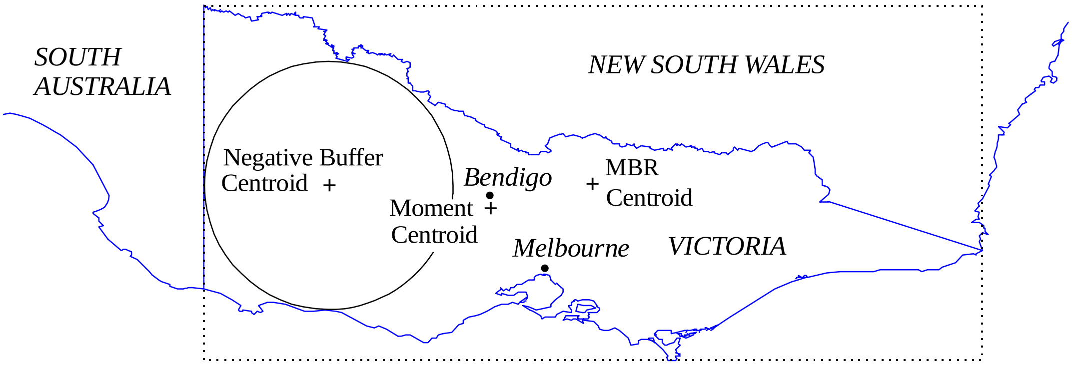

Figure 7
Negative Buffer Centroid, Minimum Bounding Rectangle (MBR) Centroid and Moment Centroid

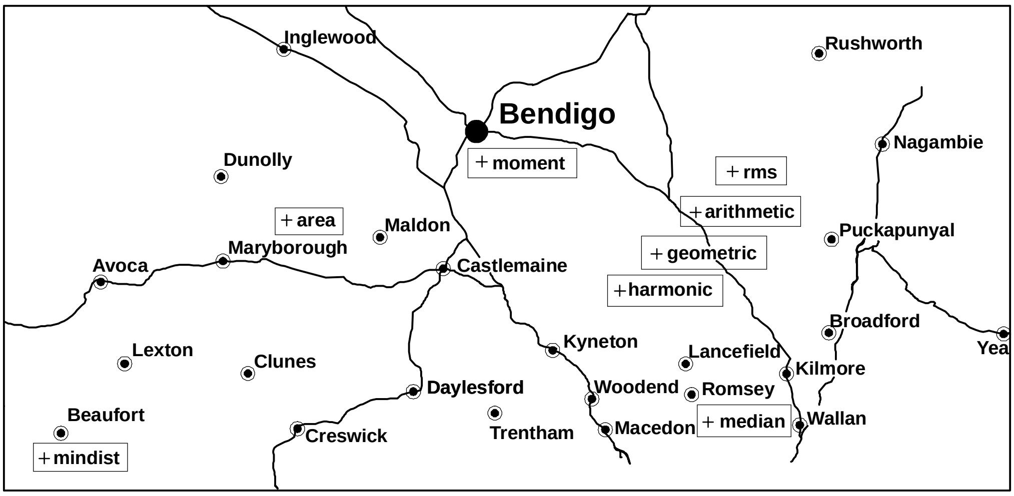

Figure 8
The Moment, E-W Area and Minimum Distance Centroids
and the five Average Centroids

CENTROID OF MELBOURNE

Melbourne, the capital of Victoria, is a large urban metropolis with a population of approximately 3.5 million. In May, 2001, officers of the Department of Natural Resources and Environment, using Victoria's digital cadastral database (VicMap Digital) created a digital outline of Urban Melbourne. This polygon, consisting of 5733 points (containing an area of 1718 km^2) defines a boundary between urban Melbourne and surrounding rural areas. Its delineation depended on criteria such as the extent of drainage and sewerage schemes, electricity and water supply schemes, housing density and land allotment areas. Its complex shape describes Melbourne's urban sprawl, which is generally in an easterly and south-easterly direction from the central business district and in places is 35-40 km from Melbourne central.

The data set was provided in Australian Map Grid 1966 (AMG66) coordinates. These values were converted to AGD66 values (ϕ, λ) and then to east and north (*E*,*N*) Cartesian coordinates (kilometres) related to an equal area cylindrical projection of a sphere of radius *R* using formulae (15) and (16) and the associated parameters

The Moment Centroid for Melbourne computed from the derived *E*,*N* coordinates is shown in Table 10 rounded to the nearest 0.5 km. The latitude and longitude values are computed by using the inverse relationships obtained from (15) and (16), which are then rounded to the nearest 15 seconds of arc.

Centroid	*E* (km)	*N* (km)	**Latitude**	**Longitude**
Moment	5007.5	1089.5	-37° 51' 00"	145° 04' 30"

Table 10 Coordinates of Moment Centroid of Melbourne

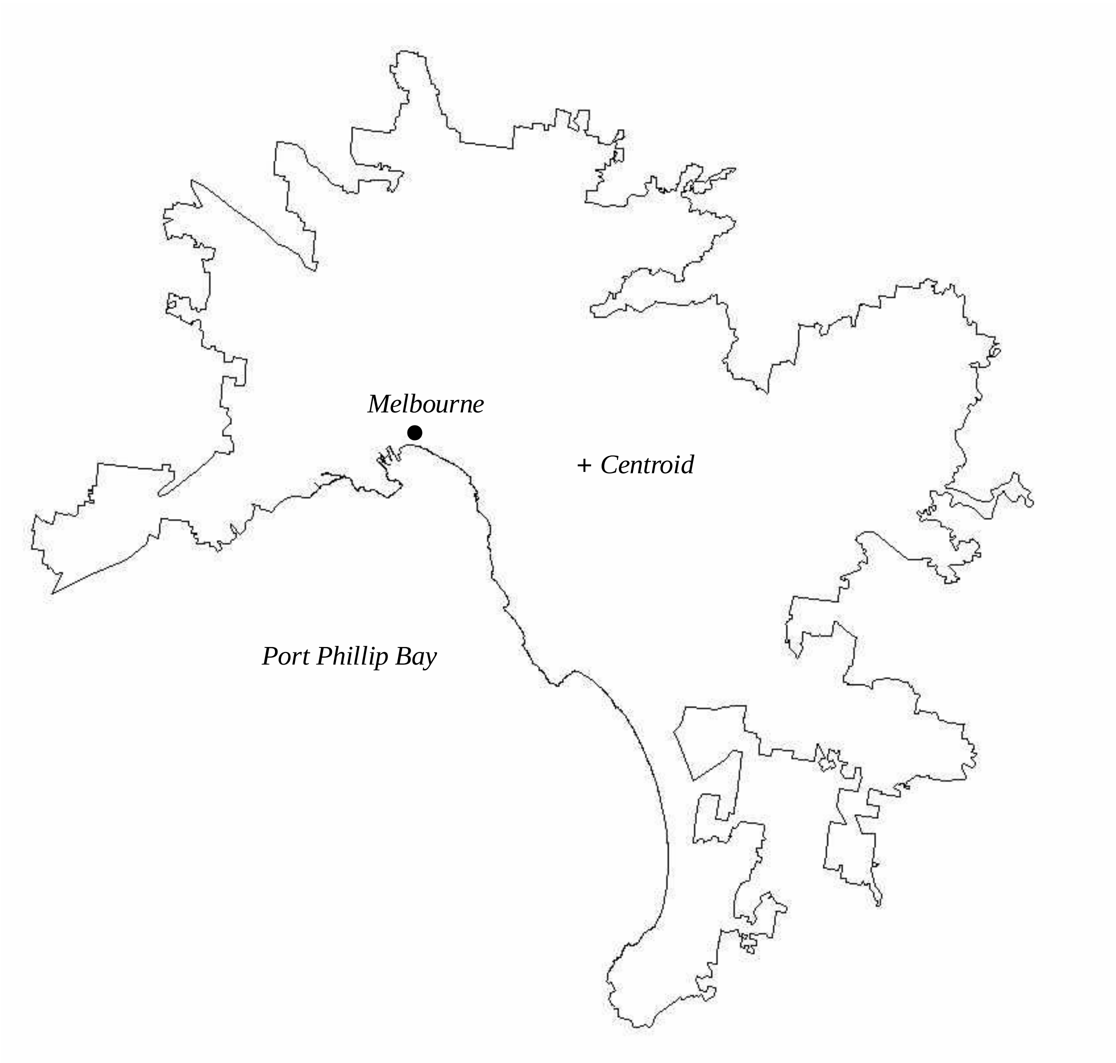

Figure 9
Moment Centroid of Urban Melbourne

CONCLUSION

This paper has provided ten working definitions for the centroid of a plane polygon. Others could no doubt be constructed, but in the opinion of the authors, the Moment Centroid, which is intuitively appealing, is the most appropriate.

In the context of requiring a centroid for the traditional cartographic purpose of labelling a polygon, consideration of how the centroid is defined is largely academic. This is particularly so when the computed point may have to be relocated to ensure it is within the polygon object or for cartographic license in producing a pleasing output.

In the context of defining the 'official' centre of an administrative or physical region, the outcome is far from purely academic interest. The work associated with this paper in determining the centroids of Victoria and Melbourne attracted the attention of the popular media in at least one national television program, two statewide newspapers, a regional newspaper, two radio interviews and a government department publication. In the case of the (Moment) centroid of Victoria at Mandurang south of Bendigo, the local council and tourism authority are considering the erection of a substantial monument, signposting and other promotional material. The owner of the property near which the centroid falls is considering how best to capitalise on the windfall, with a Bed and Breakfast accommodation facility likely. In the case of the centroid of Melbourne (located at Ferndale Park in the suburb of Glen Iris), a plaque has been placed and its dedication was attended by a Government Minister, Departmental Heads, local government officials and the media. Clearly, the centroid of a State or a Capital city has a high curiosity value and potential as a tourist attraction providing significant economic benefit and civic pride.

In the case of the Victoria, the ten centroids canvassed, span a region 275 km East-West by 100 km North-South and are included in seven different municipalities. Had *Map Info* or *ARCView* been used to calculate the centroid without consideration of the definition they implement, the location would be nearly 100 km from the position provided by *ARCInfo*. If you were a representative of one of those municipalities, where would you like the centroid to be?

REFERENCES

Apostol, Tom M., 1967. *Calculus,* Vol.1, 2nd edn, Blaisdell Publishing Co., London.

Ayres, F., 1968. *Differential and Integral Calculus,* Shaum's Outline Series, McGraw-Hill Book Company, New York.

Carrothers, G.A.P., 1956. 'An historical review of the gravity and potential concepts of human interaction.' *Journal of the American Institute of Planners,* Spring 1956, pp.94-102.

DGI (Department of Geographic Information), 1988. 'The Centre of Australia', in: From the DGI, *Queensland Surveyors Bulletin,* Vol.4, August 1988, pp.32-3.

ISA (Institution of Surveyors, Australia), 1994. 'The Lambert Centre', *The Australian Surveyor,* Vol.39, No.3, September 1994, p.215.

NGS (National Geodetic Survey), 1986. *Geodetic Glossary,* United States National Geodetic Survey, Rockville, Maryland, USA.

OGC, 1999. The OpenGIS™ Abstract Specifications, Topic 1: Feature Geometry, Version 4. OpenGIS™ Project Document Number 99-101.doc, Open GIS Consortium, http://www.opengis.org

Prescott, George W., 1995. *A Practioner's Guide to GIS Terminology: A Glossary of Geographic Information System Terms,* Data West Research Agency, Washington, USA.

Reichmann, W.J., 1961. *Use and Abuse of Statistics,* Penguin Books, Middlesex, England.

SOED, 1993. *The New Shorter Oxford English Dictionary,* ed Lesley Brown, Clarendon Press, Oxford.

Spiegel, M.R., 1968. *Mathematical Handbook of Formulas and Tables,* Schaum's Outline Series, McGraw-Hill Book Company, New York.

SoV (State of Victoria), 1982. *Atlas of Victoria,* ed J.S. Duncan, Victorian Government Printing Office, Melbourne, Victoria, Australia.

APPENDIX A

The algorithm for computing the area of a polygon, given as equation (13), can be derived by considering Figure A1, where the area is the sum of the trapeziums *bBCc*, *cCDd* and *dDEe* less the triangles *bBA* and *AEe*.

The area can be expressed as

$$\begin{aligned} 2A = & \left[(x_2 - x_1) + (x_3 - x_1)\right]\left[(y_2 - y_3)\right] \\ & + \left[(x_3 - x_1) + (x_4 - x_1)\right]\left[(y_3 - y_4)\right] \\ & + \left[(x_4 - x_1) + (x_5 - x_1)\right]\left[(y_4 - y_5)\right] \\ & - (x_2 - x_1)(y_2 - y_1) \\ & - (x_5 - x_1)(y_1 - y_5) \end{aligned} \qquad \text{(A1)}$$

Expanding (A1) then cancelling and re-arranging terms gives

$$\begin{aligned} 2A = & \, x_1(y_5 - y_2) \\ & + x_2(y_1 - y_3) \\ & + x_3(y_2 - y_4) \\ & + x_4(y_3 - y_5) \\ & + x_5(y_4 - y_1) \end{aligned}$$

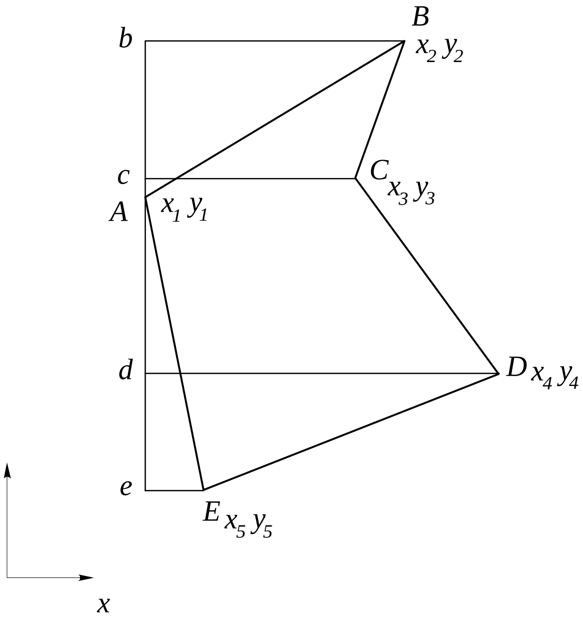

Figure A1

which can be expressed as $2A = \sum_{k=1}^{n} \left\{ x_k (y_{k-1} - y_{k+1}) \right\}$ (A2)

In Figure A2, the coordinate origin is shifted to *A* where $x_1' = y_1' = 0$ and the area, using (A2), is

$$2A = y_2' x_3' + y_3' x_4' - y_3' x_2' + y_4' x_5' - y_4' x_3' - y_5' x_4' \qquad \text{(A3)}$$

Considering each side of the polygon to have components $\Delta x_k, \Delta y_k$ for $k = 1$ to 5, (A3) can be written as

$$\begin{aligned} 2A = & \, \Delta y_1 (\Delta x_1 + \Delta x_2) \\ & + (\Delta y_1 + \Delta y_2)(\Delta x_1 + \Delta x_2 + \Delta x_3) \\ & - (\Delta y_1 + \Delta y_2)(\Delta x_1) \\ & + (\Delta y_1 + \Delta y_2 + \Delta y_3)(\Delta x_1 + \Delta x_2 + \Delta x_3 + \Delta x_4) \\ & - (\Delta y_1 + \Delta y_2 + \Delta y_3)(\Delta x_1 + \Delta x_2) \\ & - (\Delta y_1 + \Delta y_2 + \Delta y_3 + \Delta y_4)(\Delta x_1 + \Delta x_2 + \Delta x_3) \end{aligned}$$

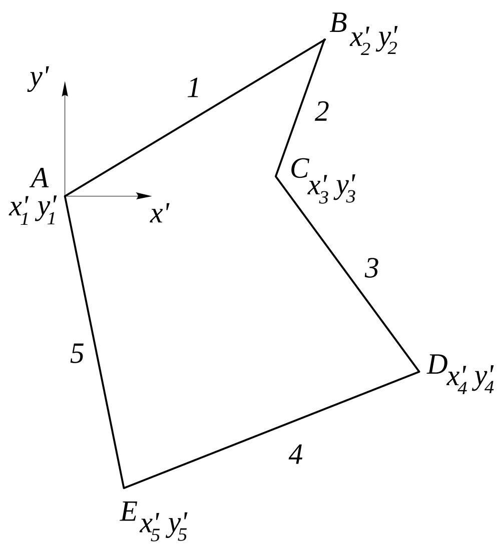

Figure A2

Expanding and gathering terms gives

$$
\begin{aligned}
2A = &\Delta y_1 \left(3\Delta x_1 + 3\Delta x_2 + 2\Delta x_3 + \Delta x_4\right) && -\Delta y_1 \left(3\Delta x_1 + 2\Delta x_2 + \Delta x_3\right) \\
&+\Delta y_2 \left(2\Delta x_1 + 2\Delta x_2 + 2\Delta x_3 + \Delta x_4\right) && -\Delta y_2 \left(3\Delta x_1 + 2\Delta x_2 + \Delta x_3\right) \\
&+\Delta y_3 \left(\Delta x_1 + \Delta x_2 + \Delta x_3 + \Delta x_4\right) && -\Delta y_3 \left(2\Delta x_1 + 2\Delta x_2 + \Delta x_3\right) \\
& && -\Delta y_4 \left(\Delta x_1 + \Delta x_2 + \Delta x_3\right)
\end{aligned}
$$

and cancelling terms and re-ordering gives

$$
\begin{aligned}
2A = &\Delta y_1 \left(0 + \Delta x_2 + \Delta x_3 + \Delta x_4\right) \\
&+\Delta y_2 \left(-\Delta x_1 + 0 + \Delta x_3 + \Delta x_4\right) \\
&+\Delta y_3 \left(-\Delta x_1 - \Delta x_2 + 0 + \Delta x_4\right) \\
&+\Delta y_4 \left(-\Delta x_1 - \Delta x_2 - \Delta x_3 + 0\right)
\end{aligned} \tag{A4}
$$

This equation for the area can also be expressed as a matrix equation

$$
2A = \begin{bmatrix} \Delta y_1 & \Delta y_2 & \Delta y_3 & \Delta y_4 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta x_3 \\ \Delta x_4 \end{bmatrix} \tag{A5}
$$

By studying the form of equations (A4) and (A5), the following algorithm for calculating the $k = n-1$ area components A_k for a polygon of *n* sides may be deduced as

$$
A_k = \tfrac{1}{2}\left\{ \Delta x_k \sum_{i=1}^{k} \Delta y_i - \Delta y_k \sum_{i=1}^{k} \Delta x_i \right\} \quad \text{where} \quad k = 1,\, 2,\, 3,\, \ldots\, n-1 \tag{A6}
$$

Equation (A6) is an efficient way to accumulate the area of a polygon given the coordinate components of the sides. By studying the algorithm, it can be seen that $A_1 = A_n = 0$ and hence the area of a polygon is accumulated without having to deal with the last side. In addition, it can be seen that each area component A_k is a triangle with one vertex at the starting point and the line *k*, with components Δx_k, Δy_k, the opposite side. This leads to the formulae for the centroids of each triangle

$$
\bar{x}_k = x_1 + \tfrac{1}{3}\left\{ 2\sum_{i=1}^{k} \Delta x_i - \Delta x_k \right\} \qquad \text{and} \qquad \bar{y}_k = y_1 + \tfrac{1}{3}\left\{ 2\sum_{i=1}^{k} \Delta y_i - \Delta y_k \right\} \tag{A7}
$$