The Role of Coordinate Systems, Coordinates and Heights in Horizontal Datum Transformations

WILL FEATHERSTONE
School of Spatial Sciences,
Curtin University of Technology,
GPO Box U1987, Perth, 6845, Western Australia.

PETR VANICEK
Department of Geodesy and Geomatics Engineering,
The University of New Brunswick,
PO Box 4400, Fredericton, E3B 5A3, Canada.

This paper is a textual version of a presentation made by the second author to the Western Australian Divisions of the Institution of Surveyors and Mapping Sciences Institute and to the University of New South Wales in 1998.

Abstract
This paper reviews the fundamental definitions of geodetic and geocentric coordinate systems, whilst clarifying the distinction between coordinates and coordinate systems. It is then argued that the transformation of coordinates from a local geodetic datum to a geocentric datum should first employ a change of the coordinate system using a six- or four-parameter transformation, followed by further modelling of the distortion in the coordinates. It is also argued that the horizontal coordinate transformation should not include height information, since this forms an entirely different coordinate in another coordinate system.

1. Introduction

As is now well known, Australia will implement the Geocentric Datum of Australia (GDA94) in the year 2000. Several papers have already been published that explain the rationale and technical arguments behind this change (eg. Featherstone, 1996; Steed, 1995; Inter-governmental Committee on Surveying and Mapping, 1994 and 1997) and these will not be duplicated here. The primary consequence of the introduction of the GDA94 is that existing coordinates, related to the Australian Geodetic Datum (which has two realisations: AGD66 and AGD84), may have to be transformed to the GDA94. Of course, there remains the option of a user or organisation retaining the AGD66 or AGD84. Nevertheless, this decision will also require the transformation of GDA94 coordinates to these datums.

In 1997, the Australian Surveying and Land Information Group (AUSLIG) released several new sets of transformation parameters with which to transform coordinates from the AGD66 and AGD84 to the GDA94. These parameters apply to several common mathematical models for datum transformations (eg. Featherstone, 1997). Of the conformal datum transformations, the seven-parameter transformation is endorsed, since it offers the highest accuracy. This transformation has been used for a number of years in Australia to transform AGD coordinates to the World Geodetic System 1984 (WGS84) in order to provide approximate initial coordinates for GPS baseline processing. Parameters for the transformation between AGD84 and WGS84 have been calculated by Higgins (1987) and between AGD66 and WGS84 in New South Wales by the Land Information Centre, for example.

In this paper, the appropriateness of the seven-parameter model for the transformation of coordinates between horizontal geodetic datums will be challenged. Instead, it will be proposed that, for rigour and conceptual clarity, a two-stage process should be used to achieve the transformation of these coordinates. The first stage changes the *coordinate system* and the second changes the *coordinates* by modelling distortions that occur mostly because of the practical realisation of the terrestrial geodetic datum. It will also be argued that heights should not be used in the transformation of horizontal coordinates, since these belong to a completely different coordinate system. Indeed, heights are not actually needed in the transformation of horizontal coordinates.

2. Basic Definitions and Terminology

Coordinates and coordinate systems are ubiquitous in virtually all aspects of surveying and mapping. Importantly, they allow the users of spatial data to easily conceptualise coordinates (ie. positions) with respect to some convenient coordinate system. Indeed, there are very few applications that can exist without the use of coordinates and coordinate systems. As such, both coordinates and coordinate systems have had, and will continue to have, an essential role to play in the spatial sciences.

A coordinate system is essentially an abstract idea that forms the ‘scaffolding’ for the representation of positions and features, and is purely independent of the geometrical objects described by them. As a simple example, a two-dimensional polygon feature can be held in a geographical information

system (GIS) in terms of geographical coordinates, map-grid coordinates or coordinates given in an arbitrary coordinate system as specified by the creator of the GIS.

There is the following distinction that can be made between coordinate systems and coordinates. A coordinate system can be chosen somewhat arbitrarily, as exemplified by the above. Importantly, a coordinate system forms a common frame of reference for the description of positions. On the other hand, coordinates are simply an ordered set of numbers that are used to describe the positions of points or features in a coordinate system. Accordingly, the terms coordinates and positions can be used interchangeably, but always refer to a specific coordinate system.

2.1 Cartesian and curvilinear coordinate systems

In geodesy, two common classes of coordinate system have been used to describe positions in relation to the Earth. These comprise the geodetic coordinate system and the Cartesian coordinate system. Historically, geodetic coordinates have been used since these are conceptually more appropriate for describing positions on or near the Earth's curved surface. However, Cartesian coordinates have taken an increasing role because of the widespread use of GPS and other satellite-based positioning systems. The geometrical relationship between these two coordinate systems is shown in Figure 1.

Figure 1. The geometrical relationship between the geodetic coordinate system (ϕ, λ, h) and the Cartesian coordinate system (X, Y, Z).

The geodetic coordinate system uses a triplet of orthogonal, curvilinear coordinates of geodetic latitude (ϕ), geodetic longitude (λ) and geodetic (or ellipsoidal) height (h). These coordinates refer to the surface of a specific ellipsoid of revolution about its minor axis. Accordingly, they are dependent upon the size, shape and three-dimensional orientation of the ellipsoid. Examples of such ellipsoids are the Australian National Spheroid or ANS (National Mapping Council, 1986), the World Geodetic System 1984 (WGS84) ellipsoid (Defense Mapping Agency, 1991) and the Geodetic Reference System 1980 (GRS80) ellipsoid (Moritz, 1980).

The geodetic coordinates are defined as follows:

- The geodetic latitude is the angle, reckoned in the meridional plane, between the ellipsoidal surface normal at the point of interest and the equatorial plane of the ellipsoid ($-90° \le \phi \le +90°$).
- The geodetic longitude is the angle, reckoned in the equatorial plane, from the Greenwich meridian to the meridian through the point of interest ($0° \le \lambda \le 180°$E and $-180°$W $\le \lambda \le 0°$).
- The geodetic height is the distance above the surface of the ellipsoid, reckoned positive away from the ellipsoid along the ellipsoidal surface normal.

The horizontal coordinate components in this coordinate system (ie. geodetic latitude and geodetic longitude) have commonly been used since they are a conceptually more convenient coordinates for describing points on or near the Earth's surface. Also, if the ellipsoid is oriented in such a way that it is a best fit to the geoid in a particular region, astronomical measurements can be assumed to be coincident with their geodetic counterparts for some applications. This is because the deflections of the vertical from the ellipsoidal normal are small enough to be neglected for some applications. However, the geodetic height is not used since it has no physical meaning. Instead, heights based on mean sea level, and thus in the Earth's gravity field, are used for practical purposes.

The three-dimensional Cartesian coordinate system comprises three orthogonal axes in the X, Y and Z directions (Figure 1). A corresponding triplet of Cartesian coordinates refers to these axes. The Cartesian coordinate system is named after René Descartes, a French mathematician in the 17th century, so is spelt as a proper noun. When applied to the Earth:

- The X axis is directed towards the intersection of the Greenwich meridian and equatorial plane.
- The Z axis is aligned towards the north pole of rotation.
- The Y axis is orthogonal to the X and Z axes and completes the right-handed coordinate system. That is, if the X axis is rotated towards the Y axis, a right-handed screw would be propelled along the Z axis.

For any geodetic coordinate system (ϕ, λ, h), there exists a representative three-dimensional Cartesian coordinate system (X, Y, Z). This is shown in Figure 1, where the Cartesian coordinate system is aligned in such a way that its axes are coincident with the major and minor axes of the ellipsoid; the third is, by definition, orthogonal. The origin of the representative Cartesian coordinate system (X=0, Y=0, Z=0) is coincident with the geometrical centre of the ellipsoid, where the minor and major axes intersect.

The relationship between coordinates in these two coordinate systems (ie. the coordinate transformation) can be derived using Euclidean geometry. Accordingly, the conversion from geodetic coordinates to Cartesian coordinates is given by:

$$X = (v + h) \cos \phi \cos \lambda \qquad (1)$$

$$Y = (v + h) \cos\phi \sin\lambda \quad (2)$$

$$Z = (v(1 - e^2) + h) \sin\phi \quad (3)$$

with

$$v = a\,(1 - e^2 \sin^2\phi)^{-1/2} \quad (4)$$

where a is the semi-major axis length of the ellipsoid and e is its first numerical eccentricity, which are effectively the coordinate transformation parameters. Together, these parameters define the size and shape of the reference ellipsoid and are integral to the definition of the geodetic coordinate system. In many cases, the flattening (f) of the ellipsoid is given instead of the first numerical eccentricity; these are related according to $e^2 = 2f - f^2$

The transformation from Cartesian to geodetic coordinates is a little more involved because iteration is usually required to solve for the geodetic latitude. Several authors have proposed efficient iterative methods (eg. Bowring, 1985). Alternatively, closed formulae, which require no iteration, have also been proposed (eg. Paul, 1973). However, the relative merits of these conversions will not be discussed here. Instead, the simple rearrangement of equations (1), (2) and (3) yields:

$$\lambda = \tan^{-1} Y/X \quad (5)$$

$$\phi = \tan^{-1} [(Z+e^2 v\sin\phi)/(X^2+Y^2)^{-1/2}] \quad (6)$$

$$h = (X^2+Y^2)/\cos\phi - v \quad (7)$$

which also requires the transformation parameters of the appropriate ellipsoid. Therefore, geodetic coordinates can easily be transformed to their corresponding Cartesian coordinates, and *vice versa*.

2.2 Geocentric and geodetic coordinate systems

In geodesy, two different classes of the above coordinate systems have been adopted. These comprise the geocentric and geodetic coordinate systems, whose origins and axes of the associated Cartesian coordinate system are different.

- The geocentric coordinate system has its origin at or close to the geocentre. The letter C will be used to denote this coordinate system. Examples of C-systems are the conventional terrestrial coordinate system, such as the International Earth Rotation Service's Terrestrial Reference Frame (ITRF), and the instantaneous terrestrial coordinate system, which are both geocentric but differ in their orientation due to the effects of polar motion.

- The geodetic coordinate system does not necessarily have its origin at or close to the geocentre. In what follows, G will be used to denote this coordinate system. An example of the G-system is the Australian Geodetic Datum, whose origin is offset from the current best estimate of the geocentre by approximately 200 metres (eg. Mather, 1970).

Associated with the C and G coordinate systems is the definition and use of a datum. Accordingly, there can be a geodetic horizontal datum or geocentric horizontal datum. The datum uses coordinates referred to the surface of some appropriately defined ellipsoid of revolution.

As stated, the ellipsoid for a local geodetic datum is usually chosen so as to give as best fit to the geoid in the region of interest. This simplifies the subsequent reduction of survey data to the datum. That is, certain corrections are made sufficiently small that they can be neglected in routine surveys. Therefore, integral to the definition of any datum, two additional parameters, typically the semi-major axis and flattening, are specified in its definition. Conceptually, the datum is a reference surface to which coordinates are referred.

Using an ellipsoid in a G-system (ie. h=0) yields a 'horizontal geodetic datum'. For instance, the Australian Geodetic Datum uses the Australian National Spheroid (National Mapping Council, 1986). On the other hand, using an ellipsoid in a C-system (ie. h=0) yields a 'horizontal geocentric datum'. Therefore, the Geocentric Datum of Australia uses the Geodetic Reference System 1980 (GRS80) ellipsoid (eg. Featherstone, 1996).

Given the above definitions, coordinates are simply ordered, numerical values that describe positions in a given coordinate system. According to the origin of the coordinate system, these coordinates can either be geocentric (C-system) or geodetic (G-system). It is also instructive to make the distinction between a coordinate system and a geodetic network. A geodetic network is simply a geometrical configuration of geodetic measurements between ground points that are independent of the coordinate system chosen. Therefore, a geodetic network can be used to produce positions in a C-system or G-system. Also, the coordinates used can be Cartesian (X, Y, Z) or geodetic (ϕ, λ, h).

3. Practical Realisation of a Coordinate System

The idea of a coordinate system must be linked to physical reality using an appropriate mathematical model. Therefore, in order to provide the infrastructure for positioning on or close to the Earth's surface, a convenient coordinate system must be realised. This consists of:

- the adoption of specific idea;

- the adoption of a position in the coordinate system (ie. its origin) with respect to the Earth;

- the adoption of an orientation of the coordinate system (ie. the directions of the coordinate axes) with respect to the Earth;

- a prescription of how to determine coordinates (positions) in the realised coordinate system and to relate these to the prescribed origin point.

There are two specific examples of the above that are of interest in geodesy. These are:

1. A classical, direct realisation of a G-system via the origin of a geodetic network. This system is essentially topocentric, with its origin being on the topography. For

example, the realisation of the AGD uses the ANS plus the origin point at Johnson Geodetic Station (ϕ_0, λ_0 and h_0). In order to orient the G-system in space, three other parameters are needed. These are the two deflections of the vertical and the geodetic azimuth at the origin station. Each of these was effectively defined for the Johnson origin for the AGD (Bomford, 1967).

2. A modern, indirect realisation of a C-system via a set of positions determined from geodetic satellites and other space-based techniques. For example, the realisation of the GDA94 uses the GRS80 ellipsoid (Moritz, 1980) and the eight points comprising the Australian Fiducial Network (AFN). The origin of the GDA94 is assumed to be close to the geocentre, since the satellites used in the establishment of the AFN are assumed to orbit about the geocentre. However, the geocentre is known to vary with time. The orientation of the GDA is implied by the positions of the AFN with respect to the Earth's mean spin axis over a specific period of time.

It is important to note that the idea of a coordinate system (whether it be Cartesian, geodetic or any other) by itself is quite useless for positioning. Instead, the surveyor must rely upon the practical realisation of the coordinate systems in order to carry out positioning.

4. Advantages and Disadvantages of Geocentric and Geodetic Coordinate Systems

Two advantages of a geocentric coordinate system (and thus a geocentric datum) are:

- The horizontal coordinates of points from different localities on the Earth can be treated as being referred (approximately) to the same coordinate system. This, most of the time, obviates the necessity of transforming coordinates from one system into another when using a multitude of these positions. This has positive consequences in applications such as national boundary demarcation and intercontinental ballistic missile guidance. It can be argued that military issues have driven the need for a single global datum. For instance, in World War 2, the multitude of different datums (and map projections) was particularly problematic in Europe.

- Coordinates and coordinate differences determined from satellite positioning systems (eg. the Global Positioning System or GPS) can usually be treated as being referred directly to the geocentric coordinate system. Importantly, this requires less work because the GPS-derived coordinates are directly compatible with the horizontal geodetic datum and coordinate transformations are not required.

The principal disadvantage of a geocentric coordinate system is that:

- It will always be only approximately geocentric. Modern, indirect realisations of geocentric coordinate systems may be subject to changes (re-definitions) when the coordinates used for the positioning and orientation of the geocentric coordinate system are re-observed and re-computed. A notable example is the case of the North American Datum 1983 (NAD83), which was originally thought to be geocentric, but is now known to be offset by a few metres. Thus, a geocentric coordinate system should either be considered to be really non-geocentric and adopted as such by a convention, or be considered as truly geocentric, and thus the subject of gradual 'improvement', and thus change. Importantly, 'the rules of the game' should be decided ahead of adopting a geocentric datum.

The advantage of a geodetic coordinate system is that:

- It is inherently immutable. This system was the order of the day before the advent of satellite positioning systems. Its realisation was direct and every G-system was really fixed and oriented with respect to the origin of the network for good. Such a system is thus quite transparent to the user, since it is understood that the system relates to something that has been optimised for a particular country. For instance, the origin of the coordinate system can be physically seen.

The disadvantage of a geodetic coordinate system is that:

- It is quite different from continent to continent (empire to empire, country to country, municipality to municipality). Thus, dealing with positions of points in different G-systems (on different geodetic datums) requires a knowledge of transformation equations and parameters, which is generally a fairly complicated matter. This is reasonably well understood by most professionals working with these data, but is not understood by the majority of lay users.

5. Terrestrial- and Satellite-derived Geodetic Networks

Only horizontal positions are considered in what follows, because horizontal terrestrial networks have been developed in a differential manner from the origin of the network. The horizontal positions have been obtained from a least-squares adjustment, often a piece-wise adjustment, of observations made by terrestrial geodetic instruments: distances, horizontal angles and/or directions, azimuths and other auxiliary observations.

An important fact that should be borne in mind is that the points belonging to horizontal geodetic networks have their horizontal positions known as accurately as it was possible to determine them, while their heights are known only approximately, or not at all. Height networks are entirely different entities, which have been designed and constructed using entirely different principles. Therefore, they should not be mixed with horizontal networks. Though heights are required in the establishment of a horizontal geodetic datum, to reduce observations from the surface of the Earth to the ellipsoid, they do not form part of the horizontal geodetic datum. For instance, the Australian Height Datum (AHD) is separate from the AGD and GDA. It was established in 1971, was independent of the AGD66 and AGD84, and will not be affected by the introduction of the GDA94 and *vice versa*.

Satellite geodetic networks are configurations of points whose coordinates have been determined by satellite positioning systems. These positions are inherently three-dimensional, and are normally computed in the Cartesian coordinate system. These positions can be readily transformed to the horizontal geodetic coordinates, latitude and longitude on the geocentric horizontal datum, using equations (1) through (7). Of note, the horizontal coordinates comprising satellite geodetic networks are generally known to a much better accuracy than their terrestrial counterparts, especially over long distances. For instance, Savage *et al*., (1996) show a systematic difference may exist between GPS and terrestrially derived distances. This is most likely to be due to scale-dependent errors in the terrestrial instrumentation. This alone provides a rationale for adopting satellite-derived coordinate systems.

When terrestrially and satellite-derived positions are to be considered side by side, the tendency has been to convert the terrestrial positions to three-dimensional positions and combine the networks in three dimensions. Indeed, this approach has been taken in Australia. This practice is considered dangerous because the heights of points belonging to terrestrial networks are clearly of very different provenance compared to their horizontal positions, as well as compared to the heights of points belonging to satellite networks. Therefore, it is recommended that the combination of these geodetic networks in the definition and realisation of horizontal datums remains a horizontal process (eg. Vanicek and Steeves, 1996).

6. Horizontal Network Distortions

As professional providers of position, it is important for surveyors to always track errors in positions. Essentially, a position should be considered as unreliable unless it has an associated estimate of its accuracy. When dealing with geodetic networks, one should be aware of the fact that the individual positions are subject to both systematic and random errors. The random errors are characterised by statistical techniques, whereas the systematic errors have to be modelled using deterministic techniques.

The systematic errors are particularly dangerous, amounting easily to tens of metres. When applied to geodetic networks, these systematic errors are generally called 'distortions', so as to distinguish them from their random counterpart. It is these distortions that complicate the combination of terrestrial and satellite positions. This is principally because one wishes to preserve the existing investment in positions and, at the same time, to take the advantages of the modern satellite positioning capabilities. Importantly, the Australian Geodetic Datum appears to contain distortions (eg. Collier *et al*., 1998).

Distortions are much worse in the case of terrestrial geodetic networks, so much so that, in comparison, the distortions of satellite networks can be often disregarded altogether. The distortions in terrestrial geodetic networks come mostly from the past practice of approximate observation reductions (eg. ignoring the geoid-ellipsoid separation or deflection of the vertical) and from approximate adjustment procedures. In Australia, the State/Territory surveying and mapping agencies are now making a concerted effort to model these distortions and produce a mathematical expression for predicting the distortion vector $\boldsymbol{D}$ as a function of position (ϕ, λ). Note that the distortion $\boldsymbol{D}(\phi,\lambda)$ is really a pair of real numbers; a horizontal vector that is applied to the horizontal position of point (ϕ, λ) to obtain its more correct horizontal position. The random errors in the modelled distortions can and should also be estimated. In Australia, these distortion models will be provided as coordinate grids of distortion vectors, from which the user can interpolate the appropriate distortion to be applied at each point.

Applying the distortion modelling process can improve the accuracy of the transformed coordinates with respect to the new coordinate system. This is because the simple conformal transformations both carry and 'smear' the systematic errors in the original coordinates into the new coordinates. As such, the full benefit of a new datum will not be realised when using this approach. Therefore, it is recommended that distortion modelling is included in the transformation process.

7. Transformations between Coordinate Systems

Now that the concepts of and distinctions between coordinate systems and coordinates have been given, together with the general notions of distortions in terrestrial geodetic networks, it is now possible to discuss the transformation of coordinate systems and the coordinates as realised via geodetic networks.

7.1 Transformation of coordinate systems

The transformation between coordinate systems, as distinct from coordinates, consists of three translations (related to the origin positions of the two systems) and three rotations (related to the alignment of the two systems). These six quantities correspond to the six degrees of freedom of any rigid body, which a three-dimensional coordinate system is. It is therefore essential to realise that a scale difference (ie. the seventh parameter in the seven-parameter transformation) has no role to play in the transformations between coordinate systems.

Any scale differences, as well as the network distortions discussed above, only come into consideration when transforming the (distorted) coordinates. Coordinate systems can never be considered as distorted, even coordinate systems that have been realised (positioned and oriented) indirectly via coordinates. The principal reason why coordinate systems must be considered undistorted is to keep their position and orientation immutable.

On the other hand, it has been common practice, both in Australia and elsewhere, to transform coordinates based on seven transformation parameters, which comprise the six degrees of freedom and an additional scale change. This scale change inherently includes a component for the distortion in the coordinates. When taking this approach, the concept of the transformation between coordinates and coordinate systems becomes blurred. Comparing the scale factor for Higgins's and AUSLIG's transformation parameters shows a notable example of this. The scale factor 'changes' from +0.0983ppm to –0.191ppm. When

interpreted alone and applied across the continent, this implies that there has been a change in distortion of around one metre. Therefore, it is argued that the transformation of coordinate systems and the transformation of the distorted coordinates, usually in the terrestrial geodetic system, are treated separately.

The basic transformation to be dealt with here is that between the practical realisation of a G-system and the practical realisation of a C-system. For the sake of clarity, this can be reduced to the transformation between their respective, representative Cartesian coordinate systems using equations (1) to (7). Importantly, the transformation of the coordinate system consists of three translations to coincide the origins and three rotations to align the three axes. Using matrix-vector notation, the transformation of coordinate systems is achieved via

$$\mathbf{r}^G = \mathbf{R}(rx, ry, rz)\,\mathbf{r}^C - \mathbf{t}^C \qquad (8)$$

where $\mathbf{r}^G = (X^G, Y^G, Z^G)^T$ is the position vector in the G-system and $\mathbf{r}^C = (X^C, Y^C, Z^C)^T$ is the position vector in the C-system, the superscript T denotes the transpose, **R** is the rotation matrix, (*rx, ry, rz*) are the angles of the misalignment of the three axes between each system, and $\mathbf{t}^C$ is the translation vector between the origins of the two systems. Essentially, $\mathbf{t}^C$ is the position vector of the origin of the G-system in the C-system, and it follows that $\mathbf{t}^C = -\mathbf{t}^G$.

For small misalignment angles, the rotation matrix can be linearised to leave a three-by-three matrix with a single rotation term in each element. Using this approximation, equation (8) is rearranged to give

$$\mathbf{R}(rx, ry, rz)\,\mathbf{r}^C = \mathbf{r}^C + \mathbf{w} \times \mathbf{r}^C \qquad (9)$$

where $\times$ is the vector cross product, $\mathbf{w} = (rx, ry, rz)^T$ is the transpose of the misalignment vector, which can be written as

$$\mathbf{w} = w\,(\cos\phi_m \cos\lambda_m, \cos\phi_m \sin\lambda_m, \sin\phi_m)^T \qquad (10)$$

where w is the magnitude of the misalignment and (ϕ_m, λ_m) are the horizontal geodetic coordinates of the misalignment axis (Vanicek and Wells, 1974). If the G-system has been realised in the classical way (via the local astronomical coordinate system at the topocentric origin of network), then $(\phi_m, \lambda_m) = (\phi_o, \lambda_o)$. That is, the misalignment axis is at the origin of the geodetic network. In this case, only one misalignment angle, $w = w_o$, appears in the transformation equation (9) and the six transformation parameters are reduced to only four.

Because of the nature of the initial orientation of any G-system, its misalignment with respect to a C-system is constrained: this misalignment rotation can only take place around the ellipsoidal normal at the origin of the geodetic network. In the unlikely event when the G-system had been oriented some other way (rather than *vis-a-vis* the local astronomic system at the origin of the network) then three unconstrained rotations would have to be used. Therefore, transformations should consider involving either four or six transformation parameters.

7.2 Determination of coordinate system transformation parameters

The six or four transformation parameters cannot be determined from the observations collected for the original positioning and orientation of the G-system. Instead, they have to be determined from the positions of a common set of points known in both coordinate systems (ie. G and C). However, this approach unavoidably includes all the random errors and distortions that affect both these positions. Accordingly, these propagate into the derived transformation parameters. Therefore, an effort must be made to model and eliminate as much of the distortions in the coordinates as possible before the coordinates are used for the transformation parameter estimation. If the distortions have not yet been reliably determined, the distortion parameters $\boldsymbol{D}(\phi,\lambda)$ can be solved for together with the four or six unknown transformation parameters. Note that these transformation parameters are immutable in time and in space. The random errors in these transformation parameters should also be estimated.

The following derivations can be used to estimate the four or six transformation parameters, assuming that the distortions have already been modelled and removed. Firstly, equation (8) is rewritten as

$$\mathbf{r}^G = \mathbf{r}^C + \mathbf{w} \times \mathbf{r}^C - \mathbf{t}^C \qquad (11)$$

or

$$\mathbf{r}_i^G - \mathbf{r}_i^C = \mathbf{w} \times \mathbf{r}_i^C - \mathbf{t}^C \qquad (12)$$

for the generalised use of $i = 1,\ldots n$ points

From the commutative law of matrix multiplication

$$\mathbf{w} \times \mathbf{r}_i^C = -\mathbf{r}_i^C \times \mathbf{w} = -\mathbf{Q}_i\,\mathbf{w} \qquad (13)$$

where

$$\mathbf{Q_i} = \begin{pmatrix} 0, & -Z_i, & Y_i \\ Z_i, & 0, & -X_i \\ -Y_i, & X_i, & 0 \end{pmatrix}$$

which is a three-by-three matrix containing the positions of each point i in the C-system, and

$$\forall i=1,\ldots n: \qquad \mathbf{r}_i^G - \mathbf{r}_i^C = -\mathbf{Q}_i\,\mathbf{w} - \mathbf{t}^C \qquad (14)$$

or

$$\forall i=1,\ldots n: \qquad \Delta\mathbf{r}_i = \mathbf{A}_i\,\mathbf{x} \qquad (15)$$

where $\mathbf{A}_i = [-\mathbf{Q}_i\,, \mathbf{I}]$ is the design matrix and $\mathbf{x} = [\mathbf{w}, \mathbf{t}^C]^T$ is the vector of the unknown parameters to be estimated.

Equation (15) yields a system of $3n$ linear observation equations for four or six unknowns, depending upon the conditions described earlier. If the variance-covariance matrix ($\mathbf{C}_{\Delta r} = \mathbf{C}_{rG} + \mathbf{C}_{rC}$) is known, then the least-squares estimate of the transformation parameters is

$$\mathbf{x} = (\mathbf{A}^T\,\mathbf{C}_{\Delta r}^{-1}\,\mathbf{A})^{-1}\;\mathbf{A}^T\,\mathbf{C}_{\Delta r}^{-1}\,\Delta\mathbf{r} \qquad (16)$$

and the (random) error estimate in these transformation parameters is

$$\mathbf{C}_\mathbf{x} = (\mathbf{A}^T\,\mathbf{C}_{\Delta r}^{-1}\,\mathbf{A})^{-1} \qquad (17)$$

7.3 The role of heights in the horizontal datum transformation

In computing $\Delta\mathbf{r}_i$ above, the ellipsoidal heights h_i are assumed to have been used. However, for the reasons outlined earlier, this should not be done! Therefore, the following alternative is proposed.

$$\forall i=1,\ldots n:\ \Delta\mathbf{r}_i = (\Delta X_i, \Delta Y_i, \Delta Z_i)^{\mathrm{T}} = \mathbf{J}(\Delta\phi_i, \Delta\lambda_i, \Delta h_i)^{\mathrm{T}} \quad (18)$$

where **J** is the Jacobian and this represents an exact relationship. Substituting this into equation (15) gives

$$\forall i=1,n:\quad \mathbf{J}(\Delta\phi_i, \Delta\lambda_i, \Delta h_i)^{\mathrm{T}} = \mathbf{A}_i\,\mathbf{x} \quad (19)$$

or

$$\forall i=1,n:\quad \Delta\mathbf{r}_i = \mathbf{J}_i^{-1}\,\mathbf{A}_i\,\mathbf{x} = \mathbf{A}^*_i\,\mathbf{x} \quad (20)$$

The least-squares solution to equation (20) is

$$\mathbf{x} = (\mathbf{A}^{*\mathrm{T}}\,\mathbf{C}_{\Delta r}^{-1}\,\mathbf{A}^*)^{-1}\ \mathbf{A}^{*\mathrm{T}}\,\mathbf{C}_{\Delta r}^{-1}\,\Delta\mathbf{r} \quad (21)$$

with a (random) error estimate of

$$\mathbf{C}_{\mathbf{x}} = (\mathbf{A}^{*\mathrm{T}}\,\mathbf{C}_{\Delta r}^{-1}\,\mathbf{A}^*)^{-1} \quad (22)$$

When heights are omitted from the transformation equations, as it is argued they should be, and the common points are taken to be on the ellipsoid.

8. Transformation of Coordinates

Often surveyors are required to use coordinates determined in one network, in one coordinate system (on one datum) in the context of coordinates belonging to the other network, in the other coordinate system (on the other datum). Consider the following example: a GPS-determined position of a parcel corner is to be reconciled with the rest of parcel's boundary that is defined by points whose positions are known on the AGD, and have been derived from a terrestrial network. A transformation of the coordinate systems has to be used first to get coordinates in the desired coordinate system, then the transformed coordinates have to be distorted to fit into the fabric of existing points.

8.1 The often-used solution

Some surveying and mapping agencies put distortion models together with transformation parameters into one transformation package, notably the seven-parameter model. In such an approach, the transformation parameters often masquerade as varying with position, which must look very odd to anyone who understands the issue correctly. Of course, this coordinate transformation package can be used to transform coordinates from one datum to another and, at the same time, to distort them (albeit to a limited extent) to fit the fabric of points on the other datum (whether one wants to do this or not!).

Why this is done is not clear. What is clear, however, is that keeping the two concepts of coordinate systems and distortions in coordinates (and thus the two sets of parameters) separate, gives the user more flexibility, more insight into the working of the transformations and more appreciation for the individual error contributions. On the other hand, it requires the user to be a bit more sophisticated and knowledgeable; is this the only reason why the simplistic package is opted for?

9. Summary

Based on the difference between coordinate systems and the coordinates realised by geodetic or geocentric datums, it has been argued that the transformation should follow two stages:

1. Transformation of the coordinate systems,
2. Modelling of the distortions between the coordinates.

In addition, it has been argued that heights be excluded from a horizontal coordinate transformation, since these comprise an entirely different coordinate system.

*ACKNOWLEDGEMENT: This presentation was given whilst the second author was a recipient of a **C.Y. O'Connor Fellowship** from the Division of Engineering and Science at Curtin University of Technology. We would also like to thank the reviewer for his time taken to offer suggested improvements to this manuscript.*

References

Bowring, B.R. (1985) The accuracy of geodetic latitude and height equations, *Survey Review*, 28(218): 202-206

Bomford, A.G. (1967) The geodetic adjustment of Australia, 1963-66, *Survey Review*, 144: 52-71.

Collier P.A., Argeseanu V.S. and Leahy F.J. (1998) Distortion modelling and the transition to the GDA94. *The Australian Surveyor*, 43(1): 29-40.

Defense Mapping Agency (1991) Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems (second edition). Technical Report no. 8350.2, Defense Mapping Agency, Washington.

Featherstone, W.E. (1996) An updated explanation of the geocentric datum of Australia and its effect upon mapping, *The Australian Surveyor*, 42(2): 30-40.

Featherstone, W.E. (1997) An evaluation of existing coordinate transformation models and parameters in Australia, *Cartography*, 26(1): 13–26.

Higgins, M.B. (1987) Transformation from WGS84 to AGD84, an interim solution, *Internal Report*, Department of Geographic Information, The University of Queensland.

ICSM (1994) A new era for Australia. *Information Circular*, Inter-governmental Committee on Surveying and Mapping, Belconnen, ACT.

ICSM (1997) Where do you stand with GDA? *Information Circular*, Inter-governmental Committee on Surveying and Mapping, Belconnen, ACT.

Mather, R.S. (1970) The geocentric orientation vector of the Australian Geodetic Datum, Geophysical Journal of the Royal Astronomical Society, 22: 395-405.

Moritz, H. (1980) Geodetic Reference System 1980. *Bulletin Géodésique*, no. 54, pp. 395-405.

National Mapping Council (1986) *The Australian Geodetic Datum Technical Manual*, National Mapping Council of Australia, Belconnen, ACT.

Paul, M.K. (1973) A note on computation of geodetic coordinates from geocentric (Cartesian) coordinates. *Bulletin Géodésique*, 108: 135-139.

Savage, J.C., Lisowsky, M. and Prescott, W.H. (1996) Observed discrepancy between Geodolite and GPS distance measurements. *Journal of Geophysical Research*, 101(B11): 25547-25552.

Steed, J. (1995) The geocentric datum of Australia. *Surveying World*, 4(1); 14-17.
Vanicek, P. and Steeves, R.R. (1996) Transformation of coordinates between two horizontal geodetic datums, *Journal of Geodesy*, 70(11): 740-745.
Vanicek, P. and Wells, D.E. (1974) Positioning of horizontal datums, *The Canadian Surveyor*, 28(5): 531-538.