Demystifying Reference Systems

A Chronicle of Spatial Reference Systems in Canada

Don Junkins
Gordon Garrard

Geodetic Survey Division
Geomatics Canada
615 Booth Street
Ottawa, Ontario
Canada
K1A 0E9

information@geod.nrcan.gc.ca
http://www.geod.nrcan.gc.ca

Abstract

Geodesists like to consider themselves as scientists or mathematicians, but are more often thought of as trouble makers when they produce new coordinate values for the control survey networks. How many horizontal reference systems again? The North American Datum of 1927 (NAD27) and its derivatives, MAY76 and Compensation Géodésique du Québec 1977 (CGQ77). The Average Terrestrial System 1977 (ATS77), the North American Datum of 1983 (NAD83), the World Geodetic System 1984 (WGS84), and the many variations of the International Terrestrial Reference Frame (ITRF). There are reasons for these changes even though cartographers, hydrographers, land surveyors and GIS users would like to have a stable reference frame with coordinate values that will not change. So geodesists have both good and unfortunately bad news. The bad news is that there will continue to be adjustments and modifications to coordinates in the future, but the good news is that absolute change is diminishing and users will be able to determine if the changes are significant or not. The presentation will briefly explain the recognized horizontal reference frames in Canada and why Geodetic Survey Division (GSD) of Geomatics Canada, Natural Resources Canada (NRCan) and its partners in the provinces and United States will continue to improve them.

Introduction

With the advent of GPS-based technologies and the proliferation of low cost equipment, it is now easier than ever for users to determine their position on the surface of the Earth. To be useful, these positions must relate through some sort of a network connection to a defined reference system. However, because of changing technologies over recent decades there are several recognized definitions currently in use. This paper will provide a brief outline of the common definitions established by the Geodetic Survey Division and its partners in Canada, and how they are derived and related. By understanding how positions are related to reference systems, users can make better informed decisions about the options available to them.

This paper is intended for those who are involved in positioning and geo-referencing for GIS, mapping, and navigation, and are not geodetic experts but need to understand some of the jargon they hear from geodetic control agencies. The detail presented is a basic overview of concepts and background information, and is not a definitive document for geodesists. More complicated and related issues such as vertical datums, inertial reference frames, gravity fields, and crustal motion will only be briefly mentioned.

Concepts and Terminology

We all think we know what datums and reference systems are, but why are there so many in Canada? Where do they all come from, how are they related, and why would we possibly need another one? To answer these questions, we first need to review some of the concepts and terminology that provide a common ground for discussion of the matter.

The notions of datums, reference systems, and coordinates are all intertwined, and we often use the terms interchangeably, causing them to lose their distinction. To help clarify the terms, let's review them in the context of the processes we go through to establish coordinates for physical points on the ground. Then we can relate the various sets of coordinates to these processes, and see the patterns of relationships among them. The explanations of the terms we will give here may not be rigorous or comprehensive definitions, but they will serve the purpose for the context of this discussion.

As with all coordinates, we must first have a **coordinate system**, which is a set of rules that specify how the coordinates are to be assigned to points. The most familiar example is the two-dimensional rectangular Cartesian coordinate system for plotting on a plane sheet of paper.

Once the location of the origin and the orientation of the axes have been specified, our coordinate system is embellished and becomes a **reference coordinate system**. We do this on our piece of paper by drawing the two axes at right angles, thus defining the origin of the system as the point of intersection, and the orientation as the directions of the axes.

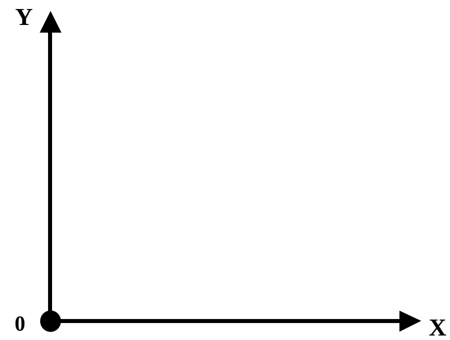

Figure 1: 2D Rectangular Coordinate System

Such a system may work well for surveying in a local area, such as for engineering surveys. However, once surveyors extend their scope, the curvature of the earth's surface becomes a significant factor which must be taken into account. Since the Earth is a three-dimensional body, any point on, above, or below its surface may easily be assigned coordinates in a three-dimensional rectangular Cartesian coordinate system. When the origin of such a system is specified as the centre of mass of the Earth (i.e. **geocentric**), and its axes aligned such that the Z-axis is parallel to the Earth's axis of rotation, and the positive X-axis intersects the Greenwich meridian, it becomes the **Conventional Terrestrial Reference System (CTRS)**. It is also often referred to as the Conventional Terrestrial System (CTS). The complete specification of a **spatial reference system** also includes such items as physical constants, conventions and other required details.

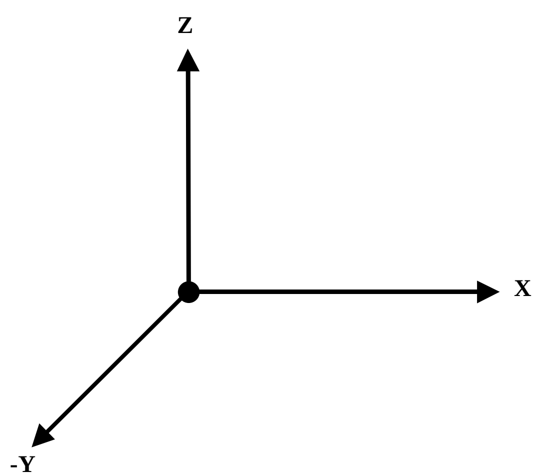

Figure 2: 3D Conventional Terrestrial System

If we now add to the CTS an ellipsoid of revolution to approximate the earth's surface, we have the conventional concept of a **geodetic datum**. The ellipsoid is really a two-dimensional surface in three-dimensional space. Thus, the traditional notion of a geodetic datum is for horizontal positioning only. The vertical component is traditionally referred to mean sea level, and will not be discussed in detail in our deliberations here.

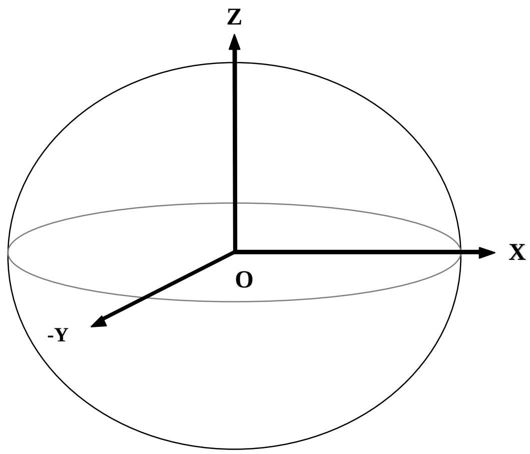

Figure 3: Geodetic Horizontal Datum

So far, the systems we have considered are conceptual in nature. To make them real, we must somehow assign numerical values for the coordinates of tangible points. This is usually done for a few points at which very comprehensive and accurate observations are made. Because they are often quite expensive to establish and maintain, the specialized procedures are used for only a very small number of points. These become the **fiducial** or **datum points** to which all other surveys may be connected and referred. By determining the their coordinates according to the principles and specifications of the CTS, and assigning values for calibration facilities, we complete the requirements for a **spatial reference frame**.

When we densify and extend this reference frame with integrated control survey networks of monumented points, we achieve our goal of fully realizing the spatial reference system. It is in this form that the user can take advantage of it for all types of geo-referencing activities, such as mapping, navigation, and GIS development. The objective of a spatial reference frame and the control survey networks is to provide the means for integrating a variety of products into a single system by facilitating reliable data compatibility.

But as we mentioned at the outset, in Canada there seems to be a confusing number of official datums or spatial reference frames, so much so that the notion of a uniform standard is obscured and we are left with a Tower of Babel preventing efficient communication and exchange of data. Let's take a look at what they are, and by detailing some basic patterns and commonalities among them, some of the mystery will be cleared away.

NAD83 - the System from 1986 until 1998

Let's start with **NAD83**, which is the spatial reference system currently adopted for most jurisdictions and applications in Canada, as is the case in the United States. When the spatial reference frame was established in 1986, it was made as close to geocentric as possible at that time. The world-wide Doppler satellite tracking network, together with VLBI and precise Doppler networks in North America and the most recent Bureau International de l'Heure (BIH) definitions of the polar axis and longitude origin all contributed to determining this reference frame. Adoption of the GRS80 ellipsoid as the horizontal reference surface most suitable for global application completed the specification of the **NAD83 datum**.

The NAD83 reference frame was originally produced by a major readjustment of all the networks in North America. In the first phase, Canada and the United States cooperated in a simultaneous adjustment in 1986, with constraints provided by the elements as described in the previous paragraph. While the U.S. included all of their networks, from highest accuracy geodetic to local and municipal traversing, Canada included only the primary geodetic network at this first stage. Lower accuracy networks were integrated by the agencies of the Canadian Geodetic Reference System Committee (CGRSC) in subsequent adjustments over the next several years, thereby densifying and updating the NAD83 network. Although these adjustments added new points and changed coordinates for some points already adjusted, they were still referred to the NAD83 datum. The datum specification and reference frame had not been changed.

Table 1: NAD83 Integration Adjustments

Name	Year	Data Included
July86	1986	Framework (CDN), US Networks
SIHBA	1989-90	Secondary Integration, East, West & North
NMIP93	1993	Secondary Densification, Western & North networks

Similarly, whenever new surveys are done that enhance the existing network, some changes to the coordinates of the network points follow as the normal course of events. In the simplest cases, the new surveys are constrained to the existing points to which they are connected, and the network is extended by adding new points. In cases where there are more complex connections with the existing network, or the new survey is more accurate or uncovers bias or blunders in the existing network, then the integration adjustment will alter existing coordinates and the network is updated as well. The datum or reference frame remain the same, however, and the network evolves. This process is known as **network maintenance**, and is why in a data base retrieval we often see coordinates with various dates or version identifiers, but still on the same datum.

No matter what reference frame we use, it will continue to grow and change. Usually the changes are within the inherent accuracies of the existing coordinates. Even when the changes are significant with respect to the assumed precision or internal accuracies, it is the reliability or external accuracies that come into play when previously undetected biases or blunders are revealed. The new surveys may provide this information either by strengthening the geometrical redundancy of the network, or by applying improved technology. In these contexts, it is acceptable to update coordinates based on new knowledge for the benefit of greater accuracy.

We have just seen how even the new NAD83 network has undergone some changes to integrate the secondary networks, and we will see that more changes are soon to come. Before we take a look at what sort of progress is in store for NAD83, first let's look back at the familiar NAD27 reference system in which so much of our legacy data is entrenched.

NAD27 - the System prior to 1986

When we look back at the widely used NAD27 reference system, we see that it was established with much older technology, but also went through many changes as it evolved.

The original NAD27 datum was established long before we had any ability to locate the centre of mass of the earth by tracking satellite orbits. With the Clarke 1866 ellipsoid selected as a best-fitting surface for North America, astronomic observations were used to position and orient it at a point in the central U.S. This was the classical technique for establishing a **local non-geocentric horizontal datum** or reference frame. Triangulation networks radiated from this datum point, known as Meade's Ranch, to provide access to the NAD27 reference frame throughout North America.

By today's standards, however, the measurements by theodolite triangulation with scale provided by taped baselines were crude and prone to accumulation of systematic error over long distances. At first, this became apparent when closing new arcs of triangulation between existing arcs as the network was being extended and reinforced. Network adjustments were carried out by laborious manual computations, and new networks were forced to absorb any misclosures because concurrent readjustment of portions of the existing networks was not a viable option. With the advent of new technologies the problems became even more apparent and acute. First, electronic distance measurement

(EDM) pointed out deficiencies in scale at a local level, and later Doppler satellite positioning revealed very significant distortions over long distances.

At the same time as the measurement technologies were advancing, so were electronic computing technologies. Least squares adjustment software made it feasible for the first time to readjust large portions of the existing network to alleviate serious problems. Major readjustments were carried out during the mid-1970's in southern Ontario and southern Quebec. The Quebec system was later extended to the rest of the province under the title Compensation Géodésique de Québec de 1977 (CGQ77). In all of these cases, the resulting coordinates superseded the previously published values and became the officially adopted values for those regions.

In a special project to test newly developed software designed to handle very large simultaneous adjustments, all of the Canadian primary triangulation network was readjusted in 1976. Although this MAY76 version was never intended to replace the existing values, it did gain favour as control for major mapping programs at the federal level for the National Topographic System (NTS) and in the province of Ontario for the Ontario Base Mapping (OBM).

Table 2: NAD27 Major Variations

Name	Data Included
NAD27 Original Realization prior to mid-1970's	National Primary Framework accumulated over many years Secondary Networks integrated by local adjustments as established over many years
Southern Ontario 1975	Primary Framework & Secondary Networks in Ontario
MAY76	National Primary Framework
CGQ77	Primary Framework & Secondary Networks in Quebec

These are only the major variations of the NAD27 reference frame. Many others occurred on a regional or local basis, carried out by different levels of responsible authorities such as federal, provincial, and municipal government agencies. As with the major variations described above, all were constrained to adjacent control networks so that they remained on the NAD27 datum and adhered to the established reference frame.

The only notable exception was the establishment of the Average Terrestrial System of 1977 (ATS77) in the Maritime provinces. In this maintenance project, they chose to readjust their networks in a geocentric reference frame that was a preliminary specification for the upcoming NAD83.

NAD83 - the System from 1998 into the Future

Looking into the future of NAD83, we see progress that significantly improves the accuracy of the geodetic control survey networks while still maintaining the same datum and reference frame.

When the NAD83 reference frame was originally realized, every possible effort was made to keep it compatible with the WGS84 reference frame used by the Global Positioning System, the surveying technique destined to prevail in the future. Origin, orientation, and the GRS80 ellipsoid all came from the same sources as those used by the U.S. Defense Mapping Agency (DMA - now the National Imagery and Mapping Agency, or NIMA) for WGS84. The two reference frames were essentially equivalent at that time, but there has been no development of WGS84 into a tangible network with published coordinates for accessible monumented points. Instead, WGS84 is the reference frame for the tracking network and the broadcast orbits of the GPS satellites.

Since then, better solutions for the geocentric reference frame have been developed. The International Earth Rotation Service (IERS) maintains a CTS called the **International Terrestrial Reference System (ITRS)** and monitors Earth Orientation Parameters (EOP) for the scientific community by obtaining results from a global network of observing stations. They use GPS and other space-based observing techniques, such as Very Long Baseline Interferometry (VLBI), Lunar Laser Ranging (LLR), Satellite Laser Ranging (SLR), and Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS). Analysis of this network of space-based measuring techniques allows the IERS to better determine the center of mass and the orientation of the earth, as well as the coordinates of the observing stations.

The **International Terrestrial Reference Frame (ITRF)** is actually a series of realizations of the ITRS, and is revised and published on a regular basis. Positions of the observing stations are now considered to be accurate to the centimetre level. In addition to establishing coordinates for these points, each frame also includes velocity components that model the crustal motion of the tectonic plates of the Earth's crust, so that changes in the positions of the observing points from epoch to epoch can be accommodated. The

WGS84 reference frame has been updated on occasion to take advantage of this progress, and is now consistent with one of the recent ITRF specifications.

As active participants in the IERS, Canada benefits by gaining intimate knowledge of the relationship between the NAD83 and ITRF reference frames at the four Canadian VLBI sites which contribute to the realization of the ITRF. These are also the critical fiducial points of our NAD83 reference frame, and were held fixed throughout all of the network integration phases described previously. As a result, we have a means to go directly between the NAD83 reference frame and any of the realizations of the ITRF. In fact, the NAD83 reference frame is now effectively defined in terms of a best fitting similarity transformation from ITRF96. Thus, it can be maintained with a high degree of accuracy and confidence through this relationship, and users with special requirements can work in whichever system is appropriate for their purposes. This is further enhanced by the fact that Canada lies entirely on the North American plate of the Earth's crust. Hence, the NAD83 reference frame is not being significantly distorted in the horizontal component by tectonic motion, although there is significant vertical change due to post glacial rebound.

Both the NAD27 and NAD83 datums and reference systems that we considered earlier are 2-dimensional horizontal systems. They produce latitude and longitude coordinates on the reference ellipsoid. This is mainly due to the fact that most of the network points were established by classical triangulation, trilateration, and traversing methods. Any related height information had to be provided by spirit leveling, or by vertical angulation. These methods produce orthometric heights with respect to a vertical datum defined by mean sea-level. To relate the mean sea-level surface to the reference ellipsoid, a geoid model must be used. It approximates the variations in the height of the equipotential surface above the reference ellipsoid due to the irregular distribution of mass in the Earth. The traditional vertical and horizontal reference systems are separate entities, with no direct geometric relationship.

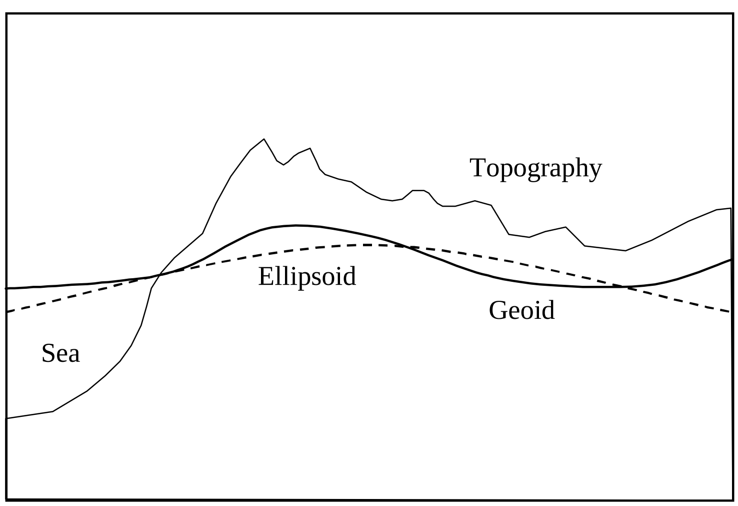

Figure 4: Geoid/Ellipsoid Relationship

But GPS satellite positioning provides three-dimensional coordinates in a CTS. When properly integrated into the reference frame, the Cartesian (X,Y,Z) coordinates can be converted geometrically to ellipsoidal (ϕ,λ,h) coordinates, where h is the height of the point above the reference ellipsoid. So we can see that any control networks established by GPS techniques and based on the 3D reference frame provided by the VLBI points would provide a 3D network on which to base further surveys.

Taking advantage of advances in high-end GPS processing techniques, geodetic control survey agencies in Canada have established three new layers of control survey networks: the continuously operated GPS sites of the Canadian Active Control System (CACS); the federal high accuracy Canadian Base Network (CBN); and the provincial High Precision Networks (HPN's) to further complement and densify the CBN. Each layer of this new 3D network is tied strongly to those above it. Each point of the CBN and HPN layers is also connected directly to the previously existing network. Although no major cooperative project is planned to take advantage of these connections, several individual geodetic control agencies responsible for the existing networks have already decided to readjust and update them and others may do so later.

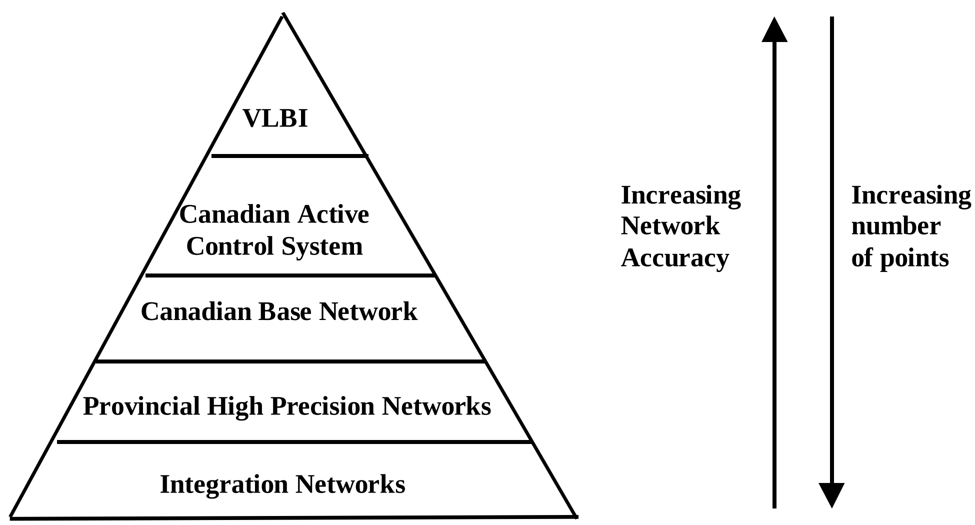

Figure 5: Hierarchy of Canadian Spatial Reference System

For the first time, we have a network that is accurate over long distances from the fundamental points of the reference frame. It provides an accessible backbone to which new surveys can be tied, and to improve the positions of the points of the existing network. All of the layers together form the realization of the Canadian Spatial Reference System (CSRS), supported by standards, calibration facilities, and all of the other requisite information to make it useful in a real world. The datum and reference frame for the CSRS is essentially unchanged from NAD83, and since it is an update from previous versions, the coordinates are said to be with respect to NAD83(CSRS).

Even though the NAD83 reference frame has been found to be non-geocentric by approximately 2 metres, it can be maintained with a high degree of accuracy and confidence through its relation to ITRF, and users with special requirements can work in whichever system is appropriate for their purposes.

What impact does this have on the NAD83 coordinates that are already in use? As we said, the existing networks were prone to bias over long distances, so we expect to see some differences where they are connected to the CBN and HPN points. These differences are a result of taking a different route from the datum points of the reference frame, and as we pointed out earlier, are most likely to be within the realistic accuracies of the old coordinates. But although we may see noticeable differences in the positions with respect to the reference frame or origin, the point-to-point or relative differences will mostly be insignificant.

To account for this concept, and the fact that we can now directly connect new local surveys to the reference frame through the CACS, CBN and HPN networks, a new set of accuracy standards has been developed. The old standards were based strictly on point-to-point relative accuracy, and ignored the very large and often unknown accuracy relative to the basic reference frame. By contrast, the new standards provide a description of both components. **Network accuracy** indicates how well a point is connected to the origin or to the datum points of the reference frame. **Local accuracy** indicates how well a point is connected to its surrounding neighbours.

Now we have two versions of the NAD83 reference frame that are potentially incompatible. How are we to reconcile the differences between them? There is no simple answer, but by objectively assessing a few factors, the situation can be put into a better perspective. The first step is to determine the network and local accuracies of both the data which the user collects and the control points to which that data is referred.

In general terms, if the change in the control point coordinates is less than the accuracy of your data, it is statistically insignificant and may usually be ignored for practical purposes. If the differences are significant, as a user you must decide whether it is more critical to make your new data consistent with your existing data, in which case local accuracy is the key factor, or whether consistency of your data with the reference frame is paramount, in which case network accuracy is the key factor. Obviously there are many possible combinations of these factors, and to deal with them exhaustively is beyond this discussion. Each situation needs to be judged on its own merit. Placing the differences in the context of accuracy to determine significance helps to keep the assessment objective.

Summary

We can see that despite all the different names floating about in connection with reference systems, there are really only two datums and basic reference frames involved in Canada, namely NAD27 and NAD83. The ATS77 system used by the Maritime provinces is a local exception. The other names reflect various major improvements, which have taken place in the normal course of maintaining the reference frames. Minor or more local improvements occur even more often, but do not receive the same degree of attention.

The CACS/CBN/HPN network layers of control have recently provided the means to again update the NAD83 reference frame by improving accuracy relative to the basic reference frame and adding the third dimension. The important thing to note is that there is no change in datum or reference frame. NAD83 continues as the official datum on which the reference system is based in most jurisdictions, and will continue to be supported for some time into the future. The improvements in network accuracy brought about by the CSRS version of the NAD83 reference frame are another step towards a broadly consistent system.

This broad consistency through improved network accuracy enables us to take better advantage of all that GPS has to offer, especially as its accuracy in long range differential positioning also evolves for all levels of users. We will no longer be as restricted to connecting to local control in order to account for local or regional distortions. It also facilitates 3D conformal transformation to and from any of the ITRF and ITRF-based reference frames (e.g. WGS84), using the well-determined relationship between NAD83 and ITRF.

Acknowledgements

The original appearance of this title and topic was a presentation made to the Canadian Hydrographic Conference held in Victoria, B.C. in March 1998. Contributors to that effort included: Denis Hains, Gordon Garrard, and Walter Gale of the Geodetic Survey Division, Geomatics Canada; Amin Kassam of Geographic Data B.C.; and Dave Doyle of the United States National Geodetic Survey. The authors of this written version for publication wish to especially thank Mike Craymer and Robert Duval of GSD for their constructive assistance in keeping the terminology correct. It is a persistent problem for us within geodetic organizations, and must drive our users mad when we are not consistent.

References

Boucher, Claude and Zuheir Altamimi, September1996. The ITRF and its Relationship to GPS. *GPS World Magazine*, *Volume 7*, *Number 9*.

Popelar, Josef Dr., 1997. Canadian Spatial Reference System, GSD internal presentation, unpublished.

Kouba, J. and J. Popelar, August 1994. Modern Geodetic Reference Frames for Precise Satellite Positioning and Navigation. *Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS94)*. University of Calgary, Alberta, Canada.

Langley, Richard B., February 1992. Basic Geodesy for GPS. *GPS World Magazine*.

Malys, Stephen, James A. Slater, Randall W. Smith, Larry E. Kunz, Steven C. Kenyon, June 1997. Status of the World Geodetic System 1984. *Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation (KIS97)*. University of Calgary, Alberta, Canada.

United States National Geodetic Survey, September 1986. *Geodetic Glossary*.