On the Correct Determination of Transformation Parameters of a Horizontal Geodetic Datum

Petr Vaníček[1], Pavel Novák[2], Michael R. Craymer[3] and Spiros Pagiatakis[4]
1) Department of Geodesy and Geomatics Engineering, University of New Brunswick, Post Office Box 4400, Fredericton, Canada E3B 5A3, Phone: (506) 453-4698, Fax: (506) 453-4943, E-mail: vanicek@unb.ca;
2) Research Institute of Geodesy, Topography and Cartography, Zdiby u Prahy Czech Republic E-mail: pnovak@pecny.asu.cas.cz;
3) Geodetic Survey Division, Natural Resources Canada, 615 Booth Street, Ottawa, Canada K1A 0E9 E-mail: craymer@nrcan.gc.ca;
4) Department of Earth and Atmospheric Science, York University, 4700 Keele Street, Toronto, ON, Canada M3J 1P3 E-mail: spiros@yorku.ca.

Abstract

In this contribution we formulate and solve the problem of determining transformation parameters in the transformation between a geodetic coordinate system (G-System) and the Conventional Terrestrial (geocentric) coordinate system (CT-system), or equivalently, between the datums associated with these systems. The transformation parameters are obtained from a set of points whose positions (coordinates) are known both, in the CT-system and in the G-system. It is shown that if the heights of these points above the horizontal datum are disregarded, as Vaníček and Steeves [1996] argued they should be, one can obtain transformation parameters of a horizontal datum positioned at “the origin of the geodetic network” (called also the “datum point”, or “fundamental datum point” by some people) and oriented with respect to the Local Astronomical coordinate system at this “origin”, to a very good accuracy. If the geodetic datum has been positioned and oriented some other way, and the misalignment of the two systems has been sought in terms of three rather than one unknown angle, one would have to pay extra attention to the spatial distribution of the common points. This is because correlations among the transformation parameters may play a more crucial role than in the simpler case described above. Even more crucial are the correlations between the transformation parameters (particularly the scale difference parameter) and parameters representing the network distortions. Because most older horizontal networks contain very significant distortions, these must be modelled either beforehand, or together with the transformation parameters. Otherwise, unmodelled distortions are likely to be absorbed by the transformation parameters giving an incorrect estimation of the actual transformation between the two datums.

Keywords: geodetic network • horizontal datum • coordinate transformation

Introduction

With the advent of satellite positioning systems, geodesists found themselves faced with the necessity of transforming positions (coordinates) from a geocentric Conventional Terrestrial coordinate system (CT-system), in which satellite-determined positions are given, into the generally non-geocentric, geodetic coordinate system (G-system), in which existing horizontal geodetic positions are known, and vice versa. To be more accurate, the satellite-determined coordinates are normally given in a coordinate system, such as the WGS-84 which would be one of the practical realizations of the CT-system. This, we feel, requires some explanation in terms of the terminology now prevalent in geodetic practice.

Nowadays, CT coordinate systems recommended for general use are regarded as being parts of "*reference systems"* such as WGS-84 or ITRS. A reference system is understood to consist of a coordinate system and a set of conventions and auxiliary models (for the earth gravity field, for atmospheric density, for tidal potential, etc.) that are to be used in the treatment of observed quantities. The coordinate system is assumed to be Cartesian and it may, or may not have a curvilinear/ellipsoidal coordinate system associated with it. Such a curvilinear coordinate system then implies that a reference ellipsoid (horizontal datum) of a certain size and shape – see below – has been adopted.

In order to be usable in practice, any coordinate system must have known positions (coordinates) of at least a few accessible (monumented) points associated with it. This association is referred to as a "*realisation of the coordinate system*". When the coordinate system has been "realised", then the reference system, of which it is part, becomes a "*reference frame*". In the sequel, we shall call these monumented points "*frame points",* because they play a specific role in "converting" a reference system to a reference frame. Generally, these points are a subset of the network of points that we shall be dealing with below. Clearly, the realisation does not have any effect on the auxiliary models, while the auxiliary models influence the realisation, i.e., the values of coordinates. An example of this terminology is the different ITRFs being different realisations of the ITRS.

Because here we are interested only in the geometrical aspects of the transformation between the CT-frame and G-frame, we shall be dealing only with the coordinate system part of the frames, leaving alone the set of conventions and auxiliary models that are an integral part of a frame. By the same token, we shall be also referring to the realizations of these systems as the "CT-system" and "G-system", without distinguishing which realization is really involved. For the same reason, we shall not be specifying which of the existing G-systems is considered.

We shall assume that the G-system we deal with here has a specific geodetic reference ellipsoid associated with it. This reference ellipsoid will be defined by the lengths of its semi-axes, a and b. Alternatively, such a geodetic reference ellipsoid may be assumed to have been defined by its major semi-axis a and flattening f, or the first numerical eccentricity e [Vaníček and Krakiwsky, 1986]. Such a reference ellipsoid must be understood to have been properly positioned and oriented within the earth and thus with respect to the CT-system, which, in turn, assures proper positioning and orientation of the G-system with respect to the CT-system. Some people want to distinguish between a "*reference ellipsoid*" and a "*horizontal datum*", the latter being a properly positioned and oriented reference ellipsoid. Since the latter is indubitably the case here, we shall be referring to the geodetic reference ellipsoid also as a "horizontal geodetic datum" for the distinction described above is moot.

Accepting these assumptions, we shall be dealing with the Cartesian G-coordinates (x^G, y^G, z^G), and the equivalent curvilinear G-coordinates (φ^G, λ^G, h^G). The standard nonlinear transformation equations [Vaníček and Krakiwsky, 1986, Eq. (15.63)]

$$\begin{bmatrix} x^G \\ y^G \\ z^G \end{bmatrix} = \begin{bmatrix} (N+h^G)\cos\varphi^G \cos\lambda^G \\ (N+h^G)\cos\varphi^G \sin\lambda^G \\ [N(1-e^2) + h^G]\sin\varphi^G \end{bmatrix}, \tag{1}$$

where N is the prime vertical radius of curvature of the reference ellipsoid at the point of interest, define the relation between the Cartesian and curvilinear G-coordinates. Most of the time, we will be referring to the ordered triplet of Cartesian coordinates $(x^G, y^G, z^G)^T$ simply as a vector $\boldsymbol{r}^G$.

The way the horizontal geodetic datum had been positioned and oriented with respect to the CT-system, i.e., the way the G-frame had been obtained from the G-system, dictates how the transformation between the G-system and the CT-system should be set up [Vaníček and Steeves, 1996]. We shall be distinguishing between G-systems (and thus between horizontal geodetic datums) positioned at the "*origin of the network*" and oriented (with respect to the Local Astronomical coordinate system at the "origin") directly, through six *"topocentric parameters"* [Vaníček and Wells, 1974], and those positioned and oriented indirectly, by means of a set of known positions. Without getting into any details here, we just wish to say that the former mode is the "classical" one, encountered in the majority of geodetic datums throughout the world, while the latter mode has been used in the establishment of the more recent geodetic datums.

The transformation between the CT-system and the G-system is usually defined in terms of the linear transformation equation

$$\boldsymbol{r}^G = \boldsymbol{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z)\, \boldsymbol{r}^{CT} - \boldsymbol{t}^{CT}, \tag{2}$$

where $\boldsymbol{R}$ denotes the rotation matrix involving the misalignment angles (ε_x, ε_y, ε_z) around the three Cartesian axes. The symbol $\boldsymbol{r}^{CT} = (x^{CT}, y^{CT}, z^{CT})^T$ stands for a position vector in the CT-system and $\boldsymbol{t}^{CT}$ represents the position vector (in the CT-system) of the centre of the reference ellipsoid (i.e., the origin of the G-system), known as the "translation vector". We note the obvious reciprocity

$$\boldsymbol{t}^{CT} = (t_x^{CT}, t_y^{CT}, t_z^{CT})^T = -\boldsymbol{t}^G, \tag{3}$$

where $\boldsymbol{t}^G$ is the position vector of the earth's centre of mass (i.e., the origin of the CT-system), in the G-system. We shall not discuss here cases where specific realisations of the CT-systems imply that the CT-system origin does not coincide with the centre of mass of the earth. We shall understand that a CT-system has its origin in the earth centre of mass by definition.

There are six "*transformation parameters*" present in Eq. (2): t_x^{CT}, t_y^{CT}, t_z^{CT}, ε_x, ε_y, ε_z. Since we are talking about a transformation between coordinate systems, there is no need for introducing a scale parameter representing the difference between the scales of coordinate systems. The scale difference is associated with transformations between coordinates and not with transformations between (coordinate) systems. Since a coordinate system is an entity separate from the point configuration described in that coordinate system, one can imagine a coordinate system that exists (by definition) regardless of the existence of any point configuration. If there have not been any measurements of distances carried out in the coordinate system, how would one be able to talk about a scale? Clearly, one could speak of a scale only after the scale had been brought in through some distance measurements associated with the determination of coordinates, or coordinate differences.

Suppose for a moment that we are willing to regard the scale implied by measured distances between the points in a configuration to "define the scale of the coordinate system". That will violate our basic assumption that the coordinate system is an entity by itself, separate from the point configuration (e.g., the network) that we may wish to describe in that coordinate system. The situation is slightly more complicated when the reference system of which the coordinate system is the essential part, is realised (positioned and oriented) by a set of "frame points" (that "convert" the system to the reference frame). Should these "frame points" be taken as being more closely related to the coordinate system or to the independent point configuration? We shall discuss this point in the next section of this paper.

It is useful to realize that, for small misalignment angles between two coordinate systems (which is always the case in practice), the "*misalignment term*", i.e., the first term on the right-hand side of Eq. (2), can be also written as [Vaníček and Carrera, 1985]

$$\boldsymbol{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z)\, \boldsymbol{r}^{CT} = \boldsymbol{r}^{CT} + \omega \times \boldsymbol{r}^{CT}, \tag{4}$$

where ω is the "*misalignment vector*" defined as

$$\omega = (\varepsilon_x, \varepsilon_y, \varepsilon_z)^{\mathrm{T}}, \tag{5}$$

and "×" denotes the vector product. The interesting geometrical insight one gets from Eq. (4) is that (for small misalignment angles) the rotated position vector $\boldsymbol{R}(\varepsilon_x, \varepsilon_y, \varepsilon_z)\, \boldsymbol{r}^{CT}$ can be obtained from the original position vector $\boldsymbol{r}^{CT}$ by a small shift $\omega \times \boldsymbol{r}^{CT}$ in a direction perpendicular to the misalignment vector ω and also perpendicular to the position vector $\boldsymbol{r}^{CT}$ itself. Generally, vector ω has an arbitrary direction and magnitude. It can be seen rather clearly from Eq. (4) that if the direction of ω happens to coincide with the direction of $\boldsymbol{r}^{CT}$, the second term on the right-hand side of Eq. (4) goes to zero vector and there is no effect from the misalignment on the position $\boldsymbol{r}^{CT}$.

It was shown by Vaníček and Wells [1974] that if the geodetic horizontal datum (and thus the G-system) is positioned and oriented the "classical" way, i.e., by means of the origin of the network, the misalignment must take the form of a rotation around the ellipsoidal normal passing through the network origin. This means that the direction of the misalignment vector ω must coincide with the direction of the ellipsoidal normal at the network origin; i.e.,

$$\omega = \omega_o = \omega_o\, (\cos\varphi_o \cos\lambda_o, \cos\varphi_o \sin\lambda_o, \sin\lambda_o)^{\mathrm{T}}, \tag{6}$$

where ω_o is the magnitude of the misalignment and (φ_o, λ_o) are the geodetic (curvilinear) coordinates of the network origin, and the vector on the right-hand side is the unit vector normal to the reference ellipsoid at the network origin.

It is interesting to realize that even when one does not expect the (small) misalignment angle to have occurred in the same direction as the normal to the reference ellipsoid at the network origin, i.e., when the geodetic horizontal datum had been positioned and oriented by means of a set of points, the transformation (2) can be written in a similar form as

$$\boldsymbol{r}^{G} = \boldsymbol{r}^{CT} + \omega_m \times \boldsymbol{r}^{CT} - \boldsymbol{t}^{CT} . \tag{7}$$

Here

$$\omega_m = \omega_m \, (\cos\varphi_m \cos\lambda_m, \cos\varphi_m \sin\lambda_m, \sin\varphi_m)^{\mathrm{T}} , \tag{8}$$

and (φ_m, λ_m) are the geodetic coordinates of the point to be determined, around the normal of which the misalignment of magnitude ω_m takes place. The sought six-tuple of transformation parameters then becomes t_x^{CT}, t_y^{CT}, t_z^{CT}, ω_m, φ_m, λ_m.

Statement and Formulation of the Problem

The position of a specific horizontal geodetic datum (reference ellipsoid) with respect to the CT-system cannot be determined directly. Therefore, we have no choice but to do it indirectly using coordinates of two sets of n points, one set in the CT-system and the other in the G-system. Now, these "*common points*" normally belong to the network, but some of them may also belong to the set of "frame points". Be that as it may, these coordinates are naturally burdened with errors (both, random and systematic) originating in the observations, as well as in some of the shortcomings of the computational procedures (systematic errors) used to derive the coordinates from the original observations [Vaníček and Steeves, 1996]. Here is where the above discussed scale difference comes into the picture; it must be taken into account, together with the other existing distortions of the coordinates that may be coming from the above discussed sources or, e.g., from geodynamical phenomena. These distortions have to be modelled and the distortion parameters estimated. We shall have more to say about this aspect a little later; for the moment, we shall assume that these distortions have been taken care of one way or another.

If we do not have such two sets of coordinates for a sufficient number of points then we cannot solve the problem. When we do, we can formulate n vectorial observation equations (equivalent to $3 \times n$ scalar equations) of the kind of Eqs. (2) or (7). The two sets of coordinates, $\{\boldsymbol{r}_1^{G}, \boldsymbol{r}_2^{G}, \ldots , \boldsymbol{r}_n^{G}\}$ and $\{\boldsymbol{r}_1^{CT}, \boldsymbol{r}_2^{CT}, \ldots , \boldsymbol{r}_n^{CT}\}$ become the "known" quantities and the transformation parameters become the "unknown" quantities.

Once we have formulated the $3 \times n$ observation equations, which, still assuming a small misalignment, can be written as, cf. Eq. (7),

$$i = 1, 2, \ldots , n \ : \ \boldsymbol{r}_i^{G} - \boldsymbol{r}_i^{CT} = \omega_m \times \boldsymbol{r}_i^{CT} - \boldsymbol{t}^{CT} , \tag{9}$$

we can attempt to solve them for the six unknown parameters t_x^{CT}, t_y^{CT}, t_z^{CT}, ω_m, φ_m, λ_m. If the geodetic horizontal datum were positioned and oriented the "classical" way then we would have only four unknown parameters t_x^{CT}, t_y^{CT}, t_z^{CT}, ω_o to solve for. In the sequel, we will be thus speaking about either "*six-parametric*", or "*four-parametric*" transformations.

It is helpful to reformulate Eq. (9) in such a way that the unknown transformation parameters appear in the usual form, as a vector pre-multiplied by a known matrix (design matrix). This can be done simply by realizing that

$$\omega_m \times \boldsymbol{r}_i^{CT} = -\boldsymbol{r}_i^{CT} \times \omega_m \,, \tag{10}$$

which can be then written as

$$-\boldsymbol{r}_i^{CT} \times \omega_m \,, = \boldsymbol{Q}_i \, \omega_m \,, \tag{11}$$

where

$$\boldsymbol{Q}^i = \begin{bmatrix} 0 & -z_i^{CT} & y_i^{CT} \\ z_i^{CT} & 0 & -x_i^{CT} \\ -y_i^{CT} & x_i^{CT} & 0 \end{bmatrix} . \tag{12}$$

Then Eq. (9) can be restated as

$$i = 1, 2, \ldots , n \; : \; \boldsymbol{r}_i^G - \boldsymbol{r}_i^{CT} = -\boldsymbol{Q}_i \, \omega_m - \boldsymbol{t}^{CT} \,, \tag{13}$$

or, more simply, as

$$i = 1, 2, \ldots , n \; : \; \boldsymbol{r}_i^G - \boldsymbol{r}_i^{CT} = \boldsymbol{A}_i \, \boldsymbol{x} \,, \tag{14}$$

where the design matrix $\boldsymbol{A}_i$ is given by

$$\boldsymbol{A}_i = [\, \boldsymbol{Q}_i \,, \boldsymbol{I} \,] \,, \tag{15}$$

and the unknown vector $\boldsymbol{x}$ consists of

$$\boldsymbol{x} = [\, \omega_m \,, \boldsymbol{t}^{CT} \,]^{\mathrm{T}} \,. \tag{16}$$

The sub-vector ω_m has to be, in the end, resolved into the three transformation parameters ω_m, φ_m, λ_m according to Eqs. (8) or (6). If a four-parametric transformation is used then $-\boldsymbol{Q}_i \, \omega_m$ in Eq. (13) must be replaced by

$$-\boldsymbol{Q}_i \, \omega_o = -\boldsymbol{Q}_i^o \, \omega_o \,, \tag{17}$$

where the vector $\boldsymbol{Q}_i^o$ is a product of the matrix $\boldsymbol{Q}_i$ with the unit vector normal to the reference ellipsoid at the network origin, given by Eq. (6). Note that ω_o on the left hand side is a vector, while ω_o on right hand side is a scalar.

Let us now return to the coordinates on the left-hand side of Eqs. (14). As we mentioned above, these coordinates are distorted and these distortions (be they a single scale difference for the entire network or many different regional scale differences or some more complicated cases) need be accounted for, either beforehand or simultaneously with the estimation of the coordinate system transformation. The problem with trying to estimate the distortion and transformation parameters independently of each other is that they are often highly correlated. Estimating the distortions first may result in some portion of the transformation parameters being absorbed by the distortion parameters. Conversely, estimating the transformation parameters first may result in some portion of the distortion parameters being absorbed by the transformation parameters. In both cases one would end up with incorrect transformation parameters.

Thus estimating both sets of parameters simultaneously seems to be the preferred choice. This can be done quite easily by adding to Eq. (14) a linear deformation model. Clearly, if the distortions are modelled by means of a linear model that contains nuisance parameters (see, for example, [Junkins, 1991]), these nuisance parameters may be resolved simultaneously with the transformation parameters that we have been working with above. The observation equations (14), containing both the transformation, as well as nuisance parameters can be solved for both kinds of parameters simultaneously. It has been a standard practice however (see, e.g., [United States Defence Mapping Agency, 1987]), to lump the distortions described by nuisance parameters together with the transformation parameters, rather then keeping them separate. This practice is misleading because it obfuscates the nature of the transformation: rather than keeping the coordinate system transformation separate from the distortions, it mixes the two things together yielding some “coordinate quasi-transformation parameters” (generally different from the transformation parameters described above) which vary from location to location. This practice is also clumsy because it precludes any reasonable attempt of an assessment of accuracy of the (coordinate system) transformation parameters as well as an assessment of correlations.

To simplify the discussion here, we shall assume that the distortions have been already modelled and that the distortion model has become part of the system of observation equations. We do not wish to discuss the distortion modelling here as it represents a problem quite different from the one we are discussing in this paper. We just want to point out that at this stage, that if some or all of the “common points” are also “frame points”, their distortions may have to be modelled independently from the other (network) points.

Now, if there are more observation equations available than the number of transformation and nuisance (for simplicity not considered any further in our derivations here) parameters sought, the least-squares approach to the solution is normally employed. For the least-squares solution, the covariance matrix of the vector of coordinate differences

$$i = 1, 2, \ldots, n \;:\; \boldsymbol{r}_i^G - \boldsymbol{r}_i^{CT} = \Delta\boldsymbol{r}_i = (\Delta x, \Delta y, \Delta z)_i^T \,, \tag{18}$$

has to be properly assembled, including the appropriate covariances. Still disregarding the linear model of coordinate deformation, the system of either four or six normal equations is formulated and solved.

Correlations among some datum transformation parameters and some nuisance parameters appear explicitly in the a posteriori covariance matrix of the two kinds of parameters. In particular, a high correlation is often experienced between the translation vector and the scale factor. The rise of all these correlations represents a very interesting and challenging problem, which should be treated in a separate paper.

There is yet another complication that should be discussed here. In our formulation above, we have been using three-dimensional positions (three Cartesian (x, y, z) or curvilinear (φ, λ, h) coordinates) to describe the position of each point needed in the formulation of the observation equations. Vaníček and Steeves [1996] argued that in the case of terrestrial geodetic horizontal networks (networks established by classical, terrestrial-based measurement techniques), the accuracy of geodetic (ellipsoidal) heights h may be significantly lower than the accuracy of the other two coordinates φ and λ, and that the height h is often missing altogether. Although accurate orthometric and geoidal heights are sometimes provided in the terrestrial position set, they are determined by completely different means than the horizontal positions and as such, in our opinion, should not be mixed with them in a purely geometrical transformation. This non-availability and incompatibility makes the practice of using three-dimensional coordinates for the determination of horizontal datum transformation parameters unnecessarily much less transparent and probably less accurate than needs be.

Consequently, Vaníček and Steeves [ibid.] have recommended that only horizontal (2-D) coordinates be used for the purpose of coordinate transformation parameter determination. They also suggest a very simple technique designed for the use of such coordinates, whereby the three-dimensional Cartesian coordinates in the CT-frame are first converted to two-dimensional coordinates on a reference ellipsoid of the same size and shape as the reference ellipsoid for the G-frame, but concentric with the origin of the CT-system. (We note that there are now two reference ellipsoids for the CT-frame: the one defined for the CT-frame and the one compatible with the G-frame.) So derived horizontal positions in the CT-frame can be then directly compared against the horizontal positions in the G-frame. We note that the vertical positions in the CT-system (heights above the reference ellipsoid) are thus not used, even though they do not suffer from the same malady as the heights in the G-system do. This is the inevitable sacrifice resulting from leaving the latter heights out of the calculations.

Let us now have a look at how the observation equations (14) change if we want to work with curvilinear rather than Cartesian coordinates so we can eliminate more easily the third coordinate, the height. The conversion of curvilinear coordinate differences into Cartesian coordinate differences is done by the following linear equations [Vaníček and Krakiwsky, 1986, Eq. (15.93)]

$$\Delta \boldsymbol{r} \;=\; (\Delta x, \Delta y, \Delta z)^T \;=\; \boldsymbol{J}\,(\Delta\varphi, \Delta\lambda, \Delta h)^T\,, \tag{19}$$

where the Jacobi matrix of transformation reads

$$\boldsymbol{J} = \begin{bmatrix} -M \sin\varphi \cos\lambda & -N \cos\varphi \sin\lambda & \cos\varphi \cos\lambda \\ -M \sin\varphi \sin\lambda & N \cos\varphi \cos\lambda & \cos\varphi \sin\lambda \\ M \cos\varphi & 0 & \sin\varphi \end{bmatrix}, \tag{20}$$

where M is the meridian radius of curvature of the reference ellipsoid at the point of interest and N has been defined earlier. In this equation, the heights h have already been set to zero, to keep the expression as simple as possible. Substituting Eq. (19) into Eq. (14), we obtain

$$i = 1, 2, \ldots, n \ : \ \boldsymbol{J}_{\mathrm{i}}(\Delta\varphi, \Delta\lambda, \Delta h)_i^T \ = \ \boldsymbol{A}_i \, \boldsymbol{x} \, , \tag{21}$$

Pre-multiplying each equation by $\boldsymbol{J}_i^{-1}$ (the inverse exists because $\boldsymbol{J}_i$ is regular for all $i = 1, 2, \ldots, n$), we get

$$i = 1, 2, \ldots, n \ : \ (\Delta\varphi, \Delta\lambda, \Delta h)_i^T \ = \ \boldsymbol{J}_i^{-1} \boldsymbol{A}_i \, \boldsymbol{x} \, , \tag{22}$$

This is the system of $3 \times n$ observation equations for the n "common points" that are to be used for the determination of the transformation parameters.

It is interesting to note that when nuisance parameters describing the position deformation are not considered, horizontal coordinates of three points known in both, the CT-system and the G-system are enough to guarantee a unique solution for six transformation parameters. Only two such points suffice to determine the four transformation parameters.

Solution

We now have two different possibilities how to solve this system of observation equations in an appropriate manner:

1. We can neglect the height differences Δh in Eqs. (22), and reduce the system of $3 \times n$ observation equations into a system of $2 \times n$ observation equations involving only the horizontal coordinate differences. We shall refer to this model as the two-dimensional or 2D model.

2. We can leave the Δh in the system of observation equations and suppress their effect in the normal equations by associating some very large a priori errors with those height differences. This results in a three-dimensional model, in which the third dimension, heights, is moot; we shall call this the 3D model.

We also tried to implement the model described by Okia [1996] — a 3D model with geodetic heights forced to zero and associated with very small a priori errors — but failed to reproduce his results. Using this setup, Okia claimed to have obtained transformation parameters exactly, i.e., without any formal errors. Our subsequent computations have shown decisively that this is not the case; the transformation parameters are actually estimated with larger errors than in the above two approaches.

When the a priori errors in Δh are chosen to be sufficiently large, the two models give the same numerical results, as it intuitively should. It does not seem to matter if the model is formulated in two or three dimensions, when, at the end, the contribution of heights in the three-dimensional model is effectively eliminated by means of stipulated large errors in Δh. Thus we will use only the smaller system of equations. We shall not discuss these two sets of normal equations here as the formulation of these equations is routine and we will go directly to the numerical results.

We shall first demonstrate the performance of our models on a set of twelve simulated "common points" and then on a set of twelve real "common points", both from Canada. For each set, coordinates are known in both the North American Datum of 1927 (NAD 27), a realization of a G-system and the "geocentric" North American Datum of 1983 (NAD 83), which represents a realization of a CT-coordinate system. To have the simulated case somewhat closer to the real case described below, we decided to use the same number of points, i.e., twelve as mentioned above. But we have distributed these simulated points more regularly (see Fig. 1) vis-à-vis the position of the origin of NAD 27 (Meade's Ranch, which had been used in 1927 for the positioning and orientation of NAD 27) then they are in the real case below.

We first chose the horizontal positions on the geocentric NAD 83, and generated uncorrelated random errors (Gaussian noise) with standard deviations σ_φ = 0.005" (corresponding to 0.15 metre) and σ_λ = 0.005" (corresponding to 0.1 metre), which were then added to the chosen positions to represent the CT-positions of the determining points. Then the (errorless) positions on NAD 83 were transformed to NAD 27 by means of three translations t_x^{CT} = +100 metres, t_y^{CT} = −100 metres, t_z^{CT} = +100 metres, and a misalignment of ω_o = −1" around the normal to the NAD 27 ellipsoid (Clarke) at the origin. Finally, these simulated errorless NAD 27 positions were burdened with uncorrelated Gaussian noise characterised by σ_φ = 0.05" (corresponding to 1.5 metre) and σ_λ = 0.05" (corresponding to 1.0 metre). To make sure that the error estimation for the computed transformation parameters works properly, we generated several different noise sequences and analyzed the estimated parameter errors for statistical consistency.

In Table 1 we show the results for three typical simulated cases of twelve "common points" using the four-parameter transformation. The transformation parameters are resolved quite well and their errors appear to be compatible with the input errors in positions as discussed above. The largest discrepancies are encountered for the y- and z-translations. These discrepancies are more variable with the choice of the random sequence than the other two because there is a relatively high correlation of −0.93 between these two, which points to a lack of resolvability due to this particular geometrical configuration. The other correlations are significantly smaller and, consequently, the other parameters are resolved better. The whole question of correlations is, of course, a very interesting one and is discussed further in Kutoglu et al. [2002].

Table 2 shows the result for the same configuration using the six-parametric transformation. The accuracy of these results is one order of magnitude worse than for the four-parametric transformation. Interestingly, the point around the normal of which the misalignment takes place is located at (−47°, 31°), (−8°, −67°), and (10°, 53°) for the three random number sequences respectively, instead of the correct location of φ_o +39.22°, λ_o −98.54°. The amount of misalignment is estimated as 1.232", 1.268", and 1.564" respectively, instead of the correct amount of 1.000". Clearly, the use of the six-parametric model does not produce very satisfactory results when the horizontal datum had been, in fact positioned using only four parameters, as was the case in this simulated example. This shows that it is important to use the appropriate transformation model when seeking the datum transformation parameters.

To show how our algorithm works with real data, we evaluate the transformation parameters for the NAD 27 geodetic datum from coordinates of twelve points obtained from the Geodetic Survey Division of Natural Resources Canada (see Fig. 2). Geodetic heights of only nine out of the twelve points were available — this illustrates one of the disadvantages of using heights of the “common points” in the parameter estimation. As we do not know the real accuracy of the twelve positions on NAD 27, which were determined by terrestrial means, we assume the following values: σ_φ = 0.05" (corresponding to 1.5 metre) and σ_λ = 0.05" (corresponding to 1.0 metre). Also we assume no correlations between the two horizontal coordinates of each point and no correlations among the twelve points for our stochastic model.

The results for the four-parametric transformation using the real data described above are given in Table 3. Clearly, we were too optimistic when assigning errors to the terrestrial positions; the value of the a posteriori variance factor (12.759) is very large. Neglecting the existing correlations among the input coordinates certainly contributes to one's uneasiness about the values for the a posteriori error estimates as one can be quite certain that there are significant correlations present. We also note that the geometrical configuration of the real points is much less favourable for the solution than the configuration of the simulated points discussed above. Last, but not least, there are definitely some systematic errors in the terrestrial positions, which we have not even attempted to model; we have included no model for the position distortions. Taking all these shortcomings into consideration, the results look fairly reasonable with the estimated errors of the parameters being about twice as large as in the simulated case. The estimated values of translation components agree reasonably well with some previous determinations — see, e.g., [Merry and Vaníček, 1974, Table 3; United States Defence Mapping Agency, 1987] — and the estimated misalignment of 0.230" agrees with the value of 0.3" estimated by [Wells and Vaníček, 1975] within the errors.

Much like for the simulated data, the six-parametric model gives unsatisfactory results also for the real data (Table 4). It places the misalignment rotation centre at (36°, 126°), and estimates the magnitude of the misalignment as being an unreasonable 8.237". The estimated errors of transformation parameters are again about one order of magnitude larger than in the four-parametric case.

Conclusions and Recommendations

There exists a rather strong motivation for disregarding the heights of the points, which belong to horizontal geodetic networks in the derivation of transformation parameters. This motivation was discussed by Vaníček and Steeves [1996] and has not been repeated here. In this paper we have shown how transformation parameters between a geodetic coordinate system (G-system) and its associated geodetic reference ellipsoid (also known as a geodetic horizontal datum), and the conventional terrestrial coordinate system (CT-system) can be determined from horizontal positions alone (the geodetic heights are set to zero). It can be done simply by taking the points common to both reference frames, i.e., the “common points” whose positions in both the CT-and the G-systems are known, and projecting them onto their respective reference ellipsoids by setting their geodetic heights to zero. Here the reference ellipsoid in the geocentric CT-system of coordinates has to have the same size and shape as the reference ellipsoid for the G-system. These two-dimensional positions are then used to derive the parameters needed for transformations from one coordinate system to the other.

The selection of the transformation model involving one misalignment angle is predicated on the procedures used in the classical geodetic horizontal datum establishment (positioning and orientation of the geodetic reference ellipsoid, or classical geodetic system realisation). This predication was discussed by Vaníček and Wells [1974], and only the consequences have been shown here. In other realisation cases three misalignment angles should be sought. Both models are used side by side and it is shown that the six-parametric model (involving three misalignment angles) performs significantly worse than the four-parametric model (involving one misalignment angle) when the horizontal geodetic datum was in fact positioned and oriented using the classical technique.

The approach we have followed here deals with transformations between coordinate systems. If one is interested in transforming coordinate values (from one reference frame to another - as one always is in geodesy) then distortions inherent in the coordinates <u>must</u> be also taken into account. These distortions include, but are usually not limited to, the scale distortions in the coordinate values. These distortions are usually parameterized and the distortion (nuisance) parameters are solved for either separately, in a sequential manner, or together with the transformation parameters. It should be mentioned that while the distortions of coordinates vary from place to place, the transformation parameters do not.

It should be borne in mind that there may exist significant correlations among the transformation and distortion (nuisance) parameters. These correlations depend predominantly on the spatial distribution of the “common points”. Sometimes, it may not be even possible to de-correlate some of these parameters, but the discussion of these cases will have to wait for another paper.

By keeping these two concepts, i.e., the transformation between coordinate systems and the distortion of coordinates, separate, a much clearer understanding of the interplay of coordinate systems can be gained. Also, a more rigorous and transparent error analysis can be brought to bear on the problem of coordinate transformations. Last, but not least, the effect of geodetic horizontal datum misalignment with respect to the CT-system on the deflections of the vertical, as well as on geodetic azimuths [Grafarend and Richter, 1977; Vaníček and Carrera, 1985] can be rigorously evaluated. To keep the two concepts separate should become a sound geodetic practice.

Acknowledgments

We are grateful to Alfred Kleusberg and Will Featherstone for many fruitful discussions of this paper. The Geodetic Survey Division of Natural Resources Canada provided us with the data used in our analysis.

References

Grafarend, E. and B. Richter (1977). The generalized Laplace condition. *Bulletin Geodesique*, 51, 4, pp. 287-293.

Junkins, D. (1991). The national transformation for converting between NAD 27 and NAD 83 in Canada. In “Moving to NAD 83”, D. C. Barnes (ed.), CISM, Ottawa, pp. 16-40.

Kutoglu, H.S., C. Mekik, H. Akcin (2002). A comparison of Two Well Known Models for 7-Parameter Transformation. The Australian Surveyor, 47(1), pp. 24-30.

Merry, C. L. and P. Vaníček (1974). The Geoid and Datum Translation Components. *Canadian Surveyor*, 28, pp. 56-62.

Okia, Y. (1996). Determination of datum transformation parameters from terrestrial and satellite networks. Ph.D. dissertation, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton.

United States Defense Mapping Agency (1987). Supplement to Department of Defense World Geodetic System 1984. Defence Mapping Agency Technical Report No. 8350.2-A, Washington, D.C.

Vaníček, P. and D. E. Wells (1974). Positioning of horizontal geodetic datums. *Canadian Surveyor*, 28 (5), pp. 531-538.

Vaníček, P. and G. Carrera (1985). Reference ellipsoid misalignment, deflection components and geodetic azimuths. *Canadian Surveyor*, 39(2), pp. 123-130.

Vaníček, P. and E. J. Krakiwsky (1986). *Geodesy: The Concepts*. 2nd corrected edition, North Holland, Amsterdam.

Vaníček, P. and R. R. Steeves (1996). Transformation of coordinates between two horizontal geodetic datums. *Journal of Geodesy*, 70(11), pp. 740-745.

Wells, D. E. and P. Vaníček (1975). Alignment of geodetic and satellite coordinate systems to the average *terrestri*al system. Bulletin Geodesique, 117, pp. 241-257.

Table 1: Four-parametric model with simulated data.

No	Parameter	Unit	$\hat{x}$	$\hat{x} - x^o$	$\hat{\sigma}_{\hat{x}}$	$t_{\hat{x}}$
1	ω_o	"	−0.989	0.011	0.086	0.016
	t_x	m	99.499	−0.501	0.507	0.976
	t_y	m	−102.909	−2.909	2.086	1.944
	t_z	m	102.547	2.547	1.817	1.965
	σ^2_o		1.364			
2	ω_o	"	−0.932	0.068	0.066	1.044
	t_x	m	100.013	0.013	0.390	0.001
	t_y	m	−100.236	−0.236	1.604	0.022
	t_z	m	100.277	0.277	1.397	0.039
	σ^2_o		0.806			
3	ω_o	"	−0.873	0.127	0.073	3.047
	t_x	m	99.540	−0.460	0.430	1.146
	t_y	m	−103.575	−3.575	1.768	4.089
	t_z	m	101.798	1.798	1.5407	1.364
	σ^2_o		0.980			

Table 2: Six-parametric model with simulated data.

No	Parameter	Unit	$\hat{x}$	$\hat{x} - x^o$	$\hat{\sigma}_{\hat{x}}$	$t_{\hat{x}}$
1	ε_x	"	0.718	0.603	0.884	0.465
	ε_y	"	0.433	−0.333	0.662	0.254
	ε_z	"	−0.903	−0.271	0.782	0.120
	t_x	m	112.435	12.435	31.087	0.160
	t_y	m	−91.813	8.187	18.249	0.201
	t_z	m	117.810	17.810	21.103	0.712
	σ^2_o		1.463			
2	ε_x	"	−0.493	−0.608	0.665	0.838
	ε_y	"	1.156	0.380	0.498	0.614
	ε_z	"	−0.172	0.460	0.588	0.611
	t_x	m	81.516	−18.484	23.387	0.624
	t_y	m	−110.789	−10.789	13.729	0.618
	t_z	m	84.683	−15.317	15.876	0.931
	σ^2_o		0.828			
3	ε_x	"	0.919	0.804	0.710	1.283
	ε_y	"	1.237	0.461	0.532	0.783
	ε_z	"	0.265	0.897	0.628	2.039
	t_x	m	69.244	−30.756	24.983	1.516
	t_y	m	−84.413	15.587	14.666	1.130
	t_z	m	118.997	18.997	16.959	1.255
	σ^2_o		0.945			

Table 3: Four-parametric model with real data.

Parameter	Unit	$\hat{x}$	$\hat{\sigma}_{\hat{x}}$
ω_o	"	0.230	0.272
t_x	m	9.515	1.716
t_y	m	−143.263	4.044
t_z	m	−205.264	4.893
σ^2_o		12.759	

Table 4: Six-parametric model with real data.

Parameter	Unit	$\hat{x}$	$\hat{\sigma}_{\hat{x}}$
ω_x	"	−3.974	2.024
ω_y	"	5.377	1.221
ω_z	"	4.810	1.032
t_x	m	−217.041	49.717
t_y	m	−230.491	46.415
t_z	m	−285.328	40.418
σ^2_o		6.355	

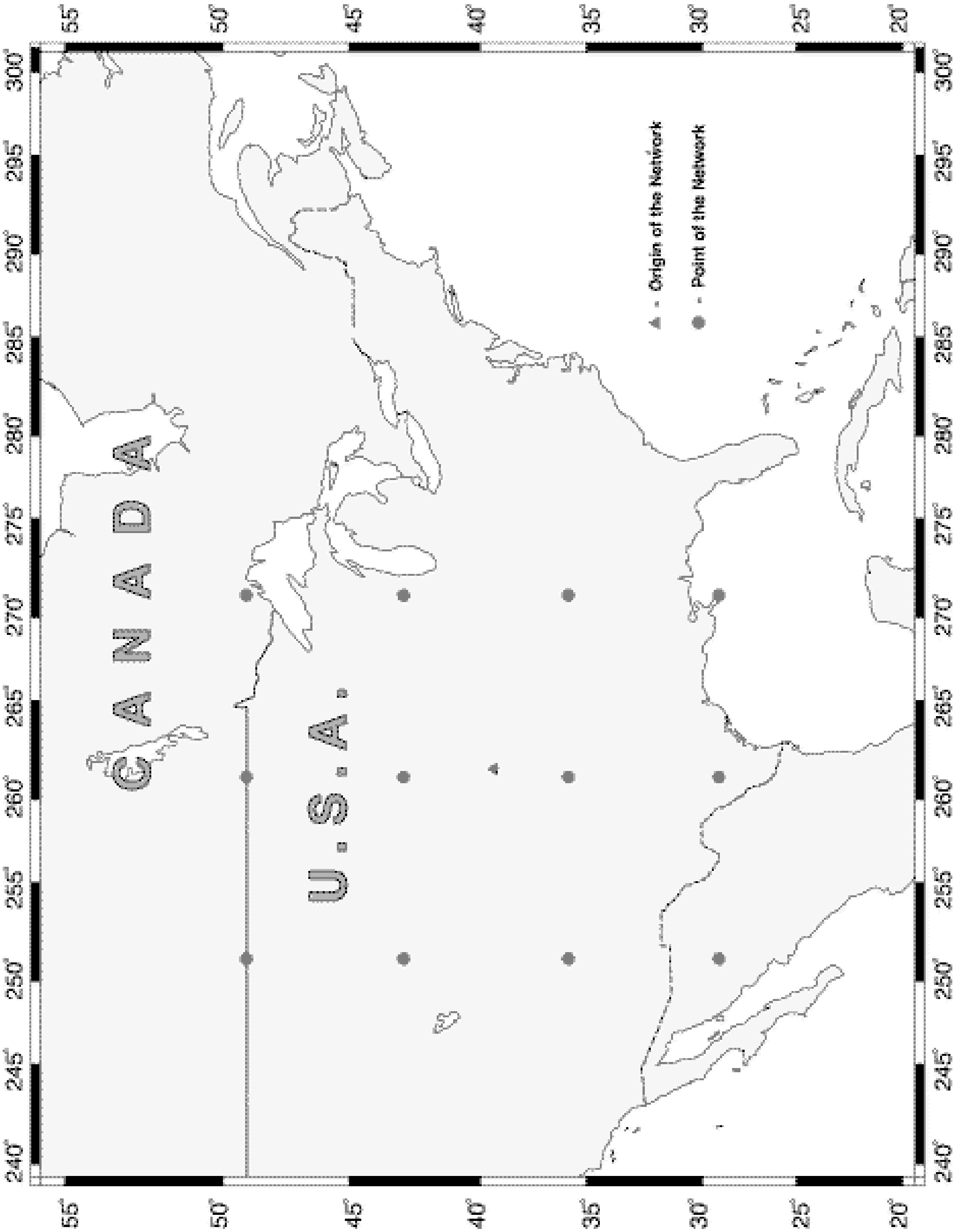

Figure 1: Configuration of the simulated network.

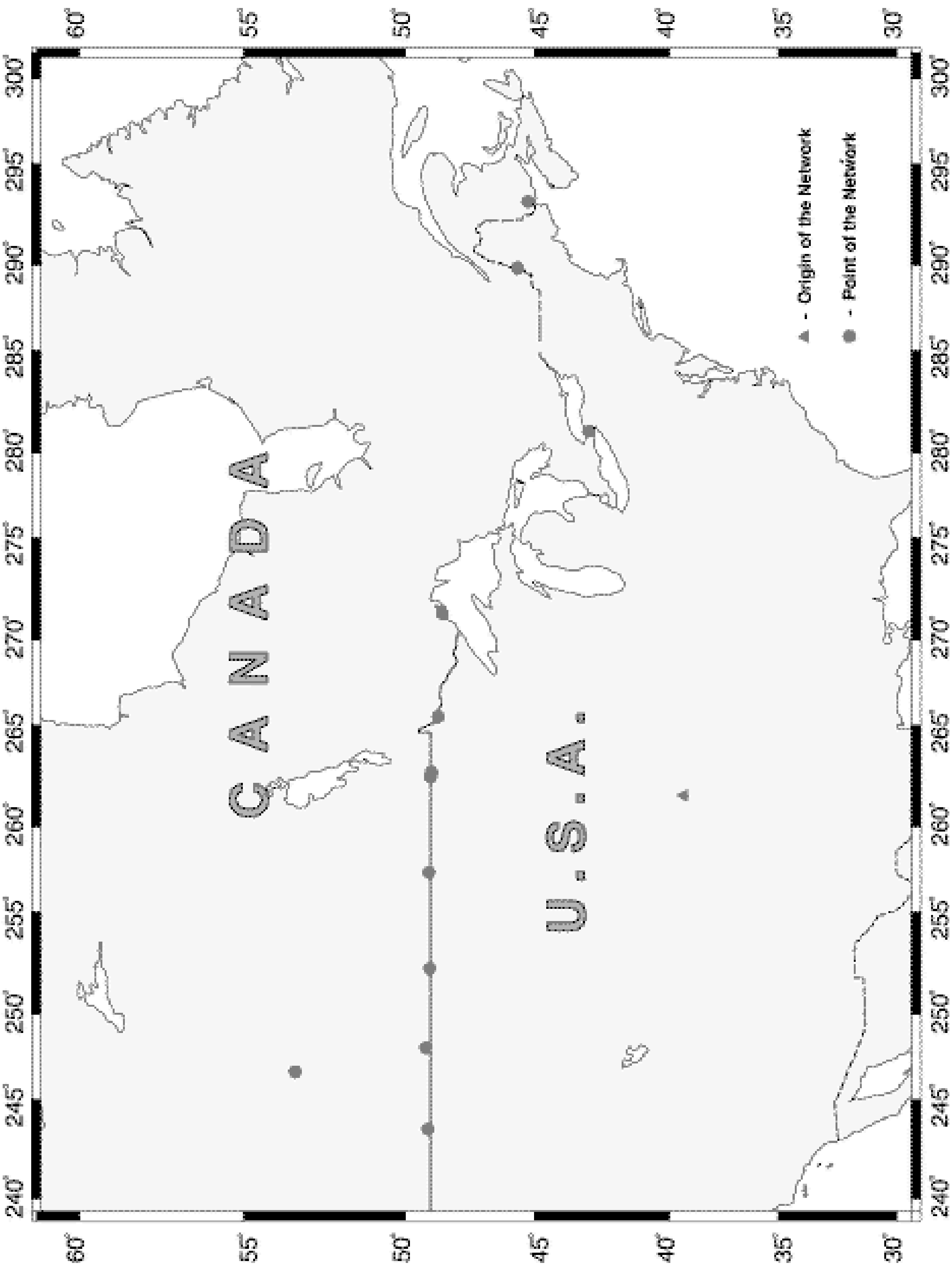

Figure 2: Configuration of the real network.