Use of the Spatial Reference Object Model to enhance Projection and Datum Transformation

Moderated Paper Session "Database Creation/Consolidation and Business Practices"
Wednesday, July 9, 2003, 1:30 PM-3:00 PM, Room 33-C (SDCC)
Esri User Conference - Paper 376
San Diego (USA) July 5 – 12, 2003

J.H.M.Hendrikse
GIS Software Development
ITC - International Institute for Geo-Information Science and Earth Observation
Enschede, The Netherlands

Abstract

This paper intends to treat the use of the Spatial Reference Object Model to enhance Projection and Datum Transformations in ArcGis.

In the daily education and research experience in our institute (ITC, The Netherlands) we are often faced with problems about how to combine information from existing paper maps with data from GPS-readings or with scanned maps or satellite imagery. The correct integration of these pieces of information (e.g. in ArcMap, ArcPad and many other GIS tools) requires a correct matching of the data as far as geographic location is concerned.

Maps from different sources, from different types and different years do often miss the needed transformation parameters to convert them accurately into a more up-to-date form or format. Projection parameters are often incomplete, let alone the parameters of the Datum transformation and the underlying spheroids. The author will show software solutions that were developed on the demand of GIS users.

The users were from different fields: map-production, education and research. The solved questions are related to projections and unknown datum parameters. The solutions require ground-control input and in certain cases a proper geoid-height model if available.

Introduction

In ArcMap there is no tool to find or compute Datum transformation parameters from measured or observed control data. In the tool described hereafter, I expect two pointfiles (in esri shape format) with 3D geographic coordinates (Lat, lon, height), each of them defined on a different but known spheroid. Common points (pertaining to identical locations) are used to calculate the Datum transformation parameters. The user can select either of the two control point layers; the selected one will serve as the 'source map' whose correction to the spatial reference of the other layer will be calculated. Optionally this transformation is carried out in the current data frame.

Motivation of this Datum shift tool

A few years ago, the National Topographic Service in The Netherlands (TDN) wanted to use about 300 topographic (military) map sheets (scale 1:25000), established in 1944, by the Allied Forces, while liberating Westeren Europe. These maps contained locations of war explosives, known at that very moment. The map grid was defined in the old French Lambert projection on the Du-Plessis spheroid, with known projection parameters but unknown, or at least uncertain Datum parameters. This system is known as 'Nord de Guerre'

It was the wish of TDN to integrate these data with the actual Dutch topographic vector data defined in the current Dutch coordinated system (stereographic double-projection on the Bessel spheroid), known as RD or Amersfoort Datum. And subsequently they wanted to convert the data to utm coordinates in the WGS84 or ETRS89 system.

I got the request to find a reliable set of datum transformation (Molodensky) parameters by comparing old and new topographic maps. My solution was to use stable ground control points that had not changed during the last 50 years, like church towers and corners of canal dikes. For these points I used the height information from the old maps, assuming heights in the Nord de Guerre compatible with those on the actual Bessel spheroid.

Figure 1. *Control points to find Datum parameters*

My method was: finding the 3 Molodensky shift parameters between un-projected points of the two systems. The control points in both systems are given as latitude, longitude, height triplets. These are input to the Molodensky equations to produce an overdetermined system with the shifts dx, dy and dz as unknowns. The methods gave satisfactory results. (See Figure 1)

A similar solution has been adopted to find the best Datum shift parameters for different regions in Mozambique. Again the local spheroid was known but the Datum Transformation parameters between the local system used in national topographic maps and the GPS measurements defined in the WGS84 system, where not accurate enough.

An extra motivation to make available a user-defined Datum transformation is the fact that the parameters of local datums provided by national survey departments become increasingly accurate (sub centimeter) And these parameters are subject to new updates every now and then.

The European countries for instance will make more and more use of the transformation parameters with respect to the ETRS, because of its independence of the relative movement of the Eurasian and American tectonic plates. This causes slight changes with respect to the parameters related to the WGS84 global Datum.

An interactive datum, defined by ground control, can be an interesting tool for cross checking the accuracy of provided datum information

---

Geodetic background

Geographic information only makes sense if we know to which location it belongs. Coordinates with respect to a well-defined reference surface are the most precise tools to determine such locations on our planet.

The mathematical ellipsoids of revolution, often called ‘spheroids’ (defined by Everest, Bessel, Clarke and other geodesists) have been for many years the shape that suits best as an approximation of the horizontal plane on which topographic features are projected to form the basis of maps. The exact definition of terrain point positions with respect to the surface and symmetry axes of such spheroids is one of the main tasks of national and international geodetic organizations.

Each nation can adopt its own spheroid that matches best the equi-gravitational zero-level surface (a surface of constant elevation). It defines the Datum for that nation. But datums can also be defined for parts of countries; e.g. North- central and south Mocambique, each have their local datum definition.

On the other hand datums can be defined for groups of adjacent countries (European ED50, North American NAD83, South American PSAD56), for continents (ETRS89) and for the whole world (WGS72, WGS84).

The mathematical relationships between different spheroids are expressed in the form of so-called datum-transformations (or simply ‘datum-shifts’).

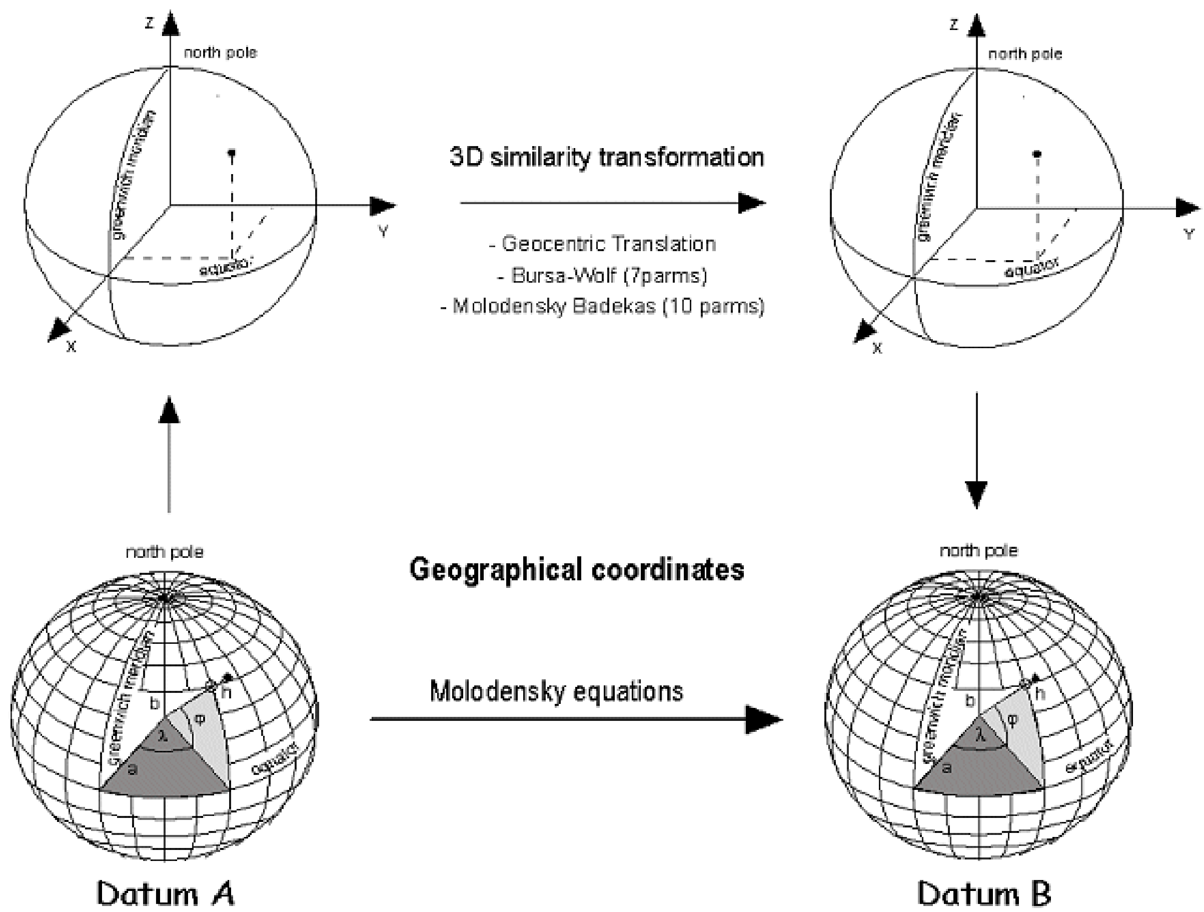

Figure 2. *Geocentric and Geographic Coordinate Systems. The latitudes in the latter are sketched as geocentric (not elllipsoidal)*

To correctly understand the datum-shift phenomenon, one should distinguish three types of coordinate systems: Geocentric, Geographic and Projected systems (as is done in the ESRI-software, where the latter two are present in almost every mapping task).

1. Geocentric coordinates

The 3 dimensional axes frame for this system is defined by the rotation axis of the earth (z-axis), the prime meridian (x-z plane) and the equator (x-y plane). This system is at the basis of all satellite positioning calculations. (See Figure 2)

2. Geographic coordinates (ellipsoidal)

These are mathematically spoken, 3-dimensional polar coordinates: two angles Phi and Lambda (latitude and longitude) and a scalar: H, the height above the chosen spheroid. In contrast to usual polar coordinates, the angle Phi is not the angle that the position radius makes with the equator plane (the ‘geo-centric’ latitude), but the angle between the local vertical and the equator plane(the ellipsoidal latitude)

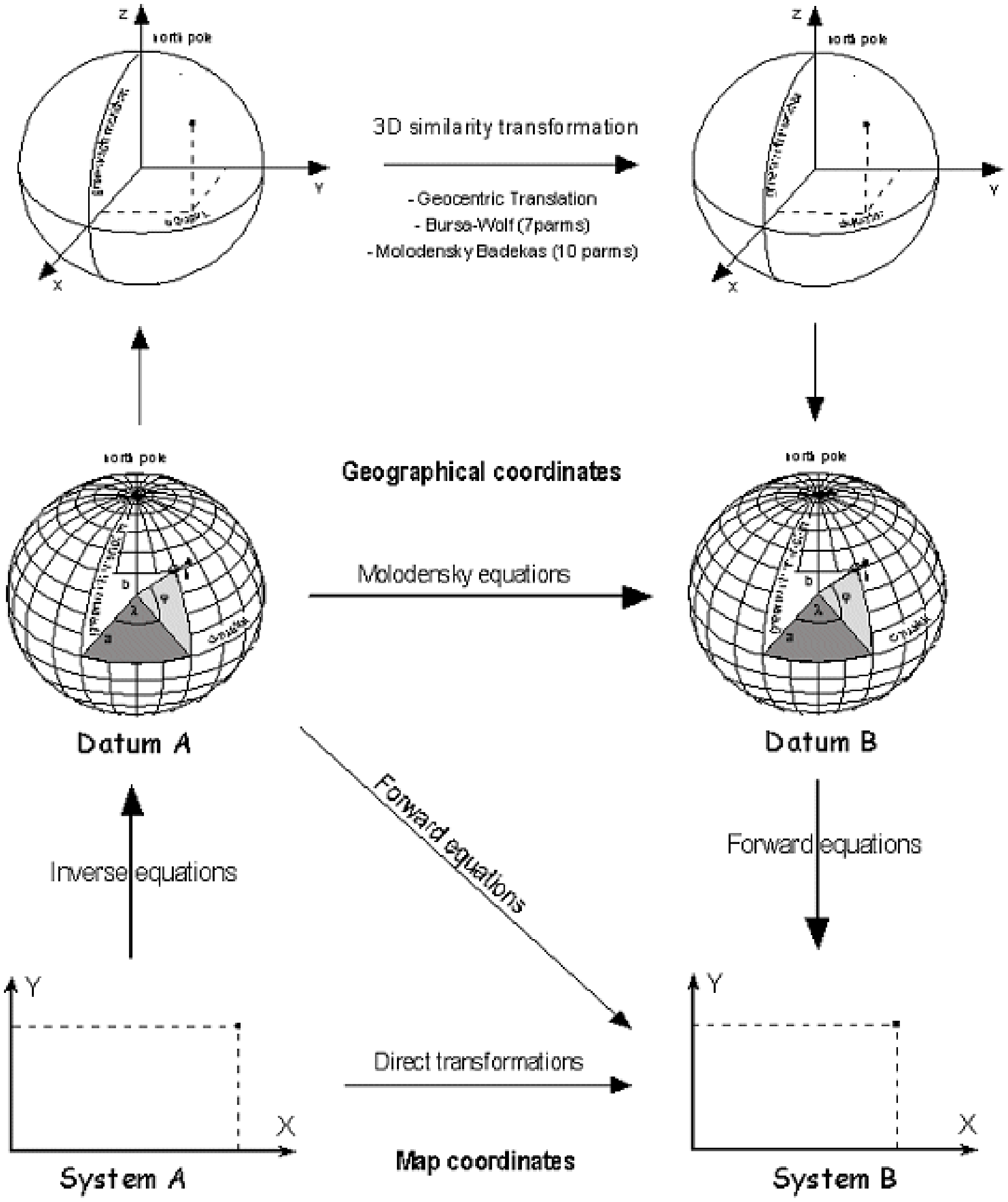

Figure 3. *Geocentric, Geographic and Projected Coordinate Systems*

3. Projected coordinates

These are 2-dimensional Cartesian coordinates defined in a map projection grid. The map projection is defined by a set of two ‘forward’ and two ‘inverse’ equations. The former convert geographic (Phi, Lambda) into rectangular (x,y) coordinates, the latter do the reverse. Examples are Gauss-Krueger, Mercator, Lambert conformal conic, Stereographic, Albers equal-area conic, to mention a few of many hundreds of projections invented in the past.

In my ArcMap tool described below, I’ll make no use of map projections because it can be left aside in the datum transformation problem. Datum transformations are essentially conversions between two different geographic coordinate systems, defined on spheroids with different position and (possibly) different shape and size.

Molodensky transformation

The relative position of two spheroids S1 and S2 can be assumed to differ only by a translation, such that the spheroids principal axes are supposed to be pair-wise parallel. Under this assumption and the possibility of different spheroid sizes (a1 <> a2) and/or different shapes (f1 <> f2), one can make use of the Molodensky or Abridged Molodensky equations to convert the ellipsoidal coordinates from S1 to S2 and back. **[1]**

Helmert transformation (similarity)

The relative positions between S1 and S2 involve also a non-parallelity (‘tilt’) of the axes-frame. Neither the rotational axis nor the prime meridian plane are supposed to be parallel. The transformation between two such systems involves three translation components Tx, Ty, Tz and three rotation angles Rx, Ry, Rz. These parameters define a transformation consisting of a 3D input vector multiplied by a composite rotation matrix and added to the given translation vector. **[2]**

In practice the transformation is extended with a seventh parameter, a scale difference DS (or scaling factor 1 + DS), that accounts for imprecision in the results obtained from the terrestrial geodetic networks, that have always been more accurate in their angular measurements than in the distance measurements. This makes the Helmert transformation a similarity (or linear conformal transformation) between two 3D vector spaces.

Moreover, the geodetic practice in case of the very small rotation angles that are used, proves that if we replace sines by the angles themselves and cosines by 1, and we neglect squares of sines, the rotation matrix becomes linear in the rotations. Thus it is well-suited for finding the seven parameters by a least squares approximation method. **[2]**

This 'linearized' Helmert transformation has the extra advantage that its inverse formulas are directly obtained by assigning an opposite sign to each-one of the 7 parameters of the transformation. The Helmert transformation, when used as a Datum conversion tool, comes in two variants:

Bursa-Wolf formulas

The Helmert equations assume one of the coordinate space-origins (of S1 or S2) as the pivot point of the rotations.

Molodensky-Badekas formulas

The Helmert equations assume a local centroid (e.g. a point in the middle of the area of interest) as the pivot-point of the rotations. These formulas differ from the Bursa-Wolf formulas only in the three translation components, rotations and scale change being identical if the same two datums are linked.

It also means that one might derive any Molodensky-Badekas translation from its corresponding Bursa-Wolf version, once a centroid has been selected. The different translations for 'BW' and for 'MB' versions can be found from each other using the rotation matrix.

The Molodensky-Badekas form has its merit mainly in a greater accuracy when defining the 7 parameters from two sets of control points. That's why it is the preferred method for the calculation of the parameters. The scale difference and rotations found, are immediately applicable as parameters in the ArcMap user-interface. The translations for the Bursa-Wolf case must be recomputed, using the displacement of the pivot (in fact its 'position vector') in addition to the translations found for Molodensky-Badekas.

'Position Vector' versus 'Coordinate Frame' method

In the current Esri software a distinction is made between the position vector method and the coordinate frame method. Both are using the Bursa –Wolf formulas, with opposite angle-orientation convention. I'll follow the position vector convention, being applied in The Netherlands and complying to my tested examples. In the tested maps the point-sets will be defined in two shape files and be loaded as two layers in a common dataframe of a map document in ArcMap.

It is possible to solve the datum transformations equations for their unknowns parameters (three shifts and if needed three rotations and a scale-change), if sufficient equations can be set up from the 3D geographic coordinates of given common points.

Thus I have customized it, using VBA with the ArcObjects possibilities. The transformation parameters are first computed from discrepancies in two shapefiles having point-sets with points at corresponding locations but slightly shifted due to two different datums on which they where projected. The two datums have each their own spheroid but unknown position with respect to WGS84.

The use of ArcObjects in ArcGis 8.3 and ArcMap

ArcObjects and Interfaces [3]

The ArcObjects concepts allow to activate and manipulate data-members, like point coordinates, and methods, like setting a reference spheroid, via interfaces to classes (more precisely to instantiated objects of these classes).

VBA is built in the ArcMap program to realize the use of these interfaces. The classes, co-classes and the needed interfaces are found in OMD's (Object Model Diagrams) that come with the ArcObjects developer Kit. The classes are reachable via the following interfaces:

- IMxDocument, IMaps, IMap, IFeatureLayer, IFeatureClass, to use and manipulate map-layers and point-features

- ISpatialReference, IGeographicCoordinateSystem to approach and use the geodetic reference systems, ‘setting’ or ‘getting’ spheroids and their size parameters, needed for my custom coordinate conversions.
- IGeoTransformationOperationSet to manipulate the predefined coordinate transformations.
- IGeoDataset to link the information from the created coordinate system to the FeatureClasses by so-called Query interface statements (comparable to dynamic typecasting)
- FeatureClass.ShapeType is a method (available via IFeatureLayer) to check whether point layers (and not lines or polygons) are selected, to compute the Datum parameters.
- IGeometryDef to check whether 2D points with height attribute or real 3D points are included in the input data layers
- IMolodenskyTransformation as special case of IGeoTransformation (see Figure 3) to put the calculated parameters
- ISpheroid to get the ellipsoids' size and shape (See Figure 4)

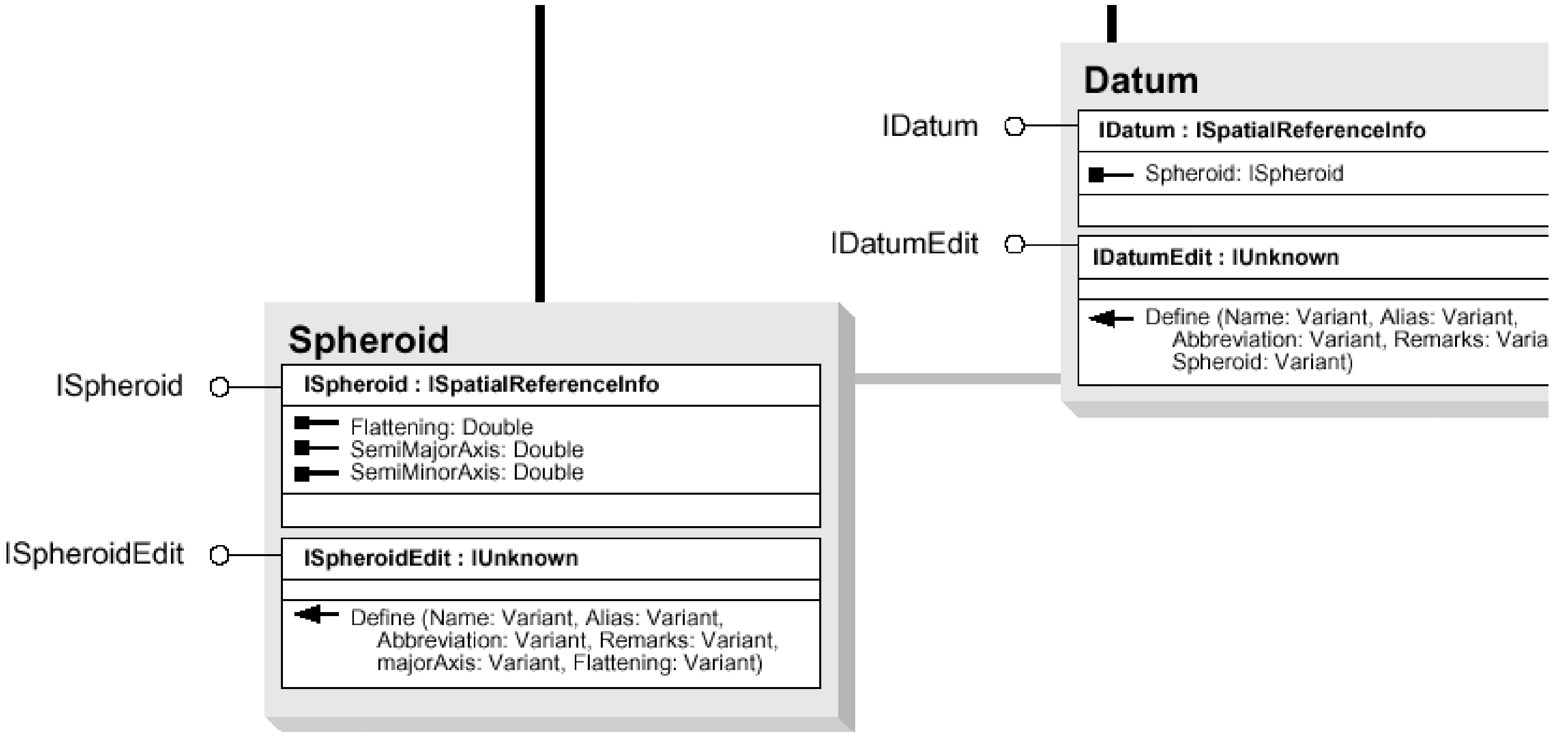

Figure 4. *Portion of the Spatial Reference Object Model Diagram to specify ellipsoids*

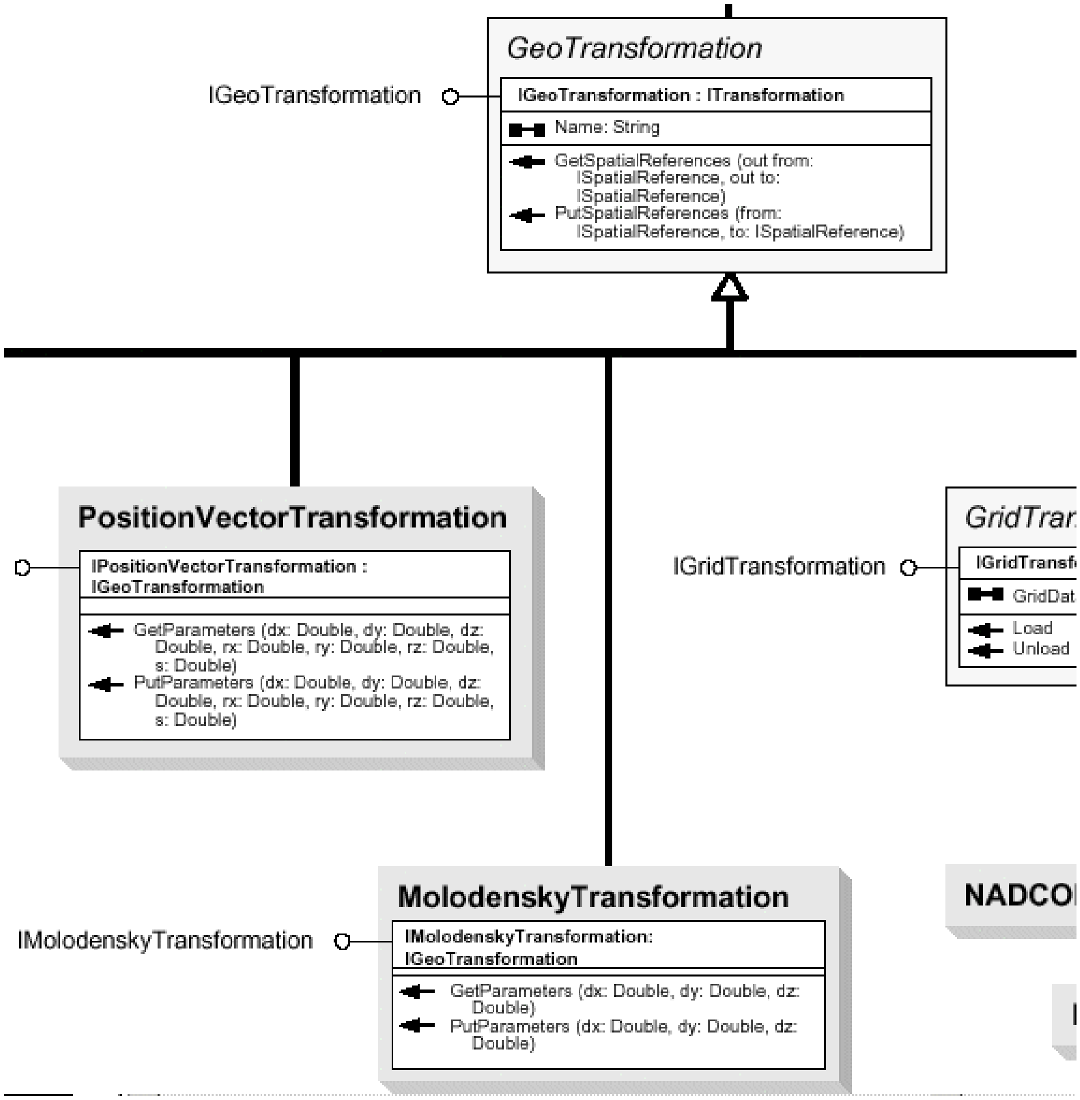

Figure 5. *Portion of the Spatial Reference Object Model Diagram to specify transformations*

Use the user-made Class concept of VBA

A class that is heavily used in the calculations of my tool is the ‘struct’ LatlonHeight, a 3D ellipsoidal coordinate that has ‘get’ and ‘set’ methods, spheroid parameters and a conversion function to find its geocentric equivalents: LLH_to_XYZ This subroutine wants a LatLonHe and returns an XYZ in the proper ellipsoidal system. It is defined in a separate Class module and could be exported to a file and possibly reused. The class is useful for all datum transformation algorithms.

Calculation and Accuracy

In the example point set, 25 points where defined on the Bessel spheroid, all with ellipsoidal height zero and spread over an area of about 200 by 300 km inside the Netherlands. (See Figure 1) The points have been transformed to the WGS84 datum, using the 7 Bursa-Wolf parameters, provided by the national geodetic commission (publicatie 30, Delft 1997). The transformed ellipsoidal latitude and longitude causes shifts in the 25 points from 150 to 170 m, if converted to metric distances in the terrain. See Figure 6.

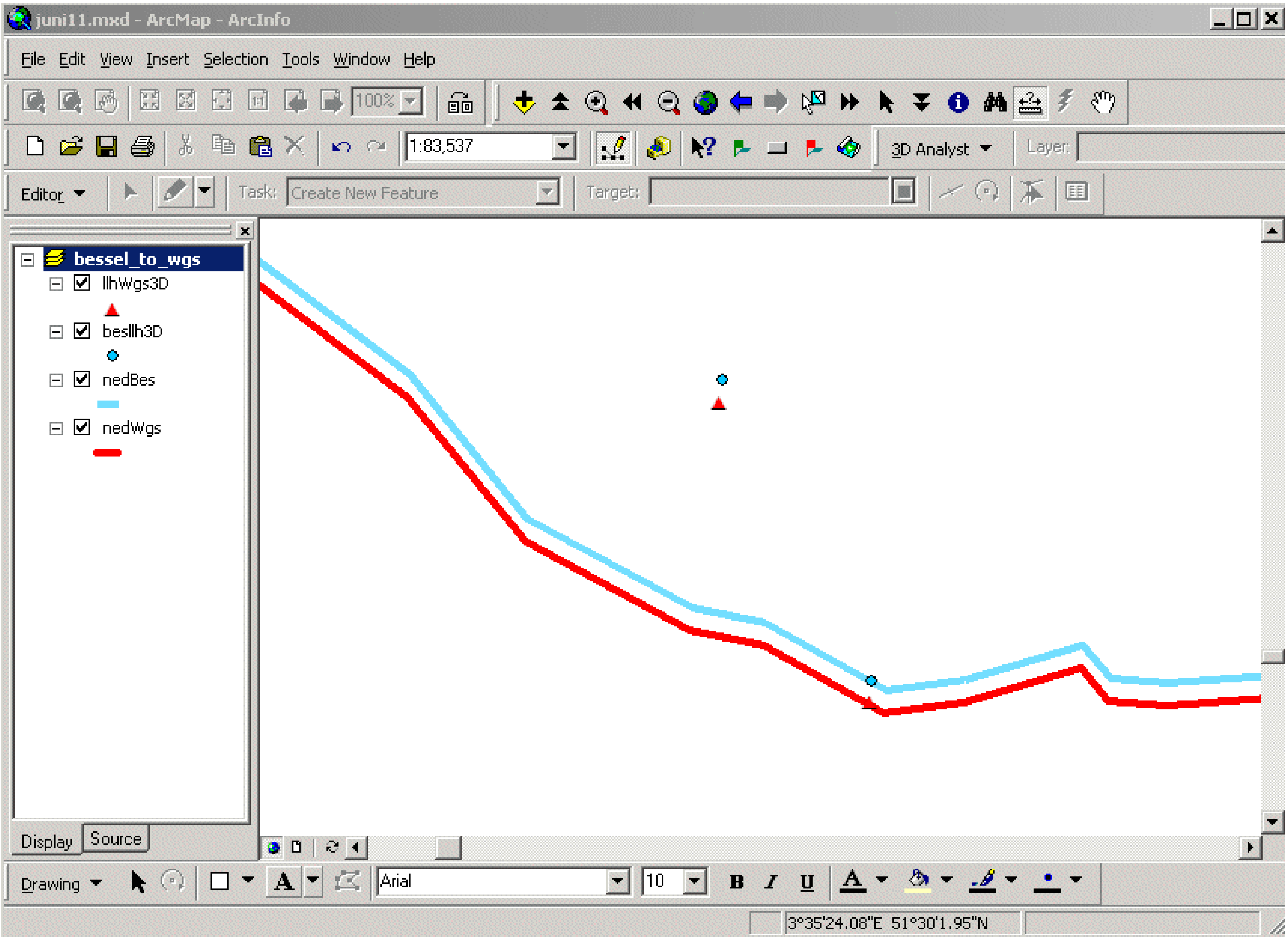

Figure 6. *Two control points and a coastline exhibiting a Datum shift of 167 m, before correction*

The Molodensky parameters are determined using 18 control points out of a set of 25. The remaining have been used as check points for evaluation. All control points are converted to corresponding 3D geocentric coordinates on the two respective spheroids. The Molodensky parameters are approximated by expressing the centroid of the points (‘center of gravity’) from the Bessel system in terms of WGS84 cartesian coordinates. This yields good estimates for the three Molodensky shifts.

With the parameters thus found, the datum correction, programmed in VBA for ArcMap has been applied. The remaining discrepancy between the control points is everywhere less than 30 mm. An inaccuracy that one can expect because the Molodensky parameters are an approximation of the translation components in the Molodensky-Badekas parameter set, using the centroid of the control points as rotation pivot.

Storing the calculated transformation

Because the SpatialReferenceEnvironment has factory functionality, it can be used to find and use predefined Geographic transformations. If one stores the transformation definition (including the newly computed parameters) in a file on disk called Geogtran, this transformation can be found through the Interface ISpatialReferenceFactory2, under the restriction that the user provides its computer system with a system variable PEOBJEDITHOME that points to the path on disk, holding the Geogtran file.

Possibilities for input and use

Choices

1. Choice between 2D-shapes with height attribute and 3D-shapes
2. PointLayer with different nr of points per layer, the first 18 points are used and matched
3. Use the geogtran file (apply 1st transformation after updating it)
4. Once the parameters have been written in the geogtran file, they can be re-used via the ArcMap User interface

User Interface Tool - Form

1. Special Button in Toolbar (included in mxd file) to activate the form
2. Title in Form Caption; Datum transformation Parameters
3. Combobox: to select one dataframe
4. Frame Info button: to check dataframe's Spatial ref and nr of loaded layers
5. Listbox containing loaded Layers to select one from
6. Layer Info button displaying Layers's Spatial ref and shape type
7. Find Molodensky shifts button, to start calculation
8. Find Bursa-Wolf params button, to start calculation
9. Write parms to geogtran file, to store the parms found, for later use
10. Apply Transformation button, to execute the transformation directly (on-line)
11. Undo transformation button, to undo the transformation
12. Close button

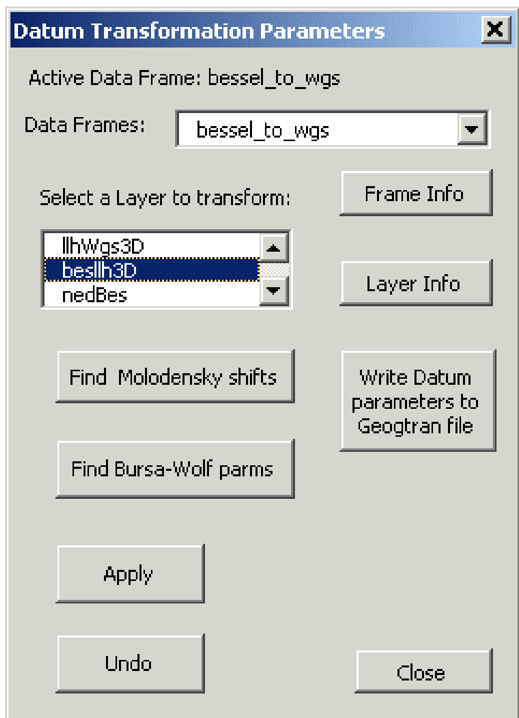

Figure 7. *The User Interface Form to select control layers for Datum shift computation*

Limitations: (and suggestions for future improvements)

1. Updating the geogtran file means deleting earlier geographic transformations found. This can be avoided by changing the path in the source code
2. Projected coordinate systems in input layers are not (yet) allowed. The ArcObject interface IProjectedCoordinateSystem and the methods Forward and Inverse enable the conversion of the planimetric coordinates
3. Real inverse Molodensky, finding the shifts from changes in Phi, Lambda and H directly is possible (C++ source available on demand)
4. Bursa-Wolf and/or Moledensky Badekas solutions using 7 x 7 matrix inversion are not yet possible (C++ source available on demand)

Source code

Source Code can be found in

http://www.itc.nl/ilwis/downloads

http://www.itc.nl/ilwis/downloads/tools/geodeticTools.asp

References

References on the Internet

[1] Nima USA,http://www.nima.mil/GandG/tm83581/toc.htm

[2] Posc USA, EUhttp://www.posc.org/Epicentre.2_2/DataModel/ExamplesofUsage/eu_cs35.html

[3] Esri USA,http://arcobjectsonline.esri.com

J.H.M.Hendrikse
GIS Software Development
hendrikse@itc.nl
ITC P.O. Box 6
7500 AA Enschede
The Netherlands