Chapter 98

An Analysis of Errors Introduced by the Use of Transformation Grids

F.G. Nievinski, M.C. Santos
Department of Geodesy and Geomatics Engineering
University of New Brunswick, P.O. Box 4400, Fredericton, New Brunswick, Canada, E3B 5A3

Abstract. In this paper we analyze the errors introduced by the use of transformation grids. A transformation grid is an intermediate step during the transformation of coordinates of points attached to distinct geodetic reference frames. The use of transformation grids simplifies the transformation when compared to the reference transformation that the grid represents.

Transformation grids have become a standard way of making transformation distortion models available for end-users. They are widely accepted by the GIS industry, being already supported by a host of commercial and free programs. They are adopted in countries like Canada and Australia, and are currently being considered for adoption in Brazil.

The work described in this paper was conducted by addressing a number of questions in the following sequence: (i) “Is there an upper bound in the error introduced by a transformation grid?”; (ii) “What is the coarsest spacing between nodes for a transformation grid to introduce only negligible errors?”; and, (iii) “How does the error introduced by a transformation grid vary spatially?”

To answer these questions we transformed a set of random test points twice, once using a transformation grid and once using the reference transformation that the grid represents. Then, we analyzed the difference between the two results. We show that: (i) yes, there is an upper bound in the error introduced by the grid; (ii) the coarsest spacing can be found by plotting error versus distance to nearest grid node; and, (iii) the error varies spatially partially in proportion to the norm of the second derivative of the reference transformation.

Keywords. Geodetic reference systems and frames, distortion modeling, transformation grids.

1 Introduction

Geodetic reference frames have intrinsic distortions due to the positioning techniques used for their materialization. The ones materialized by classical terrestrial techniques have their distortions due to the surveying techniques employed in the past. Modern, satellite-based ones have considerably less distortions. When one establishes relationships between reference frames, distortions should be taken into account. The modelling of distortions becomes even more important when relating coordinates of points between an “old” reference frame (materialized by classical techniques) and a “modern” reference frame. The distortions in the former should be modelled and taken into account.

The modelling of the distortions in the materialization of a reference frame provides better transformation results, but it makes it harder for non-expert users to apply them, especially if those users are using third-party software tools over which they have little control. That is because, as there is no commonly used form for transformations involving distortion modelling, popular software tools usually do not support them.

Transformation grids, sometimes also called “grid-shift” files, represent the field of shifts (the total shift, not the distortion only) in coordinates between two reference frames. It plays the role of a facilitating step in the transformation involving distortion modelling. The surface representing the shift field does not need to be generated many times, but only once. The grid will correspond to this surface.

Usually the procedure to use the grid is simpler than the reference transformation that was used to generate it. That is the reason why the grid makes it easier for users to apply the transformation. Also the procedure to use the grid is standardized, which makes it easier for software developers to support it in their programs.

The aim of this work is to analyze the errors introduced by the use of grids in the transformation between geodetic reference frames. The work has been conducted by addressing a number of questions, in the following sequence: (i) “Is there an upper bound in the error introduced by a transformation grid?”; (ii) “What is the coarsest spacing between nodes for a transformation grid to introduce only negligible errors?”; and, (iii) “How does the error introduced by a transformation grid vary spatially?”

This work was developed to support the transition from a classic non-geocentric reference frame to a modern geocentric reference frame in Brazil. We hope that by sharing our experience interested people or agencies facing a similar transition can benefit, in the same way we have benefited from work conducted to support similar transitions elsewhere in the world.

In Section 2 we present a background discussion about transformation of coordinates between geodetic reference frames in general, with a specific focus on transformation grids. In Section 3 we describe the data and techniques used in our analysis. Section 4 contains a discussion on the results obtained. In the final section we summarize the findings of this work.

2 Background

2.1 Transformation between geodetic reference frames

A similarity transformation is sufficient to transform points between defined reference systems (not frames[1]). A tri-dimensional similarity transformation is fully described by seven parameters, as follows: one translation in each of the three Cartesian directions, one rotation around each one these directions, and a uniform scale factor. Although the scale factor is not necessary to transform one reference system onto another (Vaní ek and Steeves, 1996), it is needed to transform *points* between reference systems.

As reference systems are idealized abstractions, we have access to them only through their materialization, called reference frames. Geodetic reference frames have intrinsic distortions due to the surveying techniques employed in their materialization. Distortions may, for instance, increase the farther away the point is from the origin, in the case of classical reference frames (IBGE, 1996). As we transform coordinates of points between two reference frames the distortions in one of them can be taken into account (assuming that the distortions in the newer, e.g., satellite-based one, are negligible compared to the distortions in the older, classical one). The modelling of distortions is important so that they do not propagate through the transformation from the old to the new reference frame, and are added back when transforming from the new to the old one, in order to guarantee a one-to-one correspondence between points in them.

Underlying the use of the similarity transformation there is the assumption that the distortions are non-existent or negligible (Collier, 2002). With the greater accuracy achieved by the use of geodetic space techniques (e.g., GPS), that assumption is no longer valid, particularly when the frames are of different nature, e.g., a classical and a space-based one. In this case, the similarity transformation alone is not enough. We have to either augment it by introducing a component to model the distortions, or choose a different transformation model.

There are many models of transformation between geodetic reference frames with distortions modelling. Collier et al. (1998) provides an overview on the problem. Junkins and Erickson (1996), Oliveira et al. (1998), and Costa et al. (1999) report additional investigations. In the general case, the transformation between coordinates can be treated as an interpolation problem (Wolberg, 1999), and then the various general spatial interpolation models (e.g., inverse distance weighting, splines, Kriging, etc.) can be applied. We have not found in the literature a commonly used model for distortion modelling.

2.1 Transformation grids

As mentioned in the previous section, there is no one common model for the transformation between reference frames with distortion modelling. That represents a challenge for non-expert users who use a variety of CAD, GIS, and Image Processing software tools provided by third parties. These tools usually offer a limited set of predefined transformation models. The inclusion of a different model is beyond the regular use of the tools, usually requiring modifications at the source code level.

One way to overcome this challenge is to introduce an intermediate form in the transformation, whose format and usage are standardized. That intermediate form is the so-called transformation grid. No matter what transformation was employed to generate it, the transformation grid can

[1] We are following the IERS usage of the terms reference systems and reference frames (see, e.g., McCarthy and Petit, chapter 4).

be used always in the same way. By "standardized" we mean that there is a specification describing it. Therefore, if we generate a grid complying with the specification, automatically our grid will be supported by the tools developed earlier that already comply with it.

The transformation grid is an array of shifts. At each node of a two-dimensional regular array the shifts in the two horizontal coordinates are given. These shifts should be added to the "old" coordinates in order to get the "new" ones. The positions of the nodes are described in the old reference frame.

The procedure to transform coordinates using a transformation grid is as follows: (1) find in the grid the four nodes nearest to the point to transform; (2) apply a bi-linear interpolation to the shifts at those four nodes; (3) sum the interpolated shifts to the old coordinates of the point to transform. The result is coordinates of the same point in the new reference frame. By an iterative procedure the same grid can be used to do the inverse transformation, i.e., from new to old coordinates.

The grid is generated by evaluating the reference transformation procedure at each of its nodes, and then computing the difference between the transformed (i.e., new) and the old coordinates. Once generated, the grid needs to be formatted in a standard way (see, e.g., Mitchell & Collier, 2000). It can be coded first in free text and later converted to a binary format.

Transformation grids were adopted in Canada, the USA, and Australia. This approach is independent of a specific software tool. Therefore, the tools developed to support the use of transformation grids in one country could be used in a different country. Nowadays there is a host of commercial and free programs for both the end-user and the programmer that support transformation grids[2].

3 Data and methods

3.1 Control points

In our investigation a set of 200 control points from the State of Paraná, in Brazil, was used. The geographic location of these points can be seen in Figure 1. Each control point has coordinates in two different Brazilian realizations of the South American Datum of 1969 (SAD69): the original one, and its 1996 realization (SAD69/96). We will be considering these two realizations as two different reference frames.

For each control point the shift in coordinates between the two reference frames was computed by simply taking the difference between them. The shifts in latitude are almost 10 times larger than the shifts in longitude, making the shifts in longitude hardly noticeable when plotted together. For a better visualization, we show in Figure 2 the normalized shifts[3].

Figure 1 Position of the control points.

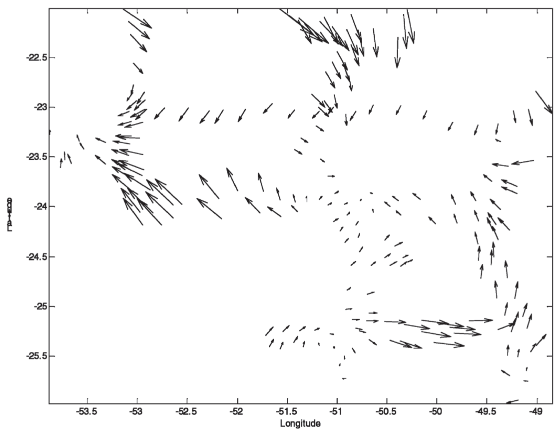

Figure 2 Normalized shifts at control points.

[2] For a list of free programs, please contact the first author; for commercial ones see ICSM (2005) and GSD (2005).

[3] From now on, whenever we show the shifts in latitude and in longitude together, we will be showing normalized shifts.

3.2 Reference transformation

To analyze the errors introduced by the grid we need a reference transformation to be used as a benchmark for the comparisons to follow. After obtaining that reference transformation we assume it as "true" or, conversely, that it does a "perfect" job modelling the observed shifts, which are then abandoned. This is valid because we are not interested in the errors introduced by the reference transformation itself, only in the errors introduced by the grid alone.

To allow us to answer the third question posed in this paper (section 4.3), we needed a transformation that would yield results with high spatial variability. Therefore we discarded simple well-behaved models such as global polynomial surfaces of low degree. Other than that, the choice was arbitrary.

The chosen reference transformation is as follows. We interpreted the shifts in latitude and longitude as two separate two-dimensional scalar fields, varying over the horizontal space. Then, we used the triangle-based bi-cubic interpolation (Fortune, 1997) to interpolate the shifts at the desired points. A sample of the reference transformation's result is shown in Figure 3. Figure 4 shows the so-called Delaunay triangulation of the control points, an intermediate result required by the chosen transformation.

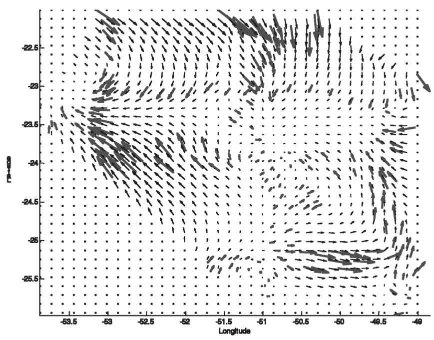

Figure 3 Sample result of the reference transformation (thin, regularly spaced arrows) against observed shifts (thick arrows).

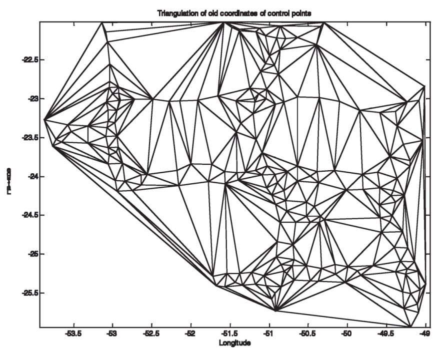

Figure 4 Delaunay triangulation of the control points.

3.3 Generation of the transformation grid

We generated an array of regularly spaced nodes (spaced 13º 30' in latitude and longitude) enclosed by the convex hull of the control points, as shown in Figure 5. Then, we evaluated the reference transformation at each grid node. The results of this evaluation are shown in Figure 6.

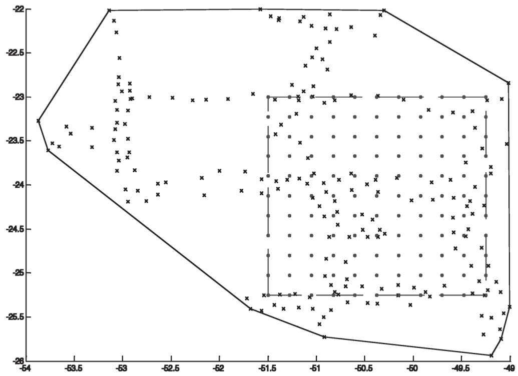

Figure 5 Control points (crosses), convex hull of control points (outer line), and grid limits (rectangular area).

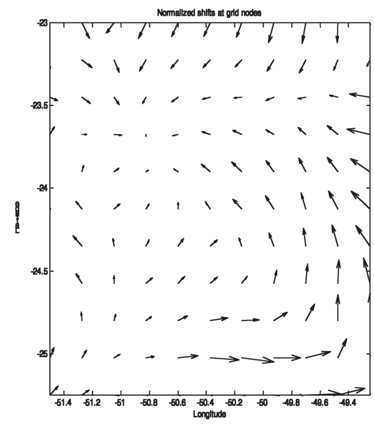

Figure 6 Modeled shifts at grid nodes.

3.4 Test points

To evaluate the performance of the grid we overlaid on it 10,000 points at random positions, as shown in Figure 7. The number of random points was chosen arbitrarily. We transformed the coordinates of each random point using (i) the reference transformation, and (ii) the grid. The difference between (i) and (ii) represents the error introduced by the grid.

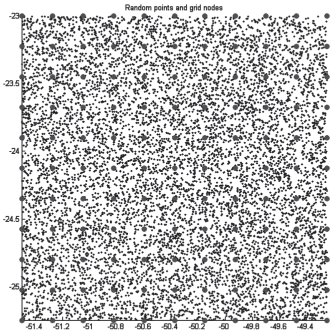

Figure 7 Random test points (thin dots) and grid nodes (thick dots).

3.5 Use of the transformation grid

We have used a program conforming to the transformation grids specification (the Canadian NTv2) to verify that the grids we were generating followed the specified text format.

The grid is to be used for interpolating bi-linearly (Press et al, 1992) the shifts given at the grid nodes. We did so with a Matlab implementation (function interp2) of that interpolator.

4 Results and discussion[4]

4.1 Is there an upper bound in the error introduced by the transformation grid?

Obviously the grid may introduce an error with respect to the reference transformation. That is so because the grid transformation model (which is made of the grid itself *and* the bi-linear interpolator to be applied on it) may not be able to capture all the variability described by the reference transformation. We would like to assess how bad this error can get.

To answer this question we analyzed the cumulative mean and maximum errors of the random test points as we increased the number of points (see Figure 8). In that figure, each dot represents an individual test point; the thin, stepped curve is the maximum error; the thick continuous line is the mean error. These error curves are *cumulative*, meaning that each value along them is calculated from the test points to the left of it.

As we sample the grid error with more and more test points, we find that there is an upper bound in the error curves. In our case, we noticed that after 5,000 test points the error stops increasing, meaning that (i) that sample is representative of the grid error, and (ii) the error is no greater than 3.5×10^{-6} degrees or 0.0126".

The upper bound on the error curves depends on a balance between grid spacing and spatial variability of the reference transformation results. The next two questions below address the problem of "tuning" a grid so that it introduces only negligible errors, in an efficient manner.

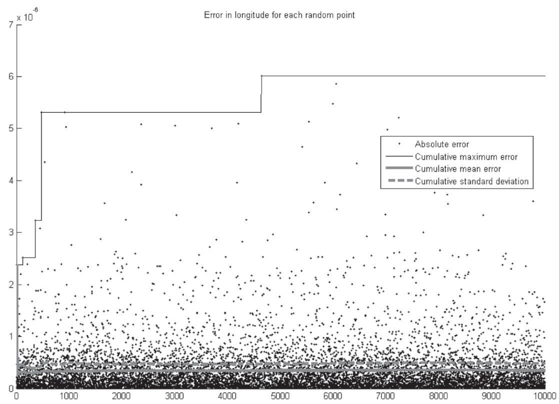

Figure 8 Absolute error at each point (thin dots), maximum cumulative error (thin, stepped curve), and mean cumulative error (thick, continuous line) versus number of test points.

4.2 What is the coarsest spacing between nodes for a transformation grid to introduce only negligible errors?

First of all it is required that the ones generating the grid define what error would be negligible in their application. This value might be based, e.g., on the error already introduced by the reference transformation itself. We will describe an example later in this section.

The question will help us to tune the grid so that the inevitable errors introduced by it do not affect the application we have in mind. Here we will assume that there is only one uniformly spaced grid covering the area of interest. In the next question we will be interested in the case of reducing the grid spacing locally (instead of globally, as we do here).

To answer the question we have obtained, for each test point, the "distance"[5] to its nearest grid

[4] Due to lack of space, in the following we have figures only for longitude. The corresponding figures for latitude show curves with similar behaviour and values 10 times larger.

[5] The "distance" is actually the Euclidean norm of the difference in geodetic coordinates between a given point and

node. We sorted the points based on that distance. Then we analyzed how the error increases as that distance increases (see Figure 9). Intuitively, the closer a point is to a grid node (i.e., the smaller that distance), the better the grid model represents the actual shift at that point or, conversely, the smaller the error is at that point.

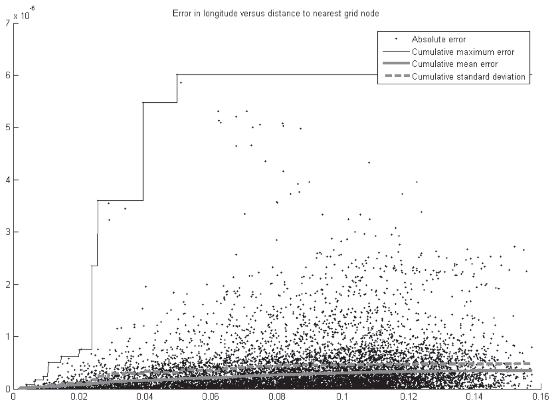

Figure 9 Error in longitude at each point (thin dots), maximum cumulative error (thin, stepped curve), and mean cumulative error (thick, continuous line) versus distance to nearest grid node.

Figure 9 can be used to find the coarsest spacing between nodes for a transformation grid to introduce only negligible errors. To do so, first, we specify the maximum acceptable error; second, we find the corresponding maximum distance to a grid node, using the error curve in Figure 9; third, we re-generate the grid using that distance times •2 (see Figure 10) as the spacing between its nodes.

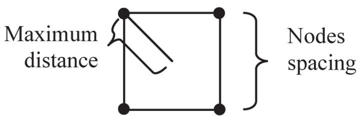

Figure 10 Relationship between maximum distance and nodes spacing (equal spacing in both directions).

As an example, the maximum acceptable error was chosen arbitrarily as 2×10^{-5} degrees, which equals 0.072". The corresponding maximum distance found in the maximum latitude and longitude error curves is 0.04º (Figure 9) and 0.025º (figure not shown), respectively. As there is only one grid, it should satisfy both latitude and longitude maximum errors. Therefore we use the most stringent maximum distance, which is 0.025º in this case. 0.035º is the spacing between grid nodes that yields 0.025º as maximum distance (i.e., 0.035º = •2 × 0.025º).

Figures 11 show the nodes of this new, denser, grid, and Figure 12 shows its corresponding error curve. Now the maximum distance is 0.025º (see horizontal axis), and the maximum error is close to 2×10^{-5} degrees (see vertical axis), as specified.

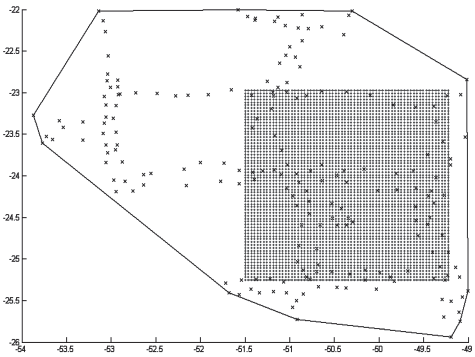

Figure 11 Nodes of the denser grid.

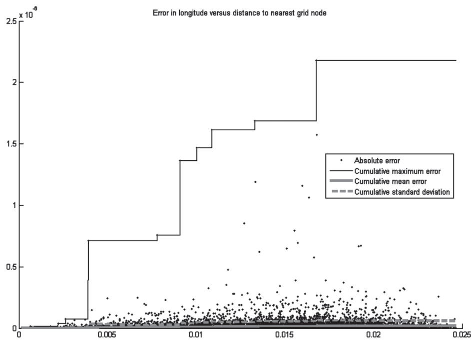

Figure 12 Error at each point (thin dots), maximum cumulative error (thin, stepped curve), and mean cumulative error (thick, continuous line) versus distance to nearest grid node – denser grid.

4.3 How does the error introduced by a transformation grid vary spatially?

In the previous question we assumed that we would be using only one, uniformly spaced grid. But transformation grids allow the use of sub-grids, i.e., denser grids covering a subset of the main grid. This property is important when a national grid is

its nearest node. It is not, e.g., the ellipsoidal distance between the two.

densified by a state or provincial grid. It is useful to be able to predict the maximum error spatially because the areas with high maximum errors would be strong candidates for sub-gridding.

To help us tackle the posed question we have used a denser[6] set of regularly spaced test data. We did so to improve visualization. The denser point set depicts the patterns in more detail, and the fact that it is regularly spaced allows us to plot the set as an image, which is a lot faster than to plot each individual point.

We expected to find a spatial portrayal of the behaviour shown in Figure 9, i.e., error increasing as a function of distance to nearest node. But what we found was the intriguing pattern shown in Figure 13. Maxima seem to concentrate near edges and vertices of the Delaunay triangulation.

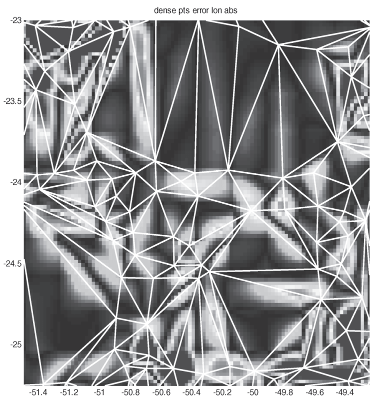

Figure 13 Error (red – large values, blue – small values) in longitude versus latitude and longitude, with edges of triangulation overlaid.

In the search for an explanation for that behavior, it was brought to our attention the existence of the following formal error bound for the 1-dimensional linear interpolation (Wikipedia, 2005):

$$|e| \le \frac{(x_1 - x_0)^2}{8} \max_{x_0 \le x \le x_1} |f''(x)|, \qquad (1)$$

where e is the error, f is the reference function being approximated by the interpolation procedure, and x_0 and x_1 define the limits of a piece of the domain on which the interpolation is taking place. The expression above shows that, in the 1-dimensional case, the piece-wise maximum error depends strongly on the absolute value of the second derivative of the reference function.

We could not find or develop an expression describing a similar dependence in the 2-dimensional case. Then we went to investigate empirically whether a similar behaviour is observed.

To do so, we calculated numerically the second gradient of shift in each coordinate, by means of the central difference numerical derivative (Press et al., 1992) in each direction. The norm of that gradient is shown in Figure 14.

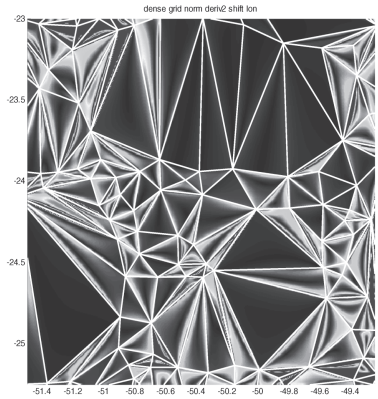

Figure 14 Norm of the second gradient of shift in longitude (red – large values, blue – small values), versus latitude and longitude, with edges of triangulation overlaid.

We see roughly the main peaks at the same positions (see, e.g., SE-, NE-, and NW-corners). Despite that, the overall matching is weak, as attested by the correlation coefficients: 0.4 and 0.3 for shifts in latitude and in longitude, respectively.

The main difference between the error field and the second gradient norm field is that the latter is better defined, with sharper variations, while the former seems like a locally-averaged version of the latter.

To verify the interpretation above, we defined grid cells, which are rectangular areas delimited by four nodes of the original grid (the grid on which the errors are based – see Figures 5 and 10). Then, we computed the maximum values of error and norm of second gradient *per grid cell*. At this time the correlation is stronger: 0.76 and 0.75 for shifts in latitude and longitude, respectively (see scatter plot in Figure 15). Therefore, in this case, the norm of the second gradient of shifts predicts partially the error per grid cell.

[6] The test points were 10 times denser than the transformation grid nodes.

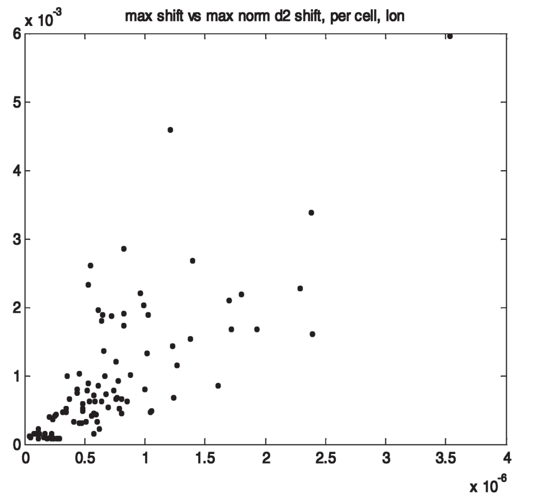

Figure 15 Scatter plot of error in longitude versus norm of second gradient of shift in longitude.

As stated at the beginning of this section, in the process of developing and tuning transformation grids, the areas with high errors would be strong candidates for sub-gridding. We should recognize that for this purpose it might be more practical to simply obtain and analyze the error field directly (Figure 13) instead of predicting it only partially.

5 Conclusions

In this paper we analyzed the error introduced by the use of transformation grids. We have shown that: (i) there is an upper bound in the error introduced by the grid; (ii) the coarsest spacing can be found by plotting the error versus distance to nearest grid node; and, (iii) the maximum error vary spatially partially in proportion to the norm of the second gradient of the shifts. We believe these conclusions and, more importantly, the analyses presented in this paper, might be useful to individuals and agencies considering, developing, or tuning transformation grids to support the transition from a classic to a modern reference frame.

Acknowledgements

Funds for this study have been provided by the Canadian International Development Agency (CIDA) in support of the Brazilian National Geospatial Framework Project (<http://www.pign.org>).

References

Australian Intergovernmental Committee on Surveying and Mapping (ICSM). *Geocentric Datum of Australia – Software*. Available at: <http://www.icsm.gov.au/icsm/gda/software.htm>. Last access on 08/May/2005.

Brazilian Institute of Geography and Statistics (IBGE) (1996). *Adjustment of the planimetric network, Brazilian Geodetic System* (in Portuguese).

Canada Geodetic Survey Division (GSD). *Commercial Software with NTv2 feature*. Available at: <http://www.geod.nrcan.gc.ca/index_e/products_e/software_e/ntv2soft_e.html>. Last access on 08/May/2005.

Collier P.A., Argeseanu V., Leahy F. (1998). "*Distortion Modelling and the transition to GDA94"*, The Australian Surveyor, Vol. 43, No. 1, March 1998.

Collier, P.A. (2002). "*Development of Australia's National GDA94 Transformation Grids*". Consultant's Report to the Intergovernmental Committee Surveying and Mapping. February, 2002.

Costa, S. M. A., M. C. Santos and C. Gemael (1999). "The integration of Brazilian geodetic system into terrestrial reference systems." Abstracts, XXII General Assembly, International Union of Geodesy and Geophysics, Birmingham, England, 19-24 July, p. A.415.

Fortune, S. (1997). Voronoi diagrams and Delaunay triangulations, *Handbook of Discrete and Computational Geometry*, J.E. Goodman, J. O'Rourke, eds., pp. 377-388, CRC Press, New York.

Junkins, D.R. and Erickson, C. *Version 2 of the National Transformation Between NAD-27 and NAD-83 and Its Importance for GPS Positioning in Canada*. Geodetic Survey Division Geomatics. Canada, 1996.

Oliveira, L.C., J.F.G. Monico, M.C. Santos and D. Blitzkow (1998). "Some considerations related to the new realization of the SAD-69 in Brazil." *Advances in Positioning and Reference Frames*, F. K. Brunner (Ed.), International Association of Geodesy Symposia, Vol. 118, Springer, Berlin, pp. 205-210.

McCarthy, D.D. and G. Petit. (2004). IERS Conventions 2003 (IERS Technical Note 32) *Frankfurt am Main: Verlag des Bundesamts für Kartographie und Geodäsie*. 127 pp., paperback, in print.

Mitchell, D.J., Collier, P.A. (2000). *GDAit Software Documentation* – Version 2.0. Available at: <http://www.geom.unimelb.edu.au/gda94/SoftDoc.pdf>. Last access on: 07/Sep/2004.

Press W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992). *Numerical Recipes in C – The Art of Scientific Computing*. Second Edition, Cambridge University Press.

Vaní ek, P., R.R Steeves (1996) "Transformation of coordinates between two horizontal geodetic datums". *Journal of Geodesy*, Vol, 70, No, 11, pp. 740-745.

"Linear interpolation." *Wikipedia: The Free Encyclopedia.* 2005. Available at <http://en.wikipedia.org/wiki/Linear_interpolation>. Last access on 01/Aug/2005.

Wolberg, G. (1990) *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, CA.