Toitu te Land whenua Information New Zealand

Development of Transformation Parameters from NZGD49 to NZGD2000

OSG Technical Report 13

Chris Crook and Merrin Pearse
Office of the Surveyor-General

16 January 2001

Contents

Foreword

Land Information New Zealand (LINZ) (Toitu te Whenua) was established in July 1996. It is a government department with roles and responsibilities in the following key areas:

Regulatory Responsibilities	**LINZ Regulatory Groups**
National spatial reference system and cadastral survey infrastructure	Office of the Surveyor-General
Topographic and hydrographic information	National Topographic/Hydrographic Authority
Land Titles	Office of the Registrar-General of Land
Setting rules for rating valuations	Office of the Valuer-General
Crown Property	Office of the Chief Crown Property Officer (Crown Property)
Assisting the government address land related aspects of Treaty of Waitangi issues	Office of the Chief Crown Property Officer (Crown Property)

The main role of the department is a regulatory one, to set guidelines and standards and manage contracts for carrying out the day to day business associated with each of the key areas.

LINZ also offers a range of services to customers related to land titles, survey plans and Crown property. Land Titles and Survey services are carried out by the Operations Group based in LINZ regional offices throughout New Zealand.

The LINZ overarching objective is to be recognised as a world leader in providing land and seabed information services.

Chief Executive:
Dr Russ Ballard
Land Information NZ
P O Box 5501
Wellington
Phone: 0-4-460 0110
Fax: 0-4-472 2244
Email: info@linz.govt.nz
Internet http://www.linz.govt.nz/

DEVELOPMENT OF TRANSFORMATION PARAMETERS FROM NZGD49 TO NZGD2000

1 Introduction

In October 1999 Land Information New Zealand (LINZ) officially announced a new geodetic datum for New Zealand – New Zealand Geodetic Datum 2000 (NZGD2000). The new datum is a geocentric, geometric datum - that is coordinates are defined in terms of XYZ positions relative to the centre of the earth, which are usually expressed as latitude, longitude, and ellipsoidal height in terms of the ellipsoid define by the Geodetic Reference System 1980 (GRS80). The implementation of this datum is described by Pearse [2000].

Before NZGD2000 geodetic surveys and most cadastral surveys were in terms of the New Zealand Geodetic Datum 1949 (NZGD49). This was calculated in 1949 using surveys dating from between 1909 and 1949 [*Lee* 1978], when it was defined in terms of the coordinates assigned to the 1st order stations. The datum was constrained by the data available - it only covers the North Island, and northern and eastern parts of the South Island. In the mid 1970s it was extended onto the West Coast of the South Island. The datum is a two-dimensional datum defining only latitude and longitude.

Although the coordinates of the 1st order NZGD49 stations are defined as being absolute and perfect, in practice they are not. There are several reasons why. At the time they were calculated they were subject to errors in the observations used, in particular in extrapolating distances measured on a few baselines into the network. Also since the earliest observations were made, there has been significant earth deformation, up to 3 metres across the country.

There are additional errors in the lower order NZGD49 coordinates that are derived from more local surveys and adjustments. The analysis in Appendix A suggests that this may be expressed in terms of the distance from the nearest 1st order control as 0.01 m + 0.003 m/km or more .

With the establishment of a new datum comes a need to transform coordinates from one datum to the other as only a small number of marks have been surveyed in terms of both datums.

Conventionally, conversions between geodetic datums have been done using a similarity transformation. This assumes that the datums differ only in the position, orientation, and scale of the reference frame on which they are defined. A similarity transformation has been recommended for transformations between NZGD49 and WGS84 (World Geodetic System 1984) by Pearse and Crook [1997]. This is equally applicable to transformations between NZGD49 and NZGD2000. However it only provides an accuracy of about +/- 4 metres, because it cannot account for the distortion in the NZGD49 datum.

For many applications the similarity transformation between NZGD49 and NZGD2000 is not accurate enough. To achieve a better accuracy requires a model that accounts for the distortion of the NZGD49 datum. There is also distortion in the NZGD2000 datum, but it is insignificant in comparison with that in NZGD49.

The approach adopted is to calculate the difference between the two datums at points on a grid spanning the country. The difference at any other point can then be inferred by interpolating from the differences at the corners of the grid cell in which it lies. The grid of differences is called a distortion grid here, as it represents the distortion in the NZGD49 datum.

The distortion model is calculated using survey marks that have been surveyed in terms of both NZGD49 and NZGD2000. At these control points the difference between the datums is known. The following sections detail the set of control points, and describe the development of a model from them.

One philosophical difficulty in this analysis is defining what is meant by an NZGD49 coordinate for a point. The problem is that at any given point a NZGD49 coordinate could be derived by observing to local 3rd and 4th order control, or by connecting to the nearest 1st or 2nd order stations, or by connecting to more remote control using GPS. Each choice of reference stations is likely to result in a different coordinate for the point. In extreme cases the differences may be of the order of metres.

For cadastral surveyors the most sensible coordinates are those that are based upon and consistent with the local geodetic control. However local control has often developed following the pattern of land usage rather than following the principles of good network design. This leads to an uneven distribution of error within the network.

2 The data set

When this study was initiated there were 2535 geodetic marks for which NZGD49 and NZGD2000 coordinates were available. The NZGD49 orders of the marks are summarised as follows:

NZGD49 order	Number of marks
1	267
2	202
3	545
4	888
5	17
6	616

For the purposes of this study the 5th and 6th order marks are not considered accurate enough, so the study uses the 1902 mark of 4th order or better.

These marks are shown in figure 1. One feature of this is the very uneven distribution of control points. This reflects the LINZ surveying programme since the establishment of NZGD2000 which has tended to concentrate on developed areas.

Figure 1: Control points used to calculate the NZGD49 distortion model. The symbol shows the NZGD49 order of the marks - solid triangle is 1st order, open triangle is 2nd order, large cross is 3rd order, and small cross is 4th order.

3 Analysis

The method used to model the relationship between NZGD49 and NZGD2000 is to cover the country with a square grid (see Figure 3) and calculate the difference between the datums at each node. Details of the method are provided in Appendix B. The main features of the method are:

- it attempts to reproduce the measured difference at each control point
- it attempts to minimise distortion that is inconsistent with the survey methods (primarily triangulation) used to observe NZGD49. That is, it allows rotation, scaling, and shifting of each grid cell, but tries to minimise distortions such as shear.
- it calculates residuals for each control mark that can be used to assess their reliability

The parameters that control the analysis are:

- the density of the grid (this is, the size of the grid cells). The final analysis used a grid size of 20km. This was chosen partly for practical reasons (finer grids required too long to calculate) and partly because almost any survey in terms of NZGD49 could be using points up to 20 km away for control, so there is little sense in attempting to model smaller scale distortions.
- the set of control points used. Control points at which the difference between NZGD49 and NZGD2000 coordinates appeared anomalous were rejected (see section 3.1).
- the accuracy specified for each control point. The model allows the control points to be weighted according to the order of the NZGD49 coordinate, as well as weighted individually. Weighting is in terms of NZGD49 order as it is assumed that the majority of the error arises from these rather than from the NZGD2000 coordinates, which have relatively high and uniform accuracy.
- the sensitivity to distortion of the grid. In effect this controls the "rigidity" of the model.

3.1 *Rejecting anomalous control marks*

The initial aim in the analysis was to identify marks at which the difference between NZGD49 and NZGD2000 coordinates appeared to be atypical. The analysis progressed from simple to more refined models, at each stage identifying and rejecting points which significantly failed to fit the model. This approach was used to avoid allowing anomalous data to excessively influence the more refined models.

The first analysis used a grid size of 50km. The residual errors (in terms of distance between calculated and measured NZGD2000 coordinates for the control marks) are shown in figure 2. It is clear from this diagram that marks with residuals over 1m are outside the typical range of errors. On this basis the 68 marks with errors greater than 1m were removed from the subsequent analysis. The rejected marks comprise 66 4th order marks, 1 3rd order mark, and 1 2nd order mark. The one NZGD49 2nd order mark was "WELL". Upon investigation it transpired that the NZGD49 coordinate of this mark was not derived from an NZGD49 adjustment - it was calculated by transforming from an NZGD2000 adjustment.

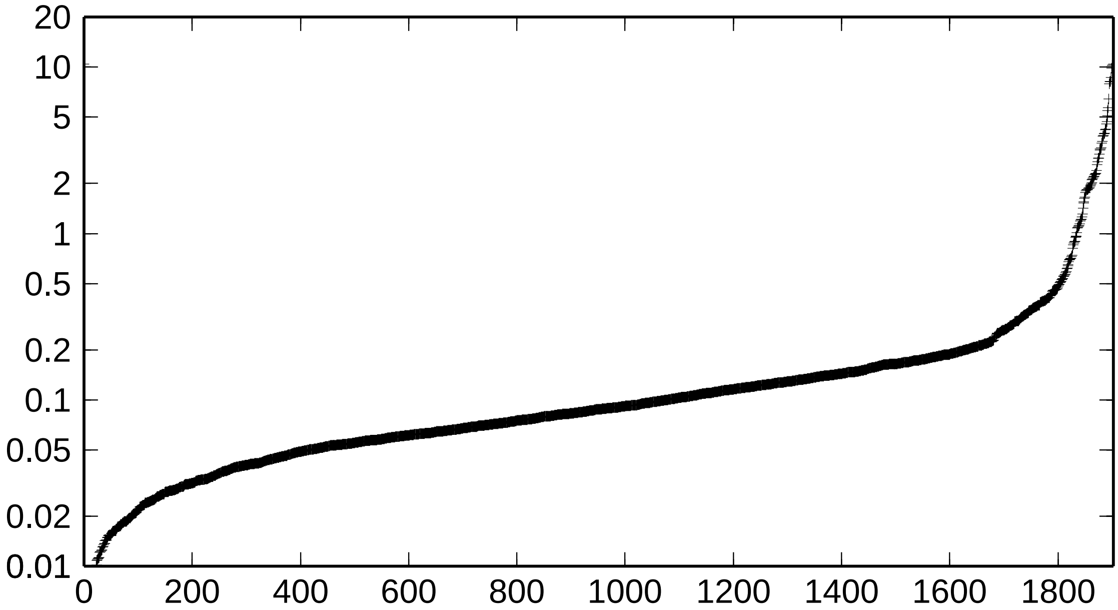


Figure 2: Plot of residuals of control marks from the initial analysis plotted in order of magnitude.

The adjustment was repeated without these points using a finer grid spacing (20km. In this model a further 23 marks appeared anomalous, with residuals of over 50cm. These comprised 4 3rd order marks and 19 4th order marks.

It was noted that 33 of the 91 rejected marks had NZGD49 order of "4c", which is defined as:

> “Coordinates determined by a hanging line fix over 100m long from a 1st, 2nd or 3rd order station. Also bench mark with geodetic coordinates obtained from unknown source.”

Since these appear to be particularly unreliable the remaining 34 order 4c marks were also rejected.

At the completion of this analysis 1777 control points remained in the data set and 125 were rejected. The high percentage of rejected marks (7%) probably has three causes:

the piecemeal extension of the NZGD49 control from the 1st order marks down to the 4th order marks;

disturbance of marks between the NZGD49 surveys and the NZGD2000 surveys; and

NZGD49 coordinates which are of lower accuracy than implied by their order.

3.2 *Relative weighting of control marks*

The initial intention when performing this analysis was to ensure a good fit at the NZGD49 1st order stations at the expense of the fit at lower order marks. The reason for doing so is that the NZGD49 coordinates of the 1st order marks are, by definition, perfect. This is modelled in the analysis by assigning much lower errors to the coordinate differences at the 1st order control points than at other points. By using a 20km grid and suitably low errors on the 1st order points it is possible to fit them to better than 1 centimetre.

However after some consideration this approach was not adopted. Instead equal weighting was given to all control points. The reasoning was that the transformation model was to be used to convert NZGD49 coordinates to NZGD2000 (or vice versa). Generally NZGD49 coordinates will be based upon the nearest available NZGD49 marks, rather than specifically upon 1st order marks. It follows that there is no special reason to favour the 1st order marks in determining the transformation model.

The transformation model is therefore defined in terms of the implementation of NZGD49, rather than in terms of the definition of NZGD49.

The difference between these two approaches is less than 5cm at 66% of control points, and less than 10cm at 90% of control points. The maximum difference is just over 40cm, which is quite significant.

3.3 *Rigidity of the transformation model*

The "rigidity" of the transformation model is defined by the weighting of the distortion constraints relative to the weighting of the control points. The approach used was to allow the control points to define the model as far as possible and downweight the distortion constraints so that they only really applied to extrapolating into areas where there was insufficient control.

The numerical relationship between weighting of the constraints and weighting of observations at the control points is somewhat arbitrary, since they are expressed in terms of very different quantities. The approach used to find the ideal relative weighting was to try a range of relative weights and assess them in terms of the residual the control points, and the shear distortion on the grid cells. The weights are as follows:

Distortion/Control weighting ratio	Residual (m)			Shear (ppm)		
	50%	95%	99%	50%	95%	99%
100	1.37	2.34	2.59	0.0	0.0	0.0
10	1.30	2.21	2.44	0.0	0.0	0.0
1	0.62	1.32	1.53	0.1	0.2	0.4
0.1	0.19	0.78	1.18	0.7	4.6	7.6
0.01*	0.05	0.21	0.36	3.0	17.1	29.0
0.001	0.03	0.14	0.30	5.4	34.1	55.8
0.0001	0.03	0.13	0.30	9.3	109.2	262.2
0.00001	0.03	0.13	0.30	21.0	494.0	2247.0

The weighting ratio used in the final analysis is 0.01. This was chosen because a lower ratio does not significantly improve the fit at the residuals, but does give a significantly greater distortion in the transformation model, and a lower ratio results in notably greater residuals.

After applying the selected weighting the distortion constraints still influenced the fit significantly in the region of the Southern Alps. This is probably due to inconsistency between the original NZGD1949 surveys on the east of the Alps, and the 1970's extension onto the West Coast on the west of the Alps. To account for this the distortion constraint was further downweighted for a number of cells spanning this region (as shown in Figure 3).

4 Calculation and characteristics of the final transformation model

The calculation method used above to generate the distortion model requires grid cells that are approximately square. To achieve this it was constructed by defining square cells in terms of the New Zealand Map Grid (NZMG) projection. However most of the software that uses distortion grids requires a model defined as a regular grid in terms of latitude and longitude coordinates.

The final NZMG distortion model was used to calculate the distortion at the nodes of a latitude and longitude grid using a 0.1° spacing in each direction. The latitude/longitude grid spacing is less than 10km, as compared to the 20km spacing on which the model was generated, so that it accurately reflects the calculated NZMG model.

This model is available on the LINZ web site at
http://www.linz.govt.nz/services/surveysystem/osgpublications/nzgd2000_gridtrans.html.

The model is supplied in the NTv2 grid format developed by Geomatics Canada and adopted by Australian Government Survey Agencies. The site also provides an online coordinate converter and Microsoft Windows based transformation software, which may be freely downloaded, for converting latitude and longitude coordinates between NZGD2000 and NZGD49.

The characteristics of the final model are summarised in the following diagrams.

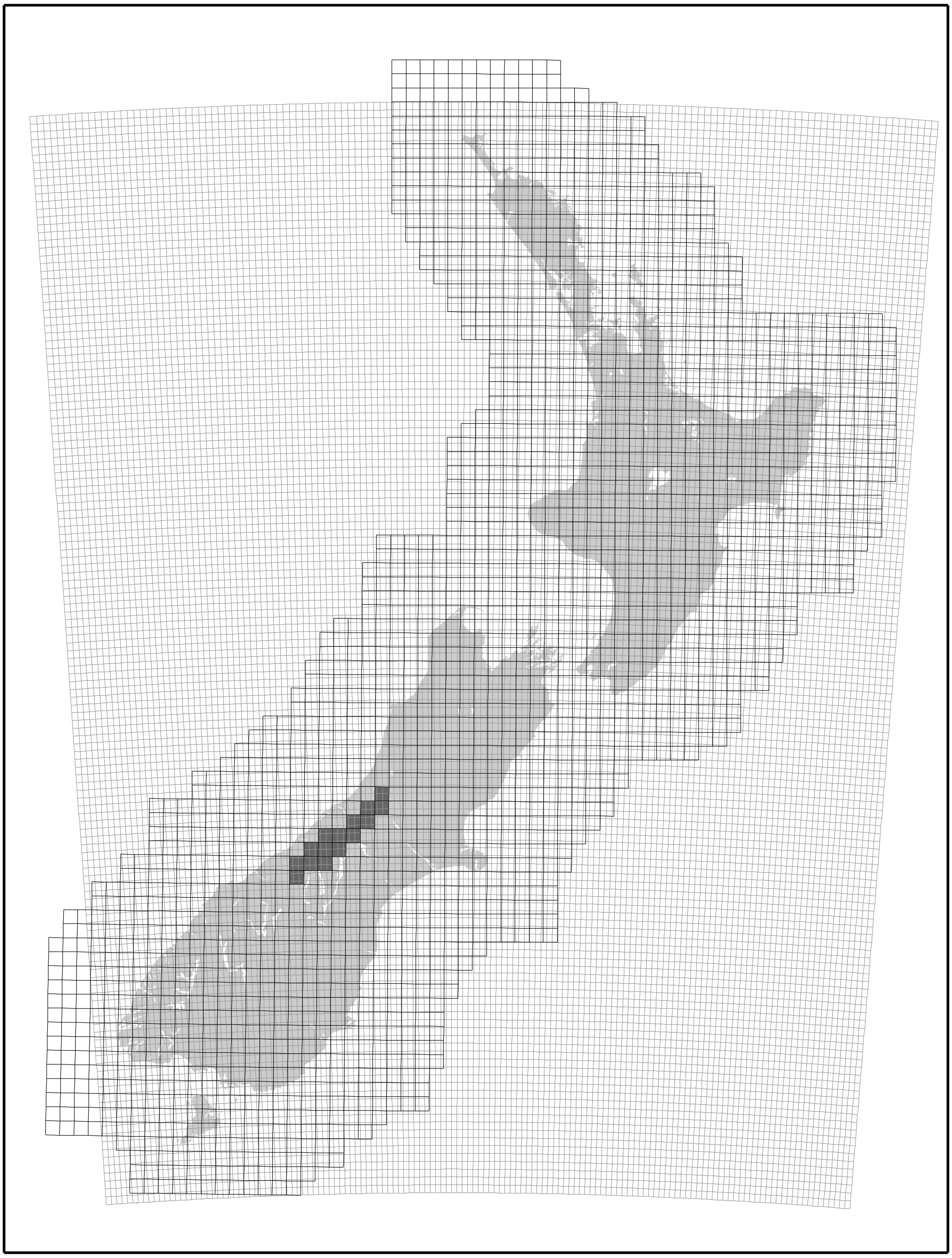


Figure 3: The course grid shown is the 20km grid of points on which the distortion model was calculated. The shaded cells are those on which the distortion constraint on the cells was downweighted to account for the difference between the original NZGD49 definition on the east of the South Island and the later extension on the West Coast. The distortion model was interpolated onto the finer grid for publication.

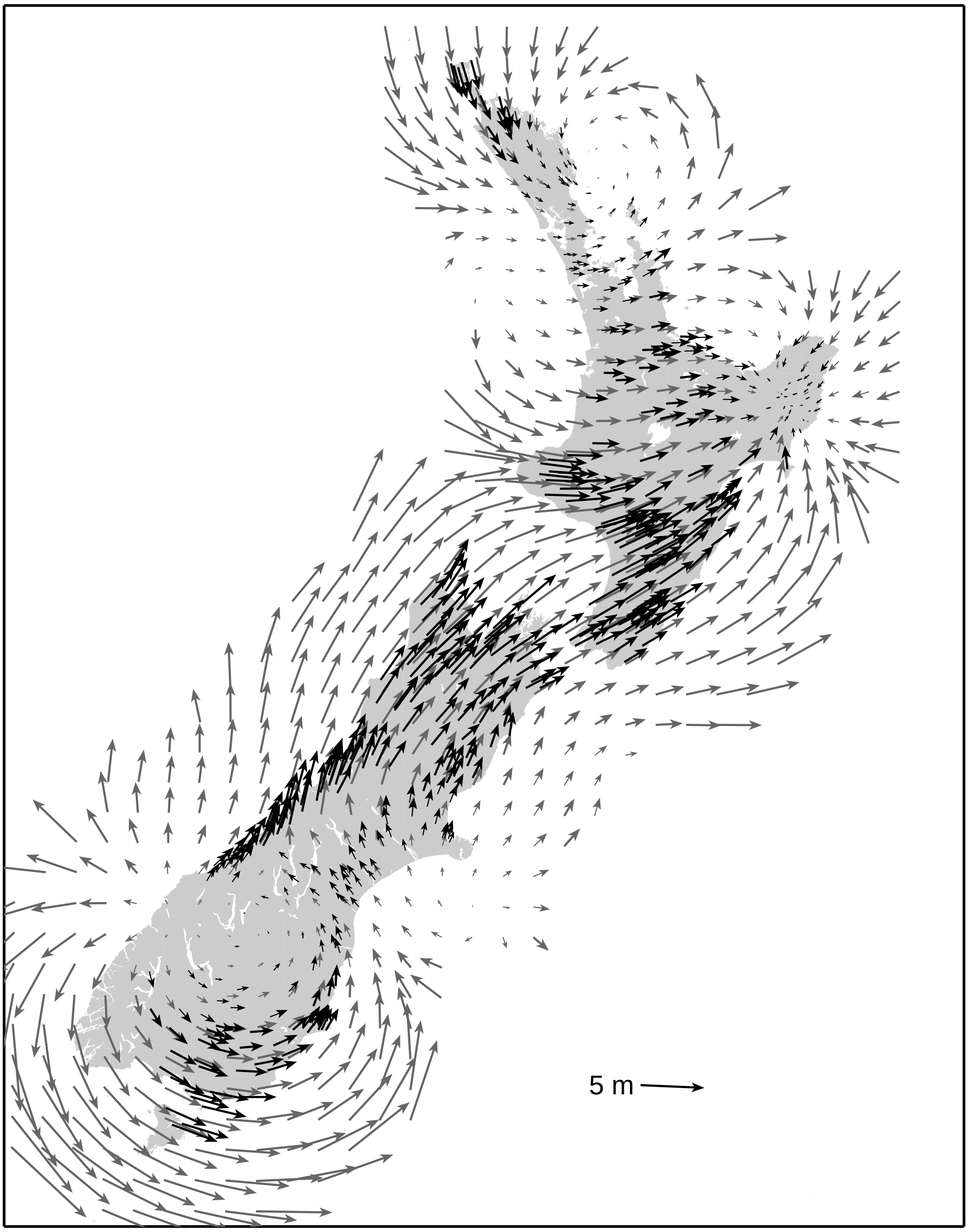


Figure 4: This shows the observed (black) and calculated (grey) differences between NZGD2000 and NZGD49. These are the residual differences after applying the 7 parameter Bursa-Wolf transformation described by Pearse and Crook [1997]. For the sake of clarity the calculated vectors are only shown at every second grid node, and the observed vectors are shown only for 1st and 2nd order NZGD49 marks.

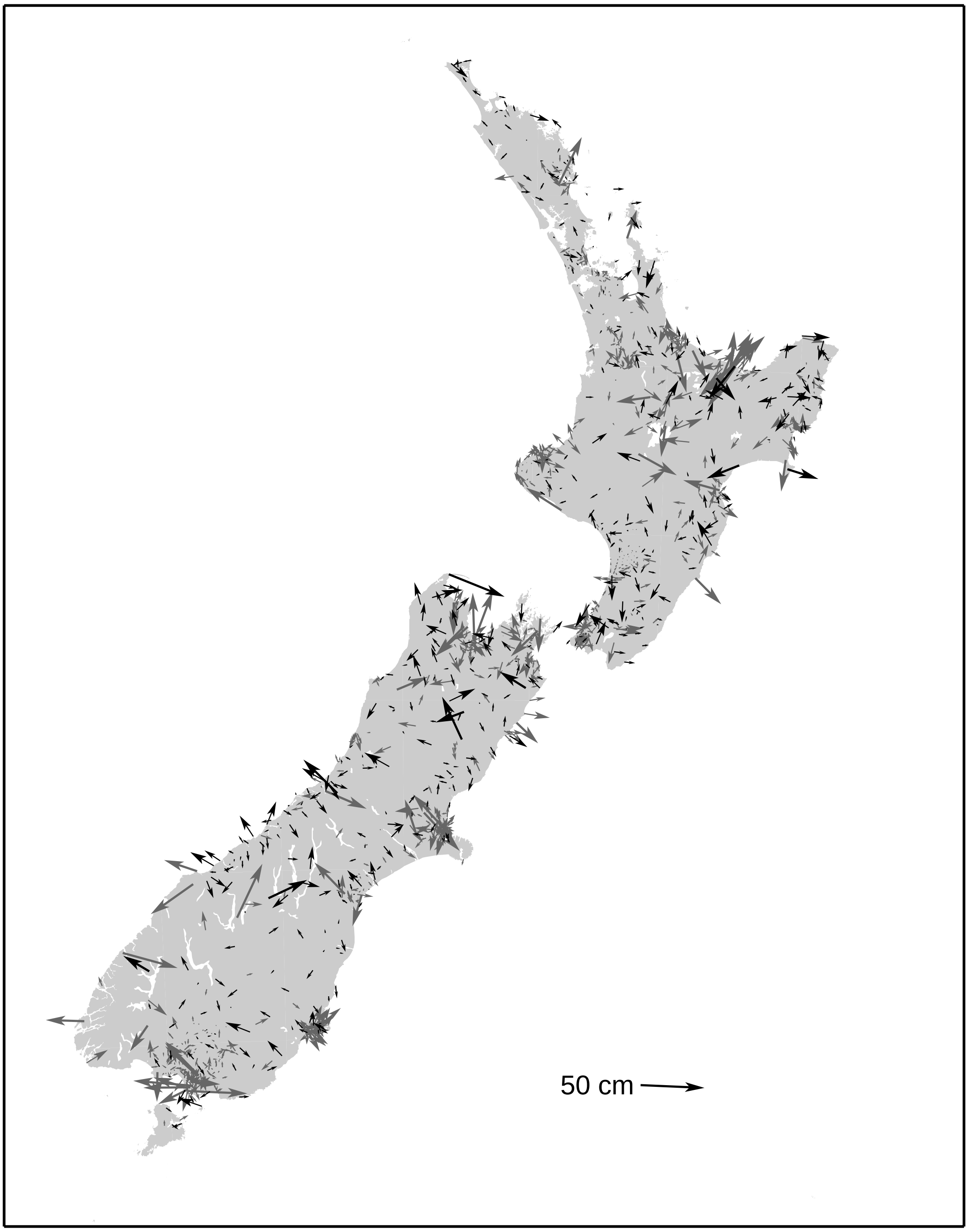


Figure 5: This shows the residual errors between the observed and calculated coordinate differences at the control points (that is, the coordinate difference after the distortion model has been applied). The NZGD49 1st and 2nd order marks are shown in black, and the 3rd and 4th order marks are shown in grey.

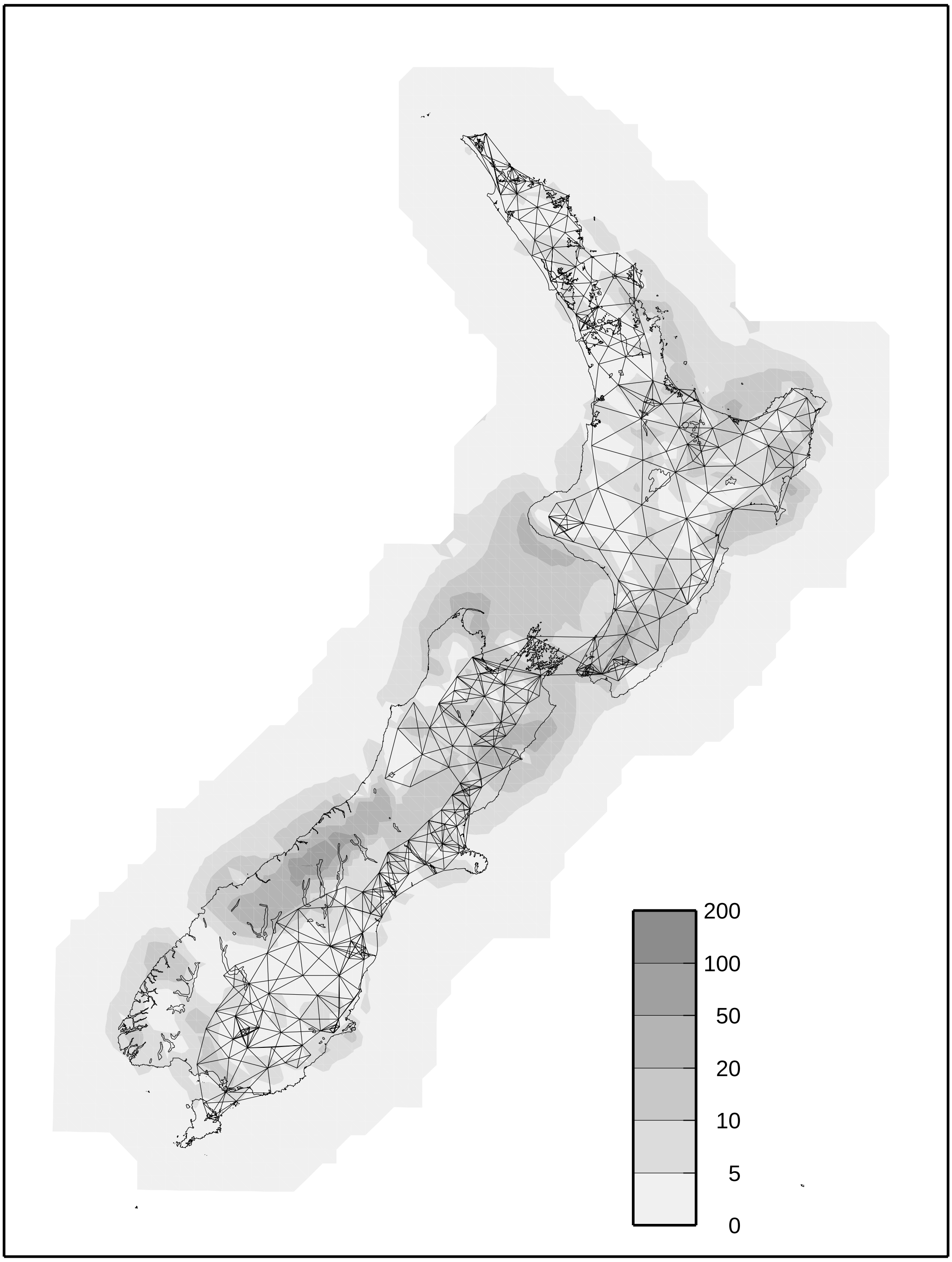


Figure 6: This shows the distortion as modelled in the NZGD49 datum. The distortion is expressed in terms of the maximum part per million shear component of "deformation" on the grid cells (see Appendix B). For reference the original NZGD49 triangulation network is also shown. This illustrates how the most significant distortion in the South Island reflects the deficiencies in the original definition of NZGD49.

5 Summary

A transformation model has been developed to allow transformations between latitude and longitude values in terms of New Zealand Geodetic Datum 1949 (NZGD49) and latitude and longitude in terms of New Zealand Geodetic Datum 2000 (NZGD2000).

The transformation model was calculated on a grid of points with a 20km spacing over the land area of New Zealand and has been interpolated onto a latitude/longitude grid with 0.1 degree spacing for publication. It accounts for the difference in the underlying datums of NZGD2000 and NZGD49 and the distortion in implementation of the NZGD49 datum.

The accuracy of the model is limited by the accuracy of the implementation of the NZGD49 datum. Based upon the control points used in this analysis the model appears to achieve an accuracy of about 0.2m at the 95% confidence level, and better than 0.4m at the 99% confidence level. In assessing these statistics it should be noted that 7% of the original control points have already been culled as having anomalous NZGD49 coordinates.

The initial estimate of the accuracy of NZGD49 coordinates from the breakdown of control into the 3rd and 4th order networks (Appendix A) predicted a mean circular error depending on the distance from the nearest 1st order control as 0.01 m + 0.003 m/km. If we take a typical distance from the nearest 1st order node as 15km, and convert this to a 95% confidence limit, this gives an expected error of about 0.14m. This estimate was derived from a single adjustment of a large amount of 3rd and 4th order data. In practice most of the NZGD49 coordinates will have been derived in a more piecemeal fashion and so will have larger errors. The 0.2m 95% confidence limit obtained from the analysis of the distortion model is therefore quite consistent with the expected errors in the NZGD49 datum.

The transformation model is as accurate a model as we can hope to achieve since it accounts for all but the expected random error in the NZGD49 datum - there is no more accurate definition of NZGD49 to transform from or to.

6 References

Lee, L.P. 1978: First-Order Geodetic Triangulation of New Zealand 1909-49 and 1973-74. *Department of Lands and Survey Technical Series Report No. 1*. Department of Lands and Survey, Wellington, New Zealand.

Pearse, M.B. and Crook C.N. 1997: Recommended transformation parameters from WGS84 to NZGD49. *Land Information New Zealand Geodetic System Technical Report 1997/11*. Land Information New Zealand, Wellington, New Zealand.

Pearse, M.B. 2000: Realisation of the New Zealand Geodetic Datum 2000. *Office of the Surveyor-General Technical Report No. 5*, Land Information New Zealand, Wellington, New Zealand.

Appendix A: Estimation of the expected accuracy of the NZGD49 datum

The distortion model attempts to predict the distortion in the NZGD49 datum based upon that observed at a number of control points. Its failure to do so will result at least partially from the inaccuracy with which NZGD49 coordinates are propagated through surveys which tie the 2nd, 3rd, and 4th order station to the 1st order stations which define the datum. The quality of this tie was investigated by analysing one example of a control network.

The example was chosen simply because the data was readily available in digital format. It is the job "JOB 1644 OTOKIA-TAIERI-SUTTON 2ND AND 3RD ORDER CONTROL". It represents a fairly typical set of observations for the breakdown of a network tying some 220 stations onto 1st order control. The observation set comprises mainly angles of various accuracies. Figure A1 below shows the layout of the network with the 1st order stations highlighted.

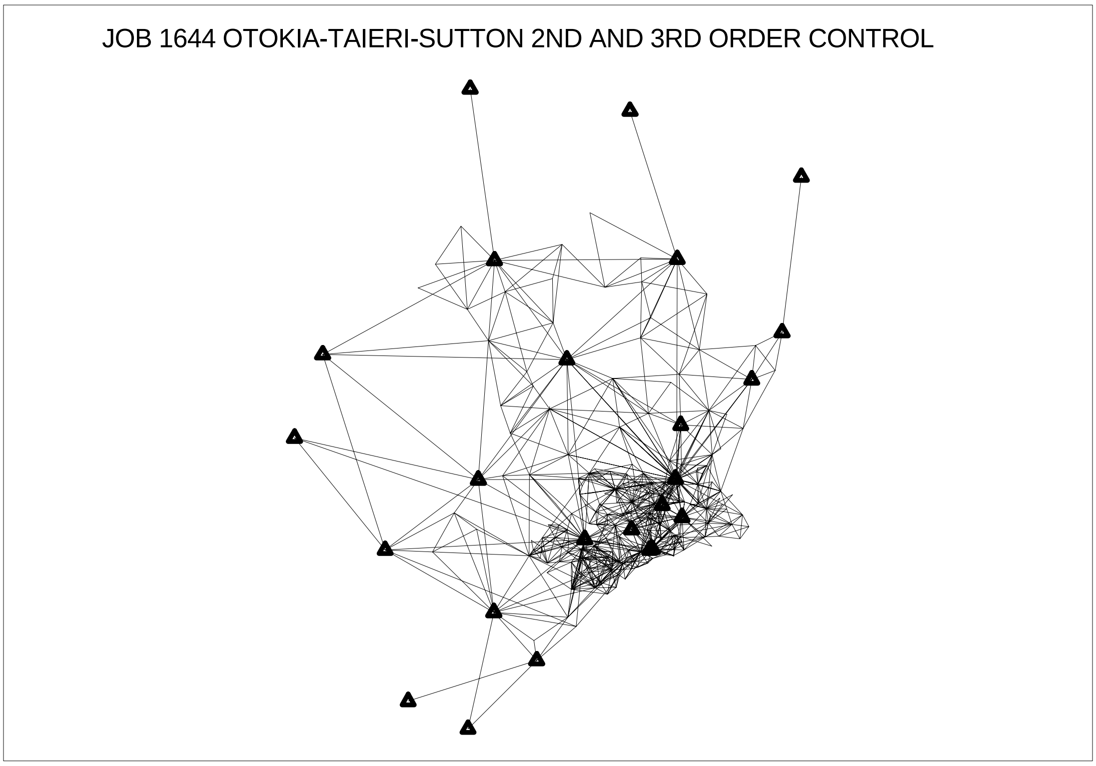


Figure A1: Showing the example network scheme

The network was adjusted to calculate *a priori* error ellipses for the stations. In the adjustment the 1st order stations were held fixed. Horizontal angle observations (or strictly directions) were assigned standard errors of 0.6″ for 1st order observations, 1.2″ for 2nd order observations, and 2″ for other observations. The data were not used to calculate coordinates or residuals - for this exercise only the configuration of the network and the observations is of interest.

The error ellipses calculated indicate how accurately the 2nd, 3rd, and 4th order points are positioned relative to the fixed 1st order points. This is a “random” error - it cannot be predicted by any observations based upon just the 1st order stations. It is the best that the distortion model could hope to achieve.

The error ellipses of the stations have been analysed in terms of their distance from the nearest 1st order station. The error ellipses are first converted to a circular expected error. Statistics were accumulated for points within 2.5km of a 1st order station, between 2.5km and 5km, and so on. Within each category the median circular expected error of the 2nd, 3rd, and 4th order stations has been calculated. The results are summarised in Figure A2

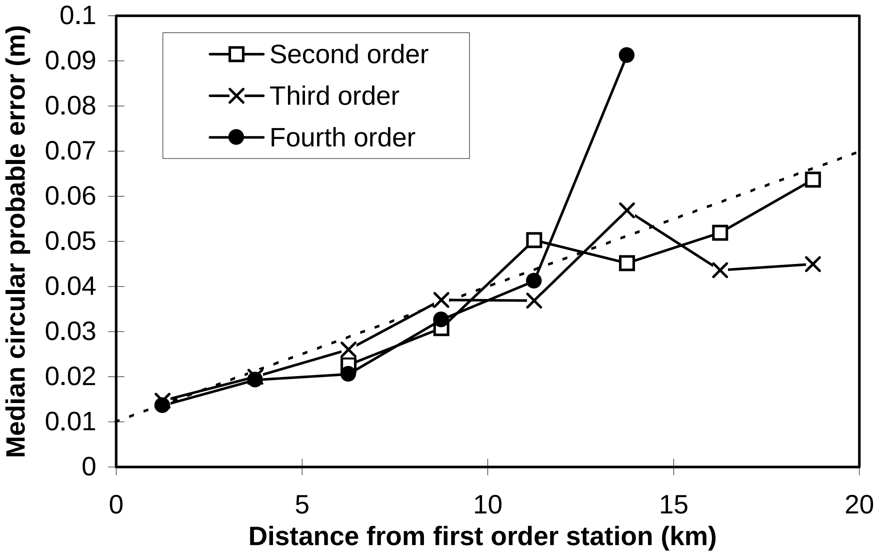


Figure A2: Median circular expected error as a function of distance from 1st order stations. The dashed line shows the calculated accuracy estimate of 0.01m + 0.003 m/km. Note that the last point of the fourth order data points which appears anomalous is based upon only four stations.

It is obvious from this plot that there is no significant difference in the accuracies with which different order marks are tied into the 1st order network. Although 3rd and 4th order marks are surveyed using less accurate observations this is compensated for by the higher density of the observations.

It is also clear that the relationship between the expected error and the distance to a control station is well approximated by a linear relationship (a straight line) as

$$E = 0.01 + 0.003 \times D$$

where E is the expected error in metres and D is the distance from the nearest control point in kilometres. This is equivalent to 0.01m + 0.003 m/km. The dashed line shows this formula on Figure A2.

This result is probably the best that can be expected, since it is derived from a comprehensive adjustment of all the available data. In those areas where control has developed and been adjusted on a more piecemeal basis the expected errors may be slightly larger.

The error of points relative to local 1st order stations will generally not be affected much by earth deformation, since there are only a few areas (in the vicinity of major active faults) where the effect of deformation is not smooth and predictable within a triangle of the 1st order network.

Appendix B: Transformation model calculation method

The transformation model is constructed by estimating the difference between the NZGD49 and NZGD2000 coordinates at points on a regular grid based upon the observed differences at an irregularly spaced set of control points.

The two main characteristics of the model developed are that it attempts to fit the observed deformation at the control points, and that it attempts to minimise the distortion in the network.

The mathematical model that has been used is based on a grid of points distributed across the country. The grid of points is defined in terms of the NZMG (New Zealand Map Grid) projection such that each grid cell is a square. In the following description we will define the grid as comprising square cells of size A metres in columns numbered 1 to N from east to west, and rows numbered 1 to M from south to north. The grid cell C_{ij} is defined as being in column i and row j. There are N+1 columns of nodes numbered from 0 to N, and having eastings x_0 to x_N, where $x_i = x_0 + i\times A$. Similarly there are M+1 rows of nodes numbered from 0 to M with northings y_0 to y_M, where $y_j = y_0 + j\times A$. The four nodes of grid cell C_{ij} are at (x_{i-1},y_{j-1}), (x_{i-1},y_j), (x_i,y_{j-1}), and (x_i,y_j).

The model defines the easting and northing correction at each node of the grid. The correction at node (x_i,y_j) is defined to be (u_{ij},v_{ij}). Corrections at other points are determined by bilinear interpolation of the easting and northing values. That is, if the coordinate of a point in cell Cij is (x,y), then the correction at that point $U_{ij}(x,y)$, $V_{ij}(x,y)$ is given by the formula

$$U_{ij}(x,y) = \{u_{i-1,j-1}(x_i-x)(y_j-y) + u_{i-1,j}(x_i-x)(y-y_{j-1}) + u_{i,j-1}(x-x_{i-1})(y_j-y)+u_{ij}(x-x_{i-1})(y-y_{j-1})\}/A^2$$
$$V_{ij}(x,y) = \{v_{i-1,j-1}(x_i-x)(y_j-y) + v_{i-1,j}(x_i-x)(y-y_{j-1}) + v_{i,j-1}(x-x_{i-1})(y_j-y)+v_{ij}(x-x_{i-1})(y-y_{j-1})\}/A^2$$

The corrections at the nodes are calculated in a large least squares adjustment. This uses two types of observations as input. Firstly, the correction calculated at the control points should be close to the measured value, and secondly the distortion of grid cells should be as small as possible. These two observations are described in more detail below.

The first type of observation comes directly from the formulae above, where U and V represent the measured corrections. The observation for control point P at (x_P,y_P) in cell C_{ij} and with measured correction (U_P,V_P) is

$$U_P = \{u_{i-1,j-1}(x_i-x_P)(y_j-y_P) + u_{i-1,j}(x_i-x_P)(y_P-y_{j-1}) + u_{i,j-1}(x_P-x_{i-1})(y_j-y_P)+u_{ij}(x_P-x_{i-1})(y_P-y_{j-1})\}/A^2 + \varepsilon_{P1}$$
$$V_P = \{v_{i-1,j-1}(x_i-x_P)(y_j-y_P) + v_{i-1,j}(x_i-x_P)(y_P-y_{j-1}) + v_{i,j-1}(x_P-x_{i-1})(y_j-y_P)+v_{ij}(x_P-x_{i-1})(y_P-y_{j-1})\}/A^2 + \varepsilon_{P2}$$

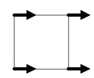
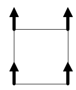
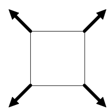
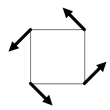
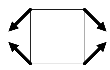
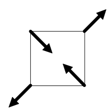
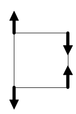
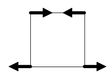
In these equations the term ε_P represents the error of the measured correction.

These observations relate the measured correction at the control point to the model corrections at the four nodes of the cell within which it lies. On their own they are not enough to calculate the model corrections, since they provide no information about nodes which are not adjacent to measured corrections.

The second type of observation specifies that there should be no distortion of each cell. The distortion of cell Cij is expressed in terms of the vector of 8 components (2 corrections at each of four nodes) as:

$$(u_{i-1,j-1}, v_{i-1,j-1}, u_{i-1,j}, v_{i-1,j}, u_{i,j-1}, v_{i,j-1}, u_{i,j}, v_{i,j})$$

There are 8 types of deformation that can be expressed in terms of these components. These are shown in the following table.

Description	Vector	Diagram
Translation east	$0.5 \times (1,0,1,0,1,0,1,0)$	
Translation north	$0.5 \times (0,1,0,1,0,1,0,1)$	
Scale	$1/\sqrt{8} \times (-1,-1,-1,1,1,-1,1,1)$	
Rotation	$1/\sqrt{8} \times (1,-1,-1,-1,1,1,-1,1)$	
Shear	$1/\sqrt{8} \times (-1,1,-1,-1,1,1,1,-1)$	
Shear	$1/\sqrt{8} \times (-1,-1,1,-1,-1,1,1,1)$	
2nd order strain	$0.5 \times (0,-1,0,1,0,1,0,-1)$	
2nd order strain	$0.5 \times (-1,0,1,0,1,0,-1,0)$	

Note that in this table each vector has an associated scale factor to convert it to a unit length vector. These vectors form an orthonormal set - which means that we can express the deformation of the cell in terms of these eight components instead of the corrections at the individual nodes.

The survey network has been largely developed by horizontal angle measurements. Any of the first four sorts of deformation (east and north translation, rotation, and scale) can be present in a survey network without affecting angles. These types of deformation are therefore considered "acceptable" - we expect them to some degree. However the last four deformations (shear and 2nd order strains) are incompatible with angle measurements - if these are present then angles are changed. These four deformations represent the distortion of the network.

In order to minimise distortion in the model four "pseudo-observations" are added for each cell which specify that the shear and 2nd order strains are equal to zero. That is

$$0 = -\alpha u_{i-1,j-1} + \alpha v_{i-1,j-1} - \alpha u_{i-1,j} - \alpha v_{i-1,j} + \alpha u_{i,j-1} + \alpha v_{i,j-1} + \alpha u_{i,j} - \alpha v_{i,j} + \varepsilon_{d1}$$
$$0 = -\alpha u_{i-1,j-1} - \alpha v_{i-1,j-1} + \alpha u_{i-1,j} - \alpha v_{i-1,j} - \alpha u_{i,j-1} + \alpha v_{i,j-1} + \alpha u_{i,j} + \alpha v_{i,j} + \varepsilon_{d2}$$
$$0 = -\beta v_{i-1,j-1} + \beta v_{i-1,j} + \beta v_{i,j-1} - \beta v_{i,j} + \varepsilon_{d3}$$
$$0 = -\beta u_{i-1,j-1} + \beta u_{i-1,j} + \beta u_{i,j-1} + \beta u_{i,j} + \varepsilon_{d4}$$

The terms ε_d represent the error in these "pseudo-observations" and are a measure of the distortion in the model. What follows this term is expressed as parts per million of the grid size. That is, if the grid size is 10km, then a distortion error of 10ppm implies the value of ε_d is 0.1m. The terms α and β are the constants scaling the deformation vectors to unit length, that is $\alpha = 1/\sqrt{8}$ and $\beta = 0.5$.

These two types of observations are compiled into a least squares adjustment that calculates the corrections at each node that minimise the weighted sum of squared residual errors. The residual errors for control point observations are weighted according to the NZGD49 order of the control point. A single weighting is used for all the distortion errors. That is, the network is treated as if it were geographically homogeneous.

One difficulty with this method is that it does result in a large set of equations to be solved - the matrix to be solved is of size equal to twice the total number of nodes in the grid. For example if the grid consists of 50 rows and columns then there are 5202 corrections (2 for each node of the grid), so that a 5202×5202 matrix must be solved to calculate them. If the grid size is halved then the matrix is quadrupled in size, and the time taken to solve it may increase by a factor of around 16. However the matrix is well suited to numerical optimisation (by minimising its bandwidth). Also the problem can be solved by a sequential approach if it becomes too big.

In practice there has been no difficulty solving these equations for a 20km grid over New Zealand using a PC, and finer grids could be solved given sufficient time or using a more sophisticated numerical method.

One refinement of the method is that the grid is limited to cells within a certain distance of control points - anything further away is simply removed from the problem.

A variation to this method that has been considered, but not investigated, is to use a triangular mesh rather than to a grid. The mesh could be based upon the survey network, and so could better reflect the density of marks and provide a more detailed model of distortion where there is information to support it. Each triangle has 6 corrections (two at each of three nodes), that can be represented by translations, rotation, scale change, and two shears. In each triangle there are therefore two "pseudo-observations" of distortion relating to the two shears. One issue with applying these observations is that they must be weighted individually to reflect the size and shape of the triangle. This would require careful consideration. Another difficulty with using a triangular mesh is that its continuation beyond the survey network (to the coast, for example) is very arbitrary. A rectangular grid of corrections is also more convenient to use - a lot of software that is available can use gridded information.

One of the main drawbacks of the gridded method is that it is a "global" method. In principle an observation at North Cape will have an effect on the model correction at Bluff. The effect will be minuscule. However it does mean that if new data are obtained the entire model must be recalculated, not just a region in the vicinity of the new data.

It would be of value to explore and compare other techniques for fitting. In particular least squares collocation offers promise, as it can include both a model (translation, rotation, and scale) and a random deviation from that model.