Data-Mining in NSW: Working Towards a New and Improved Australian Datum

Joel Haasdyk
Survey Infrastructure and Geodesy, Land and Property Information
NSW Department of Finance & Services
Joel.Haasdyk@lpi.nsw.gov.au

Tony Watson
Survey Infrastructure and Geodesy, Land and Property Information
NSW Department of Finance & Services
Tony.Watson@lpi.nsw.gov.au

ABSTRACT

Advances in positioning technologies in the last decades have given a large range of user-groups access to positioning at ever-increasing accuracies and improved spatial and temporal resolution. In this context, the limitations of our current Australian datum(s) are becoming apparent. For example, while the Geocentric Datum of Australia 1994 (GDA94) was cutting-edge when it was adopted 13 years ago, it has since been shown to have limited internal accuracy and indirect linkages to the current global reference frame, the International Terrestrial Reference Frame 2008 (ITRF2008). Since GDA94 was 'frozen' at the epoch of 1994.0, Australia has moved (and rotated) by up to 1.3 metres. The spatial community is currently discussing the best way forward to a next-generation datum, here hypothetically termed 'GDA201x'. In preparation for this new datum, each state and territory is harvesting all available observations from their archives. New computing technologies mean that state-wide and even nationwide adjustments are now routinely possible with a theoretically unlimited number of observations. In the context of New South Wales (NSW), this paper discusses the process of 'dusting-off' the observations originally used to compute GDA94, and the 'data-mining' of our archives for all new GNSS observations which have contributed to the densification of the datum since its adoption. These observations are combined, cleaned and adjusted using the phased-adjustment software 'DynaNet' and other in-house software, and constrained by the CORSnet-NSW network and several hundred AUSPOS results. It is demonstrated that by increasing the number of fiducial stations in NSW (from 13 Australian National Network stations to 113 CORS), and by increasing the number and quality of observations across NSW (from 4,000 GNSS observation to more than 62,000), we can significantly improve the accuracy of the datum across and between the states. Rigorous Positional Uncertainty (PU) can be computed from this first ever state-wide simultaneous adjustment, and with the current preliminary constraints the majority of stations have a horizontal PU of better than 20 mm, up to an order of magnitude better than in GDA94.

KEYWORDS: *GDA94, GDA201x, CORSnet-NSW, datum, data mining, phased adjustment.*

1 INTRODUCTION

Imagine that a surveyor decided to have another look inside all of those yellow project folders that are squirreled away in their office, in order to make one big list of all the observations that they had ever taken. This includes the current folders that are hiding under the phone and coffee cups, the folders from recent years filed in the compactors, and even the historical folders in the garage/offsite archive that were inherited when the practice was purchased. If they have managed to 'go digital' in the last two decades, all the better, but what about those aging CD-ROMs (or floppy discs?) that were eagerly burned years ago with that brand-new '386' computer or AppleIIc. Now consider the different file formats and programs that have probably been used over the years, and the need to find (or write) converters to translate each observation into the same format.

The task described above seems unthinkable and almost insurmountable; why would anyone do such a thing? Well, if the surveyor's new job is to provide control station coordinates over their whole area of operations, and results are expected to join seamlessly with neighbouring surveyors, then the benefits of consolidating these observations become more apparent. In a perfect world, all of this data would end up in a searchable database and be visualised in a GIS. This would serve as a good record of where observations are plentiful, and more importantly, highlight gaps in the coverage and areas in need of new control. Since the most expensive part of any survey is putting boots on the ground, existing data could be reused, where appropriate, to minimise the time spent gathering new observations (although making appropriate checks of previous work remains best practice).

Individual jobs could be amalgamated to take advantage of new computing power, which can solve for the coordinates of *all* stations in a single least squares adjustment, instead of job by job. The beauty of this approach is that existing data would be used to highlight any errors in new observations, and vice-versa. As measurement technologies improve, any new higher-precision observations would serve to make the resulting control network stronger and more accurate. Consider the pros and cons in terms of precision, distance and speed of observations when comparing the steel band vs. Electronic Distance Measurement (EDM) vs. Global Navigation Satellite Systems (GNSS). In addition, combining jobs observed at different times will give a good idea of where the land (or the monuments) might be moving.

NSW Land and Property Information (LPI) is undertaking just such a task. In this paper, it is argued that the coordinates of our control stations (the realisation of the datum) actually need to be periodically re-examined and refreshed in order to provide a backbone of suitable accuracy against which all new technologies, observations and spatial data are applied. In order to facilitate investigations into a possible datum update, the Intergovernmental Committee on Surveying and Mapping (ICSM) has mandated that each state and territory prepare a primary network of geodetic data for testing a nationwide adjustment in 2013. To this end, LPI is undertaking a project to collate and clean all available GNSS vectors in its archives and to test the first ever state-wide simultaneous adjustment of these observations.

This paper briefly outlines some of the salient features of some of the previous datums in Australia, reviews some of the drivers of datum change and introduces the use of DynaNet for large-scale simultaneous adjustments. It then describes the lessons learned, and some of the tools created during the harvesting, cleaning and adjustment of nearly a quarter of a million GNSS vectors, collected over the last 20 years, from the digital archives at LPI.

1.1 Recent History of Australian Datums

Featherstone (2013) provides a good history of recent geodetic datums in Australia. Prior to 1966, a variety of datums and ellipsoids were used for mapping and charting in Australia, leading inevitably to coordinate discrepancies and confusion. In the1960s, the Australian Geodetic Datum (AGD66), and its UTM projection – the Australian Map Grid 1966 (AMG66) – were introduced Australia-wide to remove the heterogeneity and make use of all available geodetic observations (Bomford, 1967). AGD66 was realised with the Australian National Spheroid (ANS) and coordinates adopted for the Johnston origin based on 275 astro-geodetic stations distributed over most of Australia (ICSM, 2006).

In the 1980s, new technology (notably satellite-derived geodetic data) soon highlighted deficiencies in the AGD66 coordinates (Lambert, 1981). A readjustment produced the Geodetic Model of Australia (GMA) 1982, which was subsequently adopted as the Australian Geodetic Datum 1984 (AGD84) along with its UTM projection (AMG84) again based on the ANS and Johnston origin. Differences between AGD66 and AGD84 of up to 6 metres (up to 1 metre in NSW) were noted (Allman and Veenstra, 1984). Even as AGD84 was adopted, the National Mapping Council (NMC, 1986) "*recognised the need for Australia to eventually adopt a geocentric datum*" in order to be directly compatible with emerging and global satellite technologies. AGD84 was only the first of two steps towards datum modernisation. With this in mind, many states and territories decided to hold off until the change to a geocentric datum. Only Western Australia, South Australia and Queensland adopted AGD84, resulting once again in heterogeneous coordinate datums across Australia.

In the 1990s, the ICSM resolved to adopt the Geocentric Datum of Australia (GDA94), based on a new ellipsoid (GRS80), and recommended that the implementation of this datum be completed Australia-wide by 1 January 2000 (GA, 2012a; ICSM, 2013a). GDA94 was realised in 1995 by the 3-dimensional coordinates of 8 Australian Fiducial Network (AFN) stations determined simultaneously, using GPS in the most rigorous datum available, the International Terrestrial Reference Frame (ITRF) 1992, at epoch 1994.0 (Manning and Harvey, 1992; ICSM, 2006). Through the ITRF, and for all practical purposes, GDA94 was fully compatible with GPS in terms of spheroid and datum *at the time of implementation of GDA94* (Malays and Slater, 1994; ICSM, 2006). By 2012, the GDA94 coordinates of the AFN could no longer be realised to better than 30 mm (horizontal) and 50 mm (vertical) (95% confidence – Dawson and Woods, 2010). Since 1995, these AFN stations have been re-gazetted twice, with more stations (now 21 AFN stations) and improved coordinate precision (now 7 mm horizontal, 15 mm vertical), to allow an improved realisation of the datum (Australian Government, 1998, 2012; GA, pers. comm.).

In addition, 78 national stations (13 in NSW), known as the Australian National Network (ANN), were observed with GPS and adjusted together with the AFN in a single solution to create a framework for GDA94 at 500 km intervals, with an estimated accuracy of 50 mm (95% confidence – ICSM, 2006) or 0.1 ppm (ICSM, 2006; Watson, 2006). Further adjustments were computed in a hierarchy where coordinates of higher order control were held fixed and assumed to be perfectly known in each subsequent adjustment. For example, a densification of horizontal geodetic control using approximately 71,000 terrestrial and GPS observations (Table 1) was undertaken in 12 sections of less than 1,000 stations each (computing power limiting the size of each section) to propagate the datum (ICSM, 2013b; Figure 1). This last adjustment, containing the 3,000 NSW stations commonly referred to as the 'spine' network was officially accepted by ICSM in June 1997. The majority of horizontal coordinates in the NSW spine network were estimated to be better than 250 mm (95% confidence – Watson, 2006) and were given Class '2A' and Order '0'.

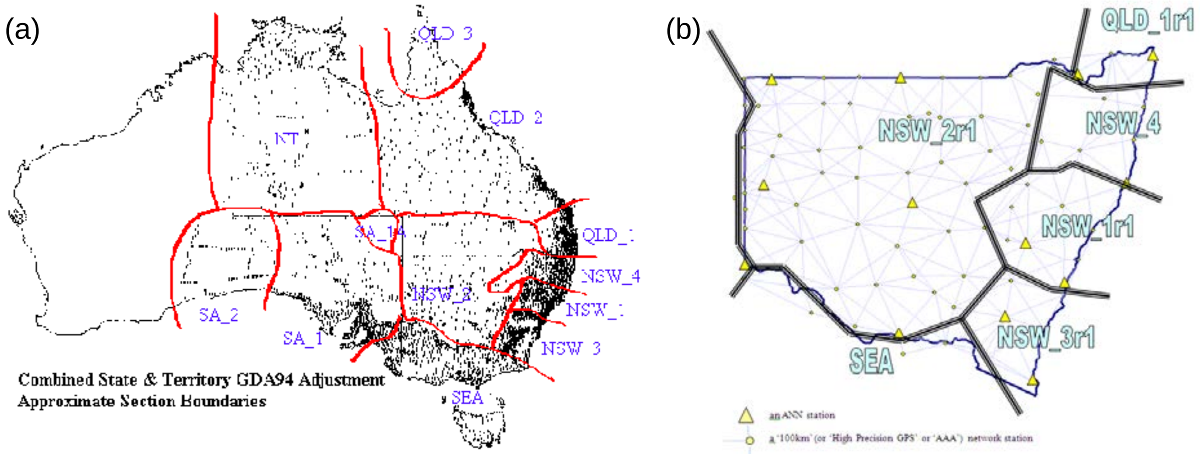

Figure 1: Datum adjustment sections for GDA94 – (a) Australia (ICSM, 2013b), and (b) NSW (Watson, 2006).

GDA94 was adopted by all states and territories, with only NSW and South Australia passing related legislation (e.g. LPI, 2013a). Any remaining stations would be transformed from AGD66 to GDA94 and then re-adjusted as more observations become available, a task which is still ongoing. Distortions across state borders were too large to allow a single national transformation model to be developed, so instead a national NTv2 grid was adopted (ICSM, 2006; LPI, 2013b).

This brief history demonstrates four main drivers of datum improvement: (1) technology changes such as terrestrial vs. satellite surveying techniques, (2) the desire for a single standard (homogenous) nationwide coordinate datum, (3) the inclusion of up-to-date geodetic observations and increased precision to smooth out distortions, and (4) the need to provide a denser realisation of the datum. The approximate numbers of stations and observations in recent Australian datums are shown in Table 1. Note that NSW contributed almost half of the stations and more than half of the observations included in the national GDA94 adjustment.

Table 1: Numbers of stations and observations in recent Australian datum definitions (Watson, 2006).

(values rounded to nearest 100)	AGD84 (NSW)	AGD84 (National)	GDA94 (NSW)	GDA94 (National)
Stations	~1,100	1,700	3,000	7,700
Directions		14,300	32,200	46,400
Distances		3,900	6,100	13,700
Azimuths		~200	~200	1,200
GPS Baselines			2,900	4,000
GPS Clusters				6,100
All Observations		18,400	41,400	71,400

1.2 Current Drivers for Datum Update

Since the adoption of GDA94 more than 13 years ago, the four drivers of datum change described above have still been in operation. Over time, and for more and more users, our world-class datum is beginning to shows its age. The motivation for datum change has once again become significant enough to cause us to carefully consider the best way forward. At the same time, it is necessary to carefully review the impacts and minimise any negative consequences associated with datum change. It is not in the scope of this paper to suggest the best format for any new datum, but it is instructive to consider some of the major issues and new drivers of datum change.

Since the last official datum update, GDA94, a significant amount of additional data has been gathered. Figure 2a, for example, shows a large number of new GNSS baselines (approximately 60,000, in blue) overlying the network of GPS observations that were used in the previous GDA94 adjustment in NSW (green). Currently in NSW, these new observations are used to provide coordinates for additional local ground control, but are not allowed to modify the coordinates of the higher order 'spine' stations. New, higher precision observations are often forced to fit the distortions in the original GDA94 adjustment.

Systematic distortions of up to 200 mm (horizontally) and 300 mm (vertically) have been demonstrated in the current datum in NSW (Figure 2b; Haasdyk et al., 2010). Positioning services such as CORSnet-NSW (Janssen et al., 2011; LPI, 2013c) and AUSPOS (GA, 2012b) provide precise coordinates in GDA94, without any reference to local ground control. As a result, an ad hoc, more homogenous realisation of the datum, dubbed GDA94(2010), has been created for use in NSW based on the direct connections of the CORS network to the AFN stations (Janssen and McElroy, 2010). CORSnet-NSW provides a backbone for positioning in NSW, of much higher density and precision than the original ANN. Removing these distortions between local control in the original GDA94(1997) and coordinates in GDA94(2010) would require a re-adjustment of the entire NSW network, without a hierarchy of fixed control, which is the subject of this paper. In the interim, a site transformation (Haasdyk and Janssen, 2012) is required to agree with the local (and legally accepted) ground control available in NSW via SCIMS (LPI, 2013d).

Soon, any internet-capable device will be able to determine an accurate position in the latest ITRF, with or without direct reference to GDA94, by directly accessing International GNSS Service (IGS) products in real time (Caissy et al., 2012). However, while GDA94 coordinates remain unchanged over time, the ITRF recognises that the Australian tectonic plate is moving at up to 7 cm/yr (~6 cm/yr in NSW) (Figure 2d; Dawson and Woods, 2010) and therefore GDA94 and ITRF diverge over time. This motion, equivalent to 1.3 metres since 1994.0 must be taken into account in order to transform these results into GDA94.

The Australian plate is not simply moving, but also rotating slightly, and deforming in places (Dawson and Woods, 2010). Significant errors can be introduced between observations taken at different epochs unless this rotation and deformation is accounted for. Stanaway et al. (2012) show that errors of up to 7 mm are introduced for baselines of only 30 km (Figure 2e) over the life of GDA94. These errors might not impact cadastral surveys directly, but engineering surveys and the provision of accurate control coordinates are definitely affected at these scales. Deformation, e.g. due to earthquakes, underground mining or water extraction, is less predictable, less systematic and occurs over a variety of time and distance scales.

Finally, creating a truly 3-dimensional datum which is coordinated across Australia is important for delivering the "*optimum benefits [of precise positioning] to the nation as soon as possible*" (ANZLIC, 2010). Figure 2c demonstrates jurisdictional boundaries where coordinate jumps are often introduced by the running of separate adjustments using different methods. In contrast, a nationwide simultaneous adjustment of all geodetic observations would result in a truly national and homogenous datum. Such an adjustment, without a fixed hierarchy of control, would allow the long-awaited introduction of Positional Uncertainty (PU – ICSM, 2007) at all stations and yield a dense network of ellipsoidal heights.

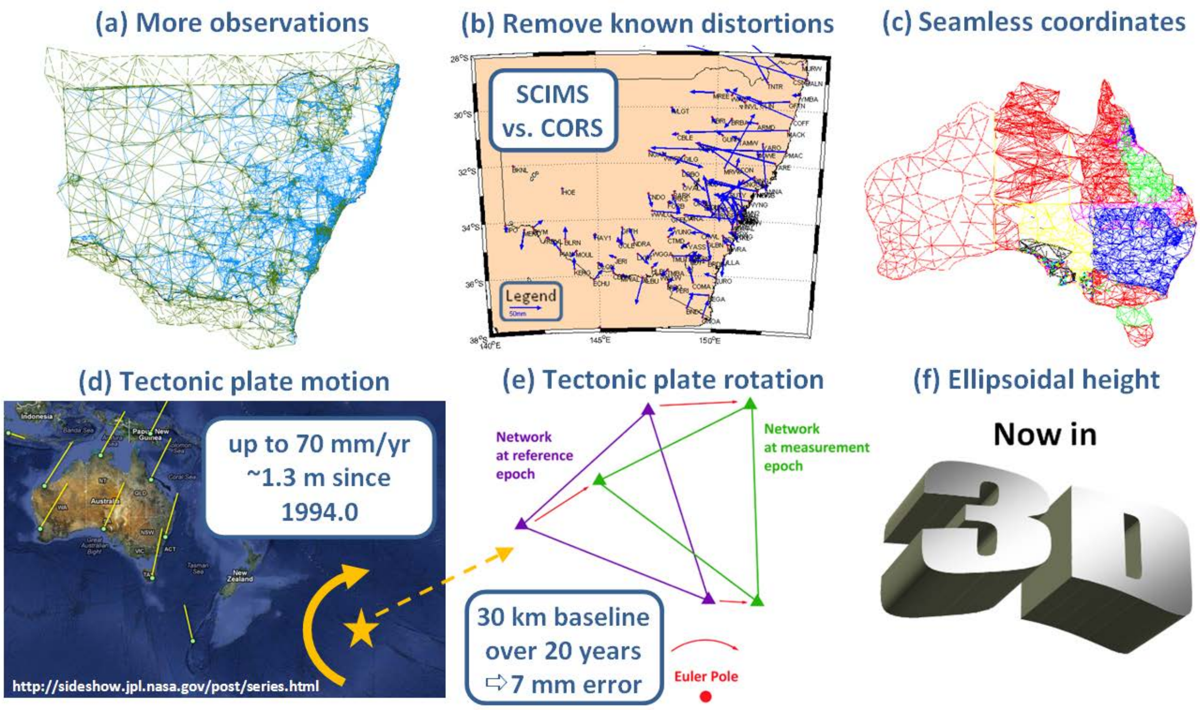

Figure 2: Some current drivers of datum change.

1.3 Dynamic Phased-Adjustment Least Squares Methodology and DynaNet

In the past, computing power has been a limiting factor in the simultaneous adjustment of large survey networks. Instead, large networks such as the most recent Australian national GDA94 adjustment (ICSM, 2013b, Figure 1) had to be segmented into ‘sections’ of ‘adjustable size’, which were treated individually and later recombined. This can lead to discrepancies across section boundaries, and also results in the loss of any relationships between stations in different sections.

However, least squares methodologies that are *not* limited by the number of stations or observations are possible, and have been recently re-examined. Leahy and Collier (1998) describe a ‘dynamic phased-adjustment’ that can perform a rigorous adjustment on a network of any size, and return a rigorous assessment of the quality of the adjusted station coordinates and observations. This allows the computation of absolute Positional Uncertainty (PU) as well as relative uncertainties across the entire network (ICSM, 2007). DyanNet, a program that can perform this dynamic network adjustment, is currently under development by the ICSM, building on the original work of the Department of Geomatics at the University of Melbourne. DynaNet can be implemented on a desktop computer, with adjustments of similar size to the GDA94 adjustment running in a matter of hours.

1.4 Challenging Paradigms

There is no doubt that the suggested adjustment ‘en-masse’ of all available observations challenges many paradigms. Survey and geodetic control networks are generally regarded as being static to provide control for lower order networks, but in a simultaneous adjustment, all stations and observations stand on their own merits (according to the weighting supplied), and any new observation should improve the adjustment accordingly. In this way, there is a 2-way interaction between all layers of control, in which ‘higher order’ control can influence ‘lower order’ control *and vice versa*.

Another implication of a simultaneous adjustment of all data is that new observations can be incorporated into the adjustment immediately after they are available and validated. A 'dynamic adjustment' allows for the regular or even continual upgrading of coordinate values. While modifying control values with any frequency might seem unpalatable at first, artificially constraining network stations and forcing new observations to fit the existing control will introduce distortions and tensions in the network as discussed above.

The result of increasing the frequency of adjustment would be increased homogeneity and reliability of all survey control coordinates. To date, the coordinates of the NSW spine network '2A0' control as determined in 1997 have been regarded as untouchable. While new observations and adjustments are allowed to affect the coordinates of subsidiary control as required, only exceptional circumstances will result in the change of spine coordinates. This is the reason why any distortions present during the initial densification of the datum in 1997 are still present today and being highlighted by new observations of better precision.

Other surveying maxims that are being challenged are the dismissal of the 'purpose of survey' (all observations are included in this adjustment out of context) and the neglect of network design (in terms of geography and repetition of observations). The implication of these paradigm changes should certainly receive proper attention and discussion in other forums.

1.5 A Comparison of Existing and Proposed Datum Methodologies

In light of the above discussions, Table 2 summarises the differences described between the existing and proposed datum definitions.

Table 2: A comparison between existing and proposed datum definitions.

	GDA94 as defined in 1997 ***(as applied in NSW)***	**GDA201x** ***(as applied in NSW)***
Primary Control	8 AFN *(1 in ACT/NSW)* 78 ANN *(13 in ACT/NSW)*	150+ CORS at final CORSnet-NSW network
Dimensions	3D with Ellipsoidal Height *(but 3D to ANN only within NSW)* AHD71 separate	3D with Ellipsoidal Height AHD71 remains separate
Observation Type (and Number)	GPS *(~ 3,000)* Terrestrial *(~38,000)*	GNSS *(60,000+)* No terrestrial at this time
Density	1st order spine control *(~3,000 stations)*	All control *(~20,000 stations and counting)*
Sectioning	YES: *(6 sections used to adjust NSW control, see Figure 1b)*	NO: Simultaneous adjustment of all control
Positional Uncertainty (95%)	Estimated: *(majority of NSW spine control at better than 250 mm)*	Computed for all control: Majority at better than 20 mm
Relative Uncertainty	Within individual adjustments	Between all control Full VCV computed
Update Frequency	Frozen *(at NSW spine control),* ad hoc local infill	Live / as required
Tectonic Plate Motion/Rotation	Ignored	Applied
Deformation	Limited application (only current coordinates available)	Monitored / applied

2 METHOD

The methodology for gathering, adjusting, testing and cleaning all GNSS vector data at LPI appears relatively straightforward (Figure 3). However, actually accomplishing the few steps required to collate and clean this data in a single adjustment was extremely time consuming, and indeed is still ongoing. Perhaps an initial aversion to opening all those 'big yellow folders' and piecing together a single adjustment for all the observations was well founded. The sheer number of observations precluded manual data mining and analysis. New software tools had to be developed for data and metadata gathering, for translation of the data, for adjustment, and for assessment of an adjustment of unprecedented size.

Not surprisingly, many of the same issues that were discussed during the original GDA94 adjustment were encountered here again, but on a significantly larger scale. These include stations with standpoint errors, non-standard names and archives in a multitude of data formats. For this study, only GNSS data has been collected; terrestrial data such as directions, distances and levelling is currently excluded.

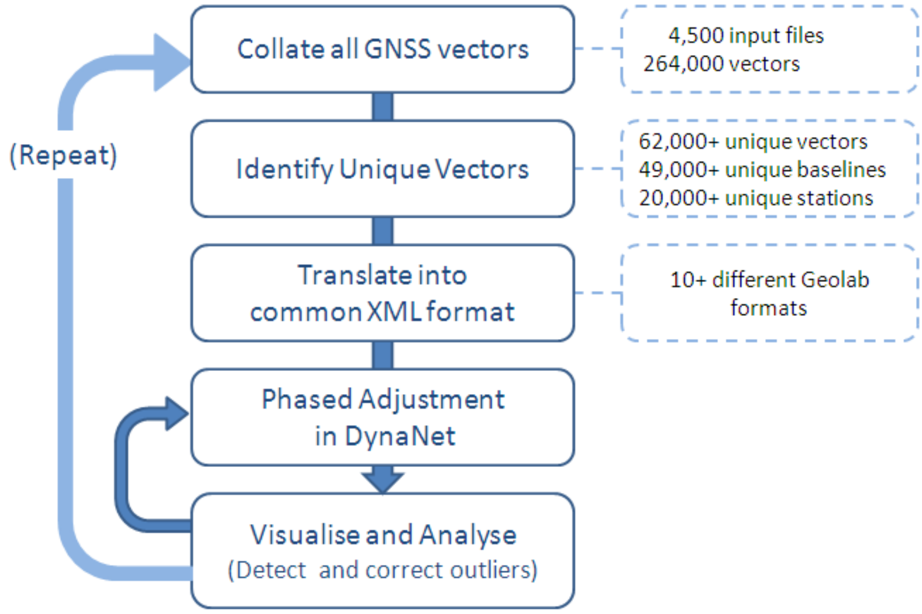

Figure 3: Flowchart of GNSS vector acquisition and cleaning

2.1 Data Sourcing (GNSS Vectors)

Historically, GPS observations and adjustments have been stored digitally within job-specific folders in the archives of LPI, and held uniquely at several offices around NSW. The large number of folders and significant repetition of information precluded manual investigation of these archives. Instead, a number of software 'spiders' were written to automatically scour every directory and file in these digital archives in search of GNSS observations.

The GNSS observations in these files were collated, compared and characterised according to the following criteria:

- Retain only one copy of each unique 3D vector (by spatial distance and station names).
- Adopt the weighting as per the most recent adjustment that employed this vector.
- Retain only vectors involving standard SCIMS naming (i.e. TS, PM, SS, MM).
- Exclude vectors that were flagged in the original adjustment (unless investigated).

More than 10 distinct GeoLab input formats were encountered. For instance, combinations of ΔXYZ, ΔNEH, ΔPLH, correlation matrices, covariance matrices, with and without SIGM weighting records, etc. were used variously at different times, and outliers were flagged in slightly different ways. Different weighting strategies have been employed as our understanding of GNSS processing has developed and matured over the decades. This made it difficult to combine and analyse the data. Ultimately, all formats discovered are automatically converted (by novel software) to a single format, an XML input in ΔXYZ with VCV matrices, suitable for use with DynaNet. As part of another ICSM project, a universal geodetic XML is currently under development for the future transfer of geodetic data between jurisdictions (Donnelly et al., 2013).

First in February 2012, and subsequently in August 2012, the LPI archives were data-mined and more than 264,000 GNSS vectors were located and characterised. Approximately 62,000 vectors were found to be unique, comprised of more than 49,000 unique baselines covering more than 20,000 stations (Figure 3 and Table 3). Interestingly, more than 80% of these baselines were found to have been observed on a single occasion. The most common GNSS vector length was approximately 600 m. In total, there were 21 times as many GNSS baselines as in the initial GDA94 adjustment.

Table 3: Numbers of stations and observations in GDA94 (Watson, 2006) vs. current adjustment.

(values rounded to nearest 100)	**GDA94 (NSW)**	**All GNSS Vectors (NSW)**	**Increase**
Stations	3,000	20,000+	6.8 x
GPS/GNSS Baselines	2,900	62,000+	21 x

2.2 Adjustment Constraints

Discussions with Geoscience Australia indicate that any new national adjustment would be constrained to CORS stations contributing to the Asia-Pacific Reference Frame (APREF – GA, 2012c) and also to AUSPOS solutions of significant duration. In this way, the CORS essentially become the new primary control and offer very high precision and ongoing monitoring of the stability of our datum at a density much greater than the ANN stations of GDA94.

The NSW network of GNSS observations is strongly linked to this CORS backbone by virtue of the ‘local tie surveys’ which provide connections between each CORSnet-NSW station and the local ground control in the area (Gowans and Grinter, 2013). At the time of the data mining, about 60% of the CORSnet-NSW stations had been explicitly connected to the surrounding control. As more of these connections are observed (and/or data-mined), the constraint supplied by the CORS will further improve the positional uncertainty across the network.

For the purposes of data-cleaning and initial testing, the GNSS vectors collated above are constrained by the Regulation 13 certification coordinates (GA, 2012d) of the CORSnet-NSW stations (115 in operation at the time of writing), and also 500+ AUSPOS solutions of greater than 4 hours at various stations across the state (Figure 4). The GNSS CORS are constrained to 10 mm in each X,Y,Z component, and the AUSPOS stations to 20 mm in each component, These constraints will be reviewed closely for use in any final datum product.

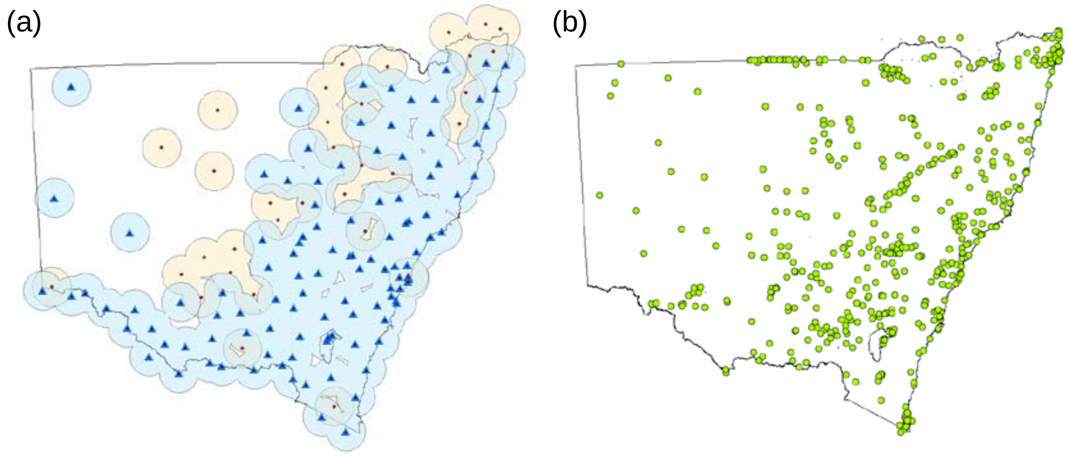

Figure 4: Datum constraining stations – (a) CORSnet-NSW stations (operational in blue and proposed in beige) with nominal 50 km operational radius, and (b) AUSPOS observations of more than 4 hours (green circles).

2.3 Adjustment and Ongoing Analysis

The simultaneous adjustment of all observations using DynaNet and the analysis of the results is an ongoing task. The large array of output information is being visually and numerically assessed using new software tools developed in house, to identify gross errors and outliers. These analyses included sorting the data before and/or after adjustment by 'observed minus expected' comparisons, observation residuals (raw and normalised), observation corrections, coordinate corrections, and vertical and horizontal Positional Uncertainty to name a few.

As expected, by combining adjustments that were previously been kept in isolation, the adjustment was initially plagued with issues relating to station naming and standpoint identification. These gross errors are generally detected by visual inspection, or identification of the 'worst' performing station or measurement. For this purpose, the display of the data geographically in ArcMap (see Figures 4-9) is invaluable for detecting patterns and outliers in the data. For example, if 'local' stations in an adjustment are labeled generically as 'PM1' in their original adjustments, then they would appear as very long and grossly incorrect vectors ending on the Far North Coast, where the true PM1 actually resides (Figure 5a). The common practice of naming a temporary station 'STN99' is fine in context, but no one ever considered the implications of combining multiple adjustments each with a STN99!

Additionally, since observations that span several decades are combined, different naming of the same station will result in two (or more) stations adjusting to coordinates in very close proximity (Figure 5b). For example, the station named 'TS5458' was re-named 'TS12033-3' when the original trigonometric station was replaced, but both names are present in the collated vectors. Of course, the inevitable typo or two were also flagged as significant outliers and easily fixed after reference to the original field notes. Note that these issues are normally detected by the quality control preceding a normal SCIMS update, but the current study has circumvented these procedures by gathering the raw vector data from the archives.

At the time of writing, the majority of these simple issues have been cleaned up, but the adjustment still continues to undergo additional cleaning. Smaller and more subtle errors are being detected through their effects on the surrounding adjustment as described in the sections below.

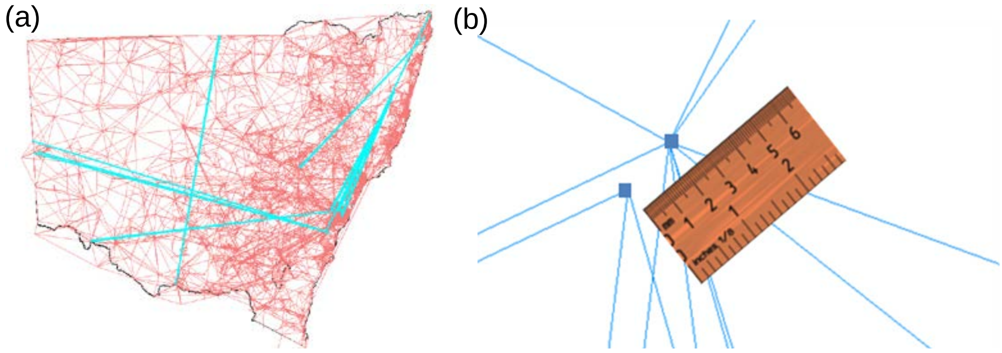

Figure 5: Incorrect station labels resulting in (a) erroneous baselines or (b) station duplication.

It is worth noting that the preparation for GDA94 involved a dedicated team of up to a dozen people working manually to collate and assess the data and the adjustment. In contrast, the current preparation of NSW data for GDA201x involves the part-time labours of two LPI staff who are primarily developing a suite of automated tools for the collation of the data, and where possible the automated detection and mitigation of errors. Most of the tools that have been developed are novel for this project, but will aid the organisation by allowing greater visualisation and interrogation of our existing dataset.

2.4 Adjustment Results

At the time of writing, only a handful of issues remain in the simultaneous adjustment of all GNSS vectors in LPI's archive. These have been flagged as orange or red vectors (Figure 6), with normalised residuals (NR) greater than 3 or 10 respectively. Recall that 99.7% of the NR values are expected to fall within ±3 (and ~95% within ±2 etc.). Therefore, a NR of ±3 was deemed a good cut-off value for the initial detection of outliers. The larger the NR, the less well the observation (and its weighting) fits the adjustment. Most (> 95%) of the observations fit well in this test dataset and the remaining outliers are generally being pushed around by a few vectors with very poor fit (see section 3).

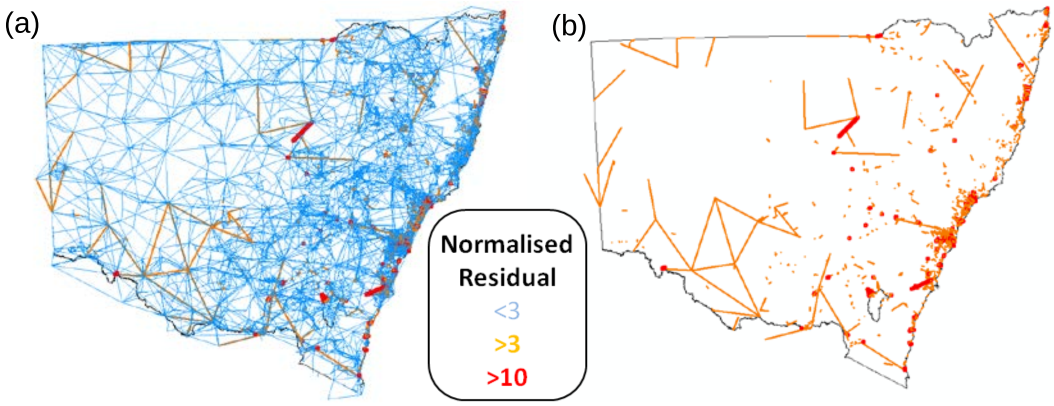

Figure 6: All GNSS baselines post DynaNet adjustment, validated observations in blue, flagged outliers in orange and red – (a) all GNSS baselines, and (b) outlier observations with NR > 3 only.

From this simultaneous adjustment, an initial assessment of PU (95% confidence) of all stations can be made. With the current CORS and AUSPOS constraints, the majority of stations have a computed horizontal PU of better than 20 mm, i.e. *more than an order of magnitude better than the estimated horizontal PU for GDA94.* As expected, stations in the western portion of NSW have higher PU (up to 100 mm), due to the relative scarcity of observations (see Figure 6).

2.5 GDA94(1997) SCIMS vs. GD94(2010) State-Wide Adjustment

By constraining the adjustment to the CORSnet-NSW Regulation 13 coordinates and AUSPOS solutions, the adjustment returns coordinates in GDA94(2010). As discussed in section 2.2, any national adjustment would be constrained in a similar way. Figure 7 shows a preliminary sample of the amount and direction by which existing SCIMS coordinates are expected to change after a new state-wide (or nationwide) adjustment. As expected, these changes mirror distortions highlighted previously by the CORSnet-NSW network (Haasdyk et al., 2010), i.e. up to 250 mm at the North Coast, and generally 100 mm or less in the Newcastle, Sydney, Wollongong, South Coast and inland areas. In some areas, this GNSS dataset is insufficient to determine these distortions.

Figure 7 demonstrates visually why a site transformation is restricted to a 'local' area where such changes from GDA94(1997) to GDA94(2010) are quite consistent (Haasdyk and Janssen, 2012). Obviously a complex transformation such as the NTv2 grid method employed between AGD66/84 and GDA94 would be required for datasets spanning larger areas.

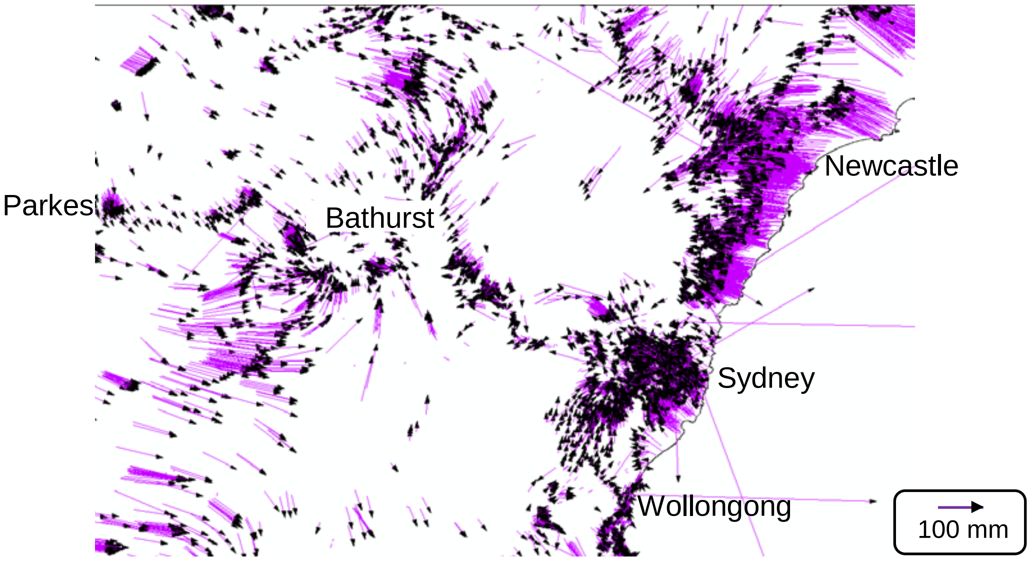

Figure 7: Expected change from GDA94(1997) SCIMS coordinates to a new adjustment in GDA94(2010) shown as purple vectors with black arrow heads. Some black arrow heads are omitted on shorter vectors for display purposes. Only established stations are included (Class C or better). Note that some large coordinate changes are present, indicating gross movement of individual stations.

3 THE NEXT STEPS

Despite the progress made thus far, there is still a lot of work to do before this dataset is free from gross errors and appropriately weighted for a simultaneous adjustment. The sections below describe the next steps for this project, and the methods by which the dataset is still being actively cleaned before submission (in whole or part) to Geoscience Australia by the middle of this year.

3.1 Identification and Elimination of Remaining Gross Errors

The task of eliminating the remaining gross errors is still ongoing, purely because it requires significant manual interaction. It is well known that the least squares method spreads errors across an adjustment, which can make the identification of offending observation(s) difficult and time consuming. Every tool that is developed in-house highlights a new class of issues and sparks additional adjustments and/or debugging of the collation and translation software.

As an example, a region on the South Coast of NSW suggested that a 2-metre shift between SCIMS coordinates and the new adjustment may be expected (Figure 8). Upon investigation, it was found that this region of the network has suffered from a single standpoint error, i.e. the GNSS vector from the archive suggested that a certain trigonometric station was occupied, but investigation of the field notes revealed that the real occupation was of an eccentric mark, 2 metres away. Correcting this gross error improves the adjustment in this region significantly by removing the erroneous stress from the network.

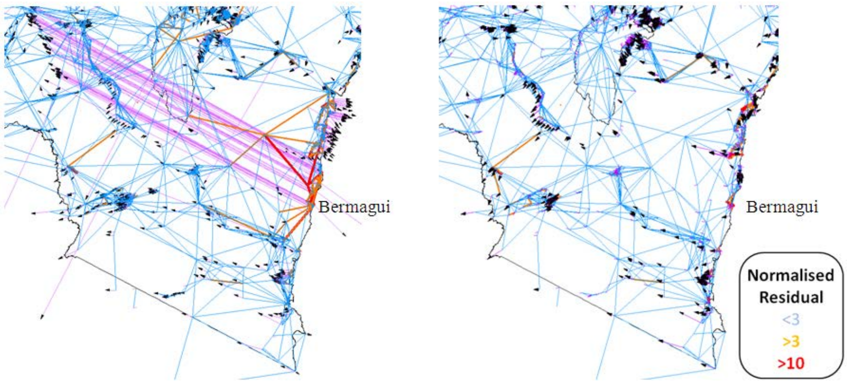

Figure 8: DynaNet adjustment before and after gross error elimination with validated observations in blue, flagged outliers in orange and red. Purple vectors with black arrow heads indicate expected change in coordinates of approximately 2 metres.

Finally, after the dataset is considered clean, it will be instructive to re-introduce known errors of various magnitudes in order to determine the smallest error that can be detected in the existing dataset via least squares, or other error detection methodologies.

3.2 Investigation of Station Movement

Investigations into the remaining outliers suggest that a number of our trigonometric stations are subject to significant land subsidence and/or movements of the monuments. For example, TS5551 Milbrodale (southwest of Singleton) has been known to move both horizontally and vertically due to nearby mining. If the adjustment assumes that this station has constant coordinates over time, observations taken at various epochs will significantly disagree with each other (Figure 9a). When this station's coordinates are solved separately for each campaign, then the outliers in the adjustment disappear (Figure 9b). TS5551 is shown to have moved more than 0.6 m horizontally and 1.6 m vertically between 2002 and 2009!

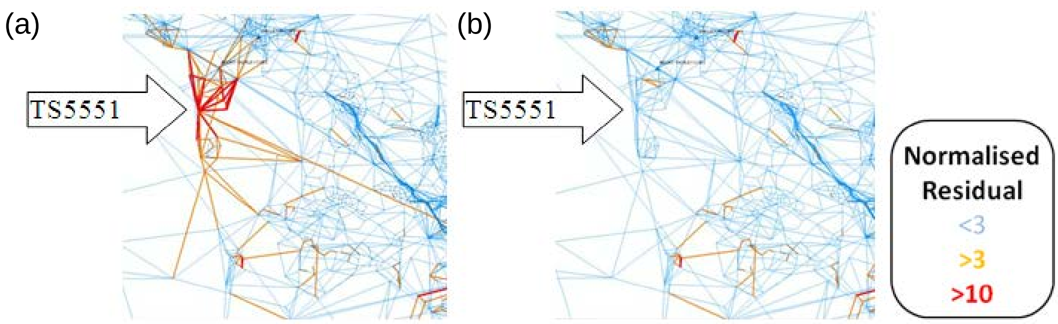

Figure 9: Effect of land subsidence on measurement residuals for observations spanning many years – (a) TS5551 assumed to be constant over time, and (b) TS5551 solved separately at each campaign observed.

It remains to be determined which other stations are subject to movement, but it is quite obvious that stations on silos, water reservoirs and other tall buildings will need to receive special attention, as well as stations near known mining and subsidence areas (Figure 10).

Figure 10: (a) TS7350 Euston showing significant measurement residuals, (b) TS1619 Coonamble on top a water reservoir, and (c) TS6697 Robinson, a concrete pillar on a building, next to significant 'earthworks'.

3.3 Inclusion of Additional Stations

As noted above, this analysis was limited to stations already issued with SCIMS coordinates and therefore following standard SCIMS naming conventions (i.e. TS, PM, SS, MM). Additional vectors have been identified with stations that served as standpoints in a GPS network but were never intended for inclusion in SCIMS, such as various GI pipes, dumpies and iron spikes. These vectors could be useful and necessary for improving the connections and redundancy of the network, and for subsequent control of associated terrestrial surveys.

3.4 Mining New Data

Since the most recent data mining event in August 2012, additional observations have been gathered by LPI staff or submitted by external organisations. Now that our methodologies are maturing, and our existing dataset is nearly clean, it will be necessary to increase the frequency of data mining so that our analysis is more current and our visualisation tools become more and more useful for the planning of future campaigns.

3.5 Mining Metadata

Since the Australian tectonic plate is not just moving but also rotating (Stanaway et al., 2012), any state-wide or national GDA201x adjustment will require vectors to be rotated to a common epoch before adjustment. For this reason, as well as for the completeness of any new database of geodetic observations, additional metadata is required for each observation. Now that the vectors of interest have been identified, more information is required about the date of observation, the GPS/GNSS solution type, etc. Gathering relevant metadata will require the development of more tools and another mining of the archives.

3.6 Distances, Directions and Levelling

At this stage, the intention is to only include GNSS vector observations in this large scale geodetic adjustment. The 20,000 marks involved represent only about 8% of the marks currently held in SCIMS, and 14% of the existing 140,000 established marks with Class C or better. Other marks in SCIMS are connected by terrestrial techniques such as directions and distances or by levelling only. Where necessary, other observations can be included in order to add marks of particular significance. Gathering all terrestrial and levelling observations could be an even larger task than the GNSS data mining described here. In the absence of a simultaneous adjustment of all observation types, subsequent adjustments (or novel transformations) will need to be employed to re-coordinate the remaining SCIMS marks as required.

While the present adjustment is 3-dimensional, the height values produced are ellipsoidal heights in GDA94(2010). The scope of this adjustment does not currently include any reference to the Australian Height Datum (AHD) and would not affect AHD71 values in SCIMS. However, through the application of AUSGeoid09 (Brown et al., 2011), orthometric levelling observations could be added to this state-wide adjustment to detect gross errors and improve the AHD71 surface.

3.7 Connection of GNSS Islands

Our investigation has identified a number of ‘islands’ of GNSS observations that are currently not connected to any of the constraining CORS or AUSPOS stations. This knowledge helps to inform our field staff when planning future campaigns in order to obtain these connections, thus strengthening the network. In addition, ‘local tie surveys’ to some existing and all new CORS still need to be observed (and/or data-mined) in order to fully constrain the network to GDA94(2010).

3.8 Review of Vector Weighting and Constraints

This GNSS archive dataset has been adjusted ‘en masse’ using the weighting as originally applied to each observation. However, these observations have been gathered over two decades, and a variety of business rules regarding the weighting of GNSS vectors have been applied over that time. It would be useful to re-weight each GNSS vector using the same algorithm (to be determined) before the final state-wide adjustment. Similarly, the constraints applied have been preliminary and for testing purposes only. A review of the constraints applied is required before any final national adjustment.

4 CONCLUDING REMARKS

History has demonstrated that the drivers for datum change generally include the increased quality and quantity of observations and the resulting user requirement to remove measurable distortions. Now, using literally a continuous stream of high-precision geodetic observations from a large number of stations, the limitations and distortions of GDA94 have been demonstrated and a new Australian datum is being carefully considered. Many new and more precise geodetic observations are available between local ground control since the last datum update. All states and territories in Australia have been mandated to collate and clean a primary network of geodetic data to facilitate investigations into a possible datum update. The intention is to adjust all observations, nationwide, in a simultaneous adjustment using the DynaNet software in order to create an improved and consistent datum across Australia, here hypothetically called GDA201x.

To this end, a collection of all geodetic GNSS observations across NSW has been created from the archives at LPI. At the time of writing, this dataset contains more than 20,000 stations and 62,000 unique baselines. Not surprisingly, a simultaneous state-wide adjustment of this dataset, combining adjustments spanning thousands of kilometres and two decades, requires a significant amount of testing and cleaning. Significant progress has already been made. The handful of errors remaining can generally be attributed to outstanding standpoint or naming errors in the dataset, and/or marks that are known to be unstable.

The computed Positional Uncertainty of the 20,000+ stations in the adjustment is better than 20 mm for the majority of marks in the network, and up to 100 mm in the western division, i.e. up to an order of magnitude better than in GDA94. These values will likely improve with more connections to CORS and AUSPOS in future adjustments. Expected horizontal coordinate changes between SCIMS and any new adjustment on GDA94(2010) are up to 250 mm at the North Coast, and generally 100 mm or less in the Newcastle, Sydney, Wollongong, South Coast and inland areas.

Through the DynaNet software, we now have the ability to incorporate new observations into a state-wide or national adjustment immediately after they become available. Therefore, it makes sense to have these data ready and cleaned, not just for this datum definition, but in preparation for future requirements. LPI aims to create a current, searchable and visualised database of its geodetic observations to assist with planning and datum maintenance. LPI is on track to submit this extensive network, far in excess of the requested primary network, for initial testing of a nationwide simultaneous adjustment in the middle of 2013.

In summary, we have ‘dusted off’ all of our yellow folders, tipped them all on the floor, sorted them out and put them back together again. The result is a new massive yellow folder with a desk of its own that is a living and changing adjustment across NSW. When soon combined with the similar ‘folders’ from all other states and territories, we will start to see a homogenous datum across Australia that is capable of adapting to the next-generation(s) of technologies and spatial data.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the decades of work that have gone into collecting, processing, adjusting and maintaining quality control over the wealth of survey information held by NSW Land and Property Information. Many thanks also to Roger Fraser (VIC DSE) for his continual work to improve DynaNet and to Steve Tarbit (QLD DERM) for a useful DynaNet-to-ArcMap utility.

REFERENCES

Allman J.S. and Veenstra C. (1984) Geodetic Model of Australia 1982, Technical Report No. 33, Division of National Mapping, Canberra, 60pp.

ANZLIC (2010) Spatial Information Council National Positioning Infrastructure (NPI) policy, http://www.lpi.nsw.gov.au/__data/assets/pdf_file/0020/146702/National_Positioning_Infrastructure_Policy_final_27-8-2010.pdf (accessed Jan 2013).

Australian Government (1998) National Measurement Act 1960 (Cth) – Recognized-value standard of measurement of position (repealed), http://www.comlaw.gov.au/Details/F2009B00151 (accessed Jan 2013).

Australian Government (2012) National Measurement Act 1960 (Cth) – Recognized-value standard of measurement of position 2012 (No. 1), http://www.comlaw.gov.au/Details/F2012L00800 (accessed Jan 2013).

Bomford A.G. (1967) The geodetic adjustment of Australia, 1963-66, *Survey Review*, 19(144), 52-71.

Brown N.J., Featherstone W.E., Hu G. and Johnston G.M. (2011) AUSGeoid09: A more direct and more accurate model for converting ellipsoidal heights to AHD heights, *Journal of Spatial Science*, 56(1), 27-37.

Caissy M., Agrotis L., Weber G., Hernandez-Pajares M. and Hugentobler U. (2012) Coming soon: The International GNSS Real-Time Service, *GPS World*, 23(6), 52-58.

Dawson J. and Woods A. (2010) ITRF to GDA94 coordinate transformations, *Journal of Applied Geodesy*, 4(4), 189-199.

Donnelly N., Fraser R., Haasdyk J. and Tarbit S. (2013) An XML schema for geodetic data transfer in Australia and New Zealand, to appear in *Proceedings of Surveying & Spatial Sciences Conference (SSSC2013)*, Canberra, Australia, 15-19 April.

Featherstone W.E. (2013) An updated explanation of the Geocentric Datum of Australia (GDA) and its effects upon future mapping, http://www.cage.curtin.edu.au/~will/datum.html (accessed Jan 2013).

GA (2012a) Geocentric Datum of Australia (GDA), http://www.ga.gov.au/earth-monitoring/geodesy/geodetic-datums/GDA.html (accessed Jan 2013).

GA (2012b) AUSPOS – Online GPS processing service, http://www.ga.gov.au/earth-monitoring/geodesy/auspos-online-gps-processing-service.html (accessed Jan 2013).

GA (2012c) Asia-Pacific Reference Frame (APREF), http://www.ga.gov.au/earth-monitoring/geodesy/asia-pacific-reference-frame.html (accessed Jan 2013).

GA (2012d) Regulation 13 certificates, http://www.ga.gov.au/earth-monitoring/geodesy/regulation-13-certificates.html (accessed Jan 2013).

Gowans N. and Grinter T. (2013) Tying it all together: CORSnet-NSW local tie surveys, *Proceedings of Association of Public Authority Surveyors Conference (APAS2013)*, Canberra, Australia, 12-14 March, 104-119.

Haasdyk J. and Janssen V. (2012) Site transformations: A block shift in thinking, *Proceedings of Association of Public Authority Surveyors Conference (APAS2012)*, Wollongong, Australia, 19-21 March, 29-47.

Haasdyk J., Roberts C. and Janssen V. (2010) Automated monitoring of CORSnet-NSW using the Bernese software, *Proceedings of XXIV FIG International Congress 2010*, Sydney, Australia, 11-16 April, 19pp.

ICSM (2006) Geocentric Datum of Australia technical manual, version 2.3, http://www.icsm.gov.au/gda/gdav2.3.pdf (accessed Jan 2013).

ICSM (2007) Standards and practices for control surveys (SP1), version 1.7, http://www.icsm.gov.au/publications/sp1/sp1v1-7.pdf (accessed Jan 2013).

ICSM (2013a) Geocentric Datum of Australia 1994 (GDA94), http://www.icsm.gov.au/gda/index.html (accessed Jan 2013).

ICSM (2013b) Adjustment of the combined state and territory geodetic networks, http://www.icsm.gov.au/gda/spine.html (access Jan 2013).

Janssen V., Haasdyk J., McElroy S. and Kinlyside D. (2011) CORSnet-NSW: Improving positioning infrastructure for New South Wales, *Proceedings of Surveying & Spatial Sciences Institute Biennial International Conference (SSSC2011)*, Wellington, New Zealand, 21-25 November, 395-409.

Janssen V. and McElroy S. (2010) Coordinates and CORSnet-NSW: Dealing with distortions in GDA94, *Position*, 50, 24-27.

Lambert B.P. (1981) Australian geodetic coordinates keeping up with the times, *The Australian Surveyor*, 30(8), 491-504.

Leahy F.J. and Collier P.A. (1998) Dynamic network adjustment and the transition to GDA94, *The Australian Surveyor*, 43(4), 261-272.

LPI (2013a) GDA, http://www.lpi.nsw.gov.au/surveying/geodesy/gda (accessed Jan 2013).

LPI (2013b) Transformation methods, http://www.lpi.nsw.gov.au/surveying/geodesy/gda/transformation_methods (accessed Jan 2013).

LPI (2013c) CORSnet-NSW, http://www.corsnet.com.au/ (accessed Jan 2013).

LPI (2013d) SCIMS online, http://www.lpi.nsw.gov.au/surveying/scims_online (accessed Jan 2013).

Malays S. and Slater J.A. (1994) Maintenance and enhancement of the World Geodetic System 1984, *Proceedings of ION GPS-94*, Salt Lake City, UT, 20-23 September, 17-24.

Manning J. and Harvey W.M. (1992) A national geodetic fiducial network, *The Australian Surveyor*, 37(2), 87-90.

NMC (1986) The Australian Geodetic Datum technical manual, Special Publication No. 10, National Mapping Council of Australia, Canberra, 65pp.

Stanaway R., Roberts C., Blick G. and Crook C. (2012) Four dimensional deformation modelling, the link between international, regional and local reference frames, *Proceedings of FIG Working Week 2012*, Rome, Italy, 6-10 May, 21pp.

Watson T. (2006) National adjustment for the implementation of the Geocentric Datum of Australia, Unpublished Report, NSW Department of Lands.