QUALITY CONSIDERATIONS FOR OPTIMAL POSITIONING WHEN INTEGRATING SPATIAL DATA

S. Hope[1] and F. Gielsdorf[2]

[1]Dept. of Geomatics, The University of Melbourne, VIC 3010 Australia
Ph. + 61 3 8344 3176, Fax +61 3 8344 3333
s.hope3@pgrad.unimelb.edu.au
[2]technet GmbH, Maassenstrasse 14, 10777 Berlin, Germany
Ph. +49 30 215 4020, Fax +49 30 215 4027
frank.gielsdorf@technet-gmbh.com

ABSTRACT

As spatial data producers are entering an era of data maintenance, new problems are emerging with respect to data quality. The availability of high accuracy positioning technologies, such as Global Navigation Satellite Systems GNSS, has facilitated the rapid collection of new data. This data is typically of higher quality than the legacy database into which it is to be integrated and, as a result, corresponding points may not coincide. The question has arisen as to how this new data can be best integrated into an existing dataset and also, as to what is the accuracy of the resulting data.

This paper considers the positional accuracy and topological consistency aspects of spatial data quality as new, higher accuracy data is integrated into an existing dataset. A methodology is developed that provides optimal positioning solutions by resolving the best fit between the new data and the legacy database, whilst preserving spatial relationships that exist among features. The method uses positional information, together with its associated accuracy, in combination with geometric and topological constraints in a rigorous process based on least squares. In addition, quality information at the point level is provided, enabling the spatial variation in positional accuracy of the resultant dataset to be portrayed. The developed method is applied in a case study to upgrade a subset of the Victorian cadastral database using data from an extensive survey in the area.

BIOGRAPHY OF PRESENTERS

Sue Hope is a PhD candidate at the University of Melbourne, supported by the Cooperative Research Centre for Spatial Information.

Dr. Frank Gielsdorf is software developer and consulting engineer at the technet GmbH as well as private lecturer at the University of Technology of Berlin.

INTRODUCTION

Many spatial datasets in use today were generated through digitisation of existing paper maps. As a result, they are of relatively low positional accuracy when compared to high resolution imagery and positioning devices based on Global Navigation Satellite Systems (GNSS). Data custodians, such as the Department of Sustainability and Environment (DSE) in Victoria, are looking for ways to upgrade their existing datasets using available higher accuracy data. However, they need a method that accounts for the differing accuracies of the input data and provides estimates of the spatial variation in quality of the upgraded dataset. In addition, they may require consideration of geometric and topological constraints in the data integration process.

This paper describes a rigorous approach to positional accuracy improvement, with an application to upgrade a subset of the Victorian cadastral database. The developed method determines optimal positions and resultant quality parameters using all of the information provided, including the associated accuracies. Where possible, spatial relationships that provide additional positioning information are also incorporated into the integration process. The result is an upgraded dataset, with updated measures of positional accuracy at the point level. In addition, geometric properties and topological relationships between features are preserved.

The paper is structured as follows. The next section describes the problems being encountered as providers of spatial datasets are moving into an era of data maintenance and introduces the motivation for the current research. Then, the concepts behind a solution to these problems are discussed. This is followed by a section that details methods to enable the detection of corresponding features based solely on geometric properties. A description of the means to incorporate additional geometric and topological constraints into the adjustment process comes next, followed by a case study detailing the use of data from an extensive survey in regional Victoria to upgrade part of the cadastral database. Lastly, the paper concludes with some closing comments.

BACKGROUND

Over the last thirty years, GIS have emerged as the pre-eminent tools for land management and spatial analysis. Their first appearance was accompanied by an urgent need for digital data. Digitisation of existing paper maps was considered to be a time- and cost-effective way to meet this demand and many digital datasets in use today were compiled in this manner. The positional accuracy of such datasets is dependent upon the scale of the source maps and the digitisation accuracy. A typical quality statement might report that 90% of well-defined points are within 1mm at plot scale of their true location (DSE, 2003). For a dataset sourced from a 1:25000 scale map, this translates to 90% of well-defined points being within 25m of their true position.

Meanwhile, other spatial technologies have evolved rapidly in recent years. For example, positioning devices using GNSS are in common usage. Handheld devices can instantaneously provide point locations with an absolute accuracy of around 10m, whereas linking to continually operating reference stations, either through post-processing or in real time, can lead to positioning solutions with centimetric level accuracy. Similarly, the availability and resolution of satellite and aerial imagery have increased markedly in recent years. For instance, the Coordinated Imagery Program currently underway in Victoria aims to coordinate access to imagery across the state with a spatial resolution of 15cm. Such initiatives, together with internet applications like Google Earth, are leading to greater use of imagery, particularly as underlays to other spatial datasets.

However, as applications of spatial data are becoming more widespread, discrepancies are increasingly apparent. Datasets obtained by digitising low scale paper maps will typically not align well with today's high resolution imagery. This is leading to marked discrepancies being observed (Fig. 1). As more data consumers underlay their datasets with imagery, custodians are recognising the need to upgrade their products.

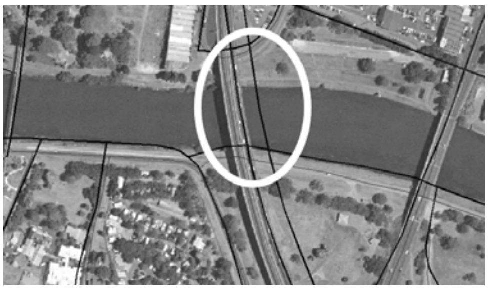

Fig. 1: Road network dataset (black lines) overlaid on a rectified orthophoto. The circle highlights a discrepancy where the road crosses the river. Source: Ramm (2006).

Similar discrepancies are observed when data points are collected using GNSS, for example in asset management, and compared to a database of lower absolute accuracy. Often, the higher accuracy data is to be incorporated into the legacy dataset. In the example of the cadastre, survey-accurate coordinates from digital sub-divisions are continually being provided for integration into the cadastral database. Simply replacing the less accurate coordinates of the bounding polygon in the existing cadastre with the new, higher accuracy locations would result in distortion of the neighbouring parcels (Fig. 2). Instead, a more rigorous means of integrating the new data is necessary.

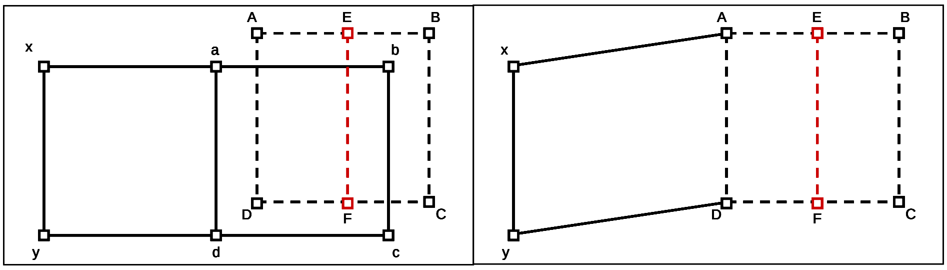

Fig. 2: (Left) cadastral boundaries (solid lines) and higher accuracy sub-division (dashed lines). (Right) polygon xady is distorted if points abcd are replaced by ABCD.

POSITIONAL ACCURACY IMPROVEMENT OF SPATIAL DATASETS

The escalating use of higher accuracy positioning technologies is leading to a recognised need for many data providers to upgrade their legacy datasets. This would also alleviate the problems being encountered when attempting to integrate local pockets of higher accuracy data into existing databases. The two problems are intertwined as the development of a rigorous method to integrate higher accuracy data into legacy datasets would itself lead to upgrade of those datasets. Indeed, such a method could be extended to provide an optimal solution whenever datasets with multiple representations of the same features are overlaid.

Data points obtained using GNSS have high absolute positional accuracy and can be used to improve the accuracy of corresponding points within an existing database. However, the relative geometry of the existing dataset is usually relatively good, so it is not sufficient to simply shift the points that have been more accurately located. The data integration method must propagate any adjustments to neighbouring points in the legacy dataset. This is a result of point coordinates in many spatial datasets showing distance-dependent correlation, arising from both the measuring and plotting processes. These correlations have to be considered during data integration, but recording covariance matrices for all of the coordinates in a spatial dataset would lead to enormous storage demands and is therefore not practical. Instead, they can be modelled by introducing relative observations between neighbouring points. These observations can be the original measurements, such as distances or alignment bases, or they can be artificially introduced by measuring the distances between adjacent points.

In addition, geometrical constraints such as the maintenance of right angles or parallel lines can be introduced as artificial observations. For example, it might be important to maintain the straightness of roadlines during upgrade of the cadastral dataset, especially as many associated utilities are mapped with reference to these lines. There may also be topological relationships existing between features, such as the end-points of utility pipelines having to meet property boundary lines. These spatial relationships can also be modelled as artificial observations within the data integration process.

If neighbourhood geometry is to be maintained through the inclusion of relative distances, as well as geometric and topological properties, there will be redundancy in the data integration process. It will not be possible to obtain a solution that perfectly meets every constraint. Instead, an adjustment problem is set up, for which an optimal solution can be determined using the method of weighted least squares. All of the available information is considered, and observations are weighted by their recorded accuracy values to determine the solution that best fits the datasets being integrated. Moreover, the method of least squares generates precision values for the calculated parameters, thereby enabling update of the positional accuracy of the upgraded dataset.

Cadastral surveyors have traditionally used least squares adjustments to determine the coordinates of parcel boundaries, and their associated precision, from a set of redundant measurements. Elfick (2001) proposes that the same methods can be used to determine a coordinated cadastre from the original survey measurements held in Land Titles records, suggesting that this may be no more expensive than the annual maintenance of the current digital cadastre database. Such a proposal has been undertaken in the Northern Territory (West & Sarib, 2001) and in New Zealand (Rowe, 2004).

However Victoria has a considerably greater number of land parcels, and this would require a large capital investment. Williamson and Hunter (1996) conclude that the benefits of a coordinated cadastre would probably not outweigh such costs and, instead, recommend an accelerated incremental approach to upgrading the Victorian cadastre. This can be done through a rigorous least squares adjustment of the existing polygon boundaries as new, highly accurate data is integrated. The use of least squares to adjust polygon boundaries after manual digitisation has been demonstrated by several researchers (for example Merrit & Masters, 1999; Tong *et al.*, 2005). These studies have incorporated geometric constraints, such as preservation of straight and parallel lines, right angles and areas, into the adjustment.

Upgrade of spatial datasets through the integration of higher accuracy data requires that links are made between the new data and the existing dataset. In the case of the cadastre being upgraded using a digital sub-division, this may be the identification of which points in the existing database correspond to those bounding the sub-division. Ideally, point or parcel identifiers are available and can be used as keys to relate the two datasets. However, more commonly, no such key is available and corresponding features across the datasets have to be manually identified. This is both time-consuming and costly; a method to automate detection of corresponding features would be beneficial.

IDENTIFYING CORRESPONDING FEATURES

Saalfeld (1988) describes a method to facilitate automated detection of identical nodes in his seminal paper on data conflation. This method is based on proximity and the directions of edges emanating from the node, termed the 'spider function'. Other researchers have proposed methods that also utilise attributes of the datasets (for example Samal *et al.*, 2004). However, the semantics used to describe attributes of datasets vary widely, whereas geometric properties remain invariant to semantics. Here, methods developed by the authors to identify corresponding points based only on geometric properties of the data are described.

The process of identifying corresponding features for data integration is an iterative one. The more salient correspondences are first identified and used in an adjustment to bring the datasets into closer alignment. Further correspondences can then be identified and the process is repeated until no further matches are found. It is therefore imperative that any mis-matches are identified early in the process as they would otherwise influence the detection of subsequent matches. To facilitate this, normalised residuals arising from the least squares adjustment can be calculated for each identified correspondence and the user alerted to any that exceed a set threshold.

The types of correspondences that can be identified depend upon the datasets being integrated. Saalfeld (1988) worked with road networks and matched intersections across the two datasets as these were the most salient features. In the case study described in this paper, points along fencelines had been collected in an extensive survey and were used to adjust a subset of the cadastral database. In this case, two types of correspondences were found to be useful, as illustrated in Fig. 3. The first of these was the matching of corner points across the two datasets. The second of these was the matching of a set of fenceline points to the line between two adjacent cadastral vertices.

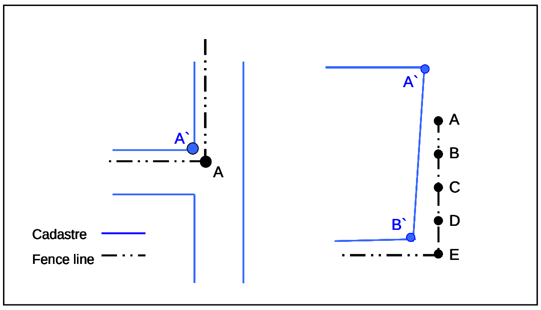

Fig. 3: (Left) matching of corner point A in the fence dataset with cadastral vertex A`. (Right) matching of fence points A-E with cadastral edge A`B`.

A prerequisite for matching corners is to first detect all of the topological corners in each dataset. In a topological sense, a corner is a set of two non-identical edges with one common point. For example, a vertex with four adjacent edges provides six topological corners (Fig. 4). Defining topological corners provides the option to match, for example, the corner of a house with the corner of a parcel, even when the house vertex and the parcel vertex have a different number of adjacent edges. Topological corners within a dataset can be found with a very efficient algorithm based on third level adjacency tensors in sparse notation.

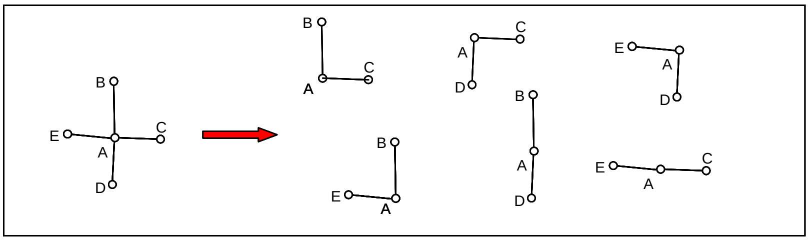

Fig. 4: Definition of topological corners

The geometrical representation of a corner is realized by the corner point and two normalized direction vectors. The parameterization is given by two point coordinates and four components of the normalized direction vectors. The information about the stochastical properties of these six parameters is stored in the corresponding covariance matrix. This is determined from a variance propagation calculation applied on the functional connection between the point coordinates and vector parameters.

Possible matches are detected by treating the six corner parameters as coordinates of a point in a six dimensional manifold. The search for matching candidates happens in a 6D window using a 6D tree. For each potential match, a χ^2-test is performed. If the match is accepted, the identity information of the two corner points is recorded in an adjacency matrix. This adjacency matrix is subsequently checked for ambiguities which are removed. For each remaining unique match, an identity observation based on coordinate differences (with the observation value 0) is generated.

The matching of a set of points to the straight line between two adjacent vertices is achieved by initially performing a quadtree search for candidate points within a rectangular window around each possible edge. If point C is collinear with points A\` and B\`, the cross product of vectors CA’ and CB\` should be zero. The cross product of the two vectors from each candidate point to the end-points of the edge is calculated and the variance of this cross product determined through error propagation. The test value for each candidate point is calculated by dividing the cross product by its variance and is normal-distributed. Whenever more than one point is identified as being collinear with an edge, an alignment constraint can be established. This establishes a local coordinate system and is a more efficient way of modelling the collinearity. If any point is detected as being collinear with more than one edge, both collinearities are removed.

INCORPORATING GEOMETRIC AND TOPOLOGICAL CONSTRAINTS

The identification of corresponding features across two datasets enables links to be made between the datasets. The adjustment process is then able to bring the datasets into alignment. However, any adjustments must be propagated to other data points that have no direct links, to maintain the relative geometry of each dataset and prevent neighbourhood distortions. If known, measurements such as distances and angles between points can be entered directly as observations into the adjustment. More commonly, these are not available and observations relating to the relative geometry have to be artificially generated.

One way to do this is to first sub-divide the plane for each dataset using a Delauney triangulation of all the data points. The coordinate differences along each triangle edge can then be entered as observations into the least squares process (Gielsdorf *et al.*, 2004). Gielsdorf (1997) has shown that if these coordinate differences are weighted as a function of the form of the adjacent triangles, the resulting point shifts across the dataset mimic the behaviour of a true rubber-sheet. The introduction of these relative distances as artificial observations in the least squares adjustment therefore results in membrane-like distortion of the surface to model the distance-dependent correlations.

There may also be specific geometric properties that can be identified within a dataset and that need to be preserved through the adjustment process. For example, a sequence of power poles may lie along a straight line; a corner in the cadastre may be known to be rectangular; or two roadlines may be parallel. These properties can be modelled as observations in the least squares adjustment. However, geometric properties are a special type of artificial observation and should be formulated so that their normalised residuals can be checked. This enables the detection of incorrect properties that might otherwise adversely affect the adjustment. Again, manual identification of each geometric constraint is a time-consuming process and the authors have developed methods to automatically detect these properties within a dataset, as detailed below.

As a first step to detecting collinearities or rectangularities in datasets, it is necessary to find all topological corners. The way to do this has already been described. The corners can then be tested for collinearity or rectangularity. The three points of the corner are collinear if the area of the triangle which they span is equal to zero. The double triangle area can be calculated as the magnitude of the cross product of the two vectors defined by the corner edges. A variance propagation from the point coordinates to the double triangle area provides the standard deviation of that value. The quotient of the triangle area and its standard deviation is normal-distributed and can be tested for significance. If the test is accepted, a straight line observation based on cross product is generated. The three points of a topological corner form a right angle if the scalar product of the two vectors is zero. The standard deviation of the scalar product is again calculated by variance propagation. The normal-distributed test value is the quotient of the scalar product and its standard deviation. If the test is accepted, a rectangularity observation based on scalar product is generated.

To detect parallel lines we use the Hessian normal form of a straight line, $\mathbf{n}^T\mathbf{x} - d = 0$, where $\mathbf{n}$ is the normal vector with $\mathbf{n}^\mathrm{T} = (n_x , n_y)$ and d is the orthogonal distance from the origin. Straight lines with the same normal vector are parallel. The three straight line parameters are treated as coordinates of a point in a three dimensional manifold. The search for candidates to be parallel lines happens in a 3D window using a 3D tree. For each of the detected candidates a χ^2-test is performed and, if this is accepted, a parallelity observation based on cross product is generated. The same approach can be used to find parallel lines with a predefined separation distance and for the matching of identical straight lines in different datasets.

It is also possible to preserve particular topological relationships that may exist between features across two datasets. One example is that of a utilities service pipe having to meet a property boundary line. If the cadastral dataset is adjusted independently of the utilities dataset, the service pipe will no longer meet the property boundary. It is possible to identify spatial relationships existing between the datasets before the cadastral adjustment, then to ensure that the appropriate shift is applied to the associated dataset. This will maintain the relationship between the utilities and the adjusted cadastre and may be the best solution when datasets are maintained separately.

However, this method does not use the topological information in determining the optimal positioning solution within the least squares process. Ideally, if the associated dataset is available during the upgrade, the relationship can be modelled as a constraint in the adjustment itself. For instance, this example could be modelled as a collinearity constraint between the end-point of the service pipe and the two cadastral vertices of the property boundary. Other spatial relationships, such as two features remaining disjoint, can only be modelled as inequalities and are not as easily incorporated into a least squares adjustment. The possibility of including topological constraints in positional accuracy improvement remains a topic of the authors' current research.

CASE STUDY – CADASTRAL UPGRADE IN REGIONAL VICTORIA

A case study was established to demonstrate the ability of adjustment techniques to improve the positional accuracy of a subset of the Victorian cadastre. An extensive survey had been undertaken in North West Victoria as part of the Wimmera Mallee Pipeline Project. In this rural region, the cadastre had been digitised primarily from 1:25000 scale maps and had a reported positional accuracy of 90% of well-defined points being within 25m of their true position. As part of their commitment to improving the overall quality of Vicmap products, the DSE requested a pilot study to improve the positional accuracy of the cadastre in that region, using the survey data.

Method

The survey dataset included approximately 14000 points representing fencelines collected along the proposed route of the pipeline. The points had been collected using real time kinematic GNSS and had a reported precision of 15mm. A polygon bounding the survey data was used to clip the cadastral dataset. The resulting subset of the cadastre contained approximately 89000 points. Both the cadastral dataset and the fenceline data had been provided as ESRI shapefiles, having no explicit topological format. Unique identifiers are required for each point in an adjustment, to ensure that its identity is not lost as the coordinates change. The software GEOgraf was used to translate the datasets into a format that provided this topology. The datasets could then be imported directly into the adjustment software Systra.

The two most suitable means of identifying links between the datasets were found to be corner matching and line matching (Fig. 3). Initially, 25 point identities linking the cadastral dataset to the more accurate fenceline points were manually identified. Points at road intersections were generally used as it was found that these were most easily matched. An adjustment was run using these point identities to pull the cadastre into better alignment with the survey data. The normalised residual for each point identity was checked and any mis-matches removed. A program written to detect collinearity observations between the datasets was then used iteratively, initially with a low threshold value to minimise the chance of mis-identifying constraints. As the datasets were brought into better alignment, more collinearities were identified. In total, 7786 such constraints were detected. These were then used in a strict least squares adjustment of the data and again the normalised residuals were checked to remove any mis-matches. Finally, a proximity adjustment was used to propagate the adjustment to points not directly linked to the fenceline data. This applied the membrane-like distortion over the cadastral dataset, thereby accounting for the correlation between neighbouring points.

Results

Visual inspection of the adjustment results indicated that the cadastre had been adjusted to better align with the survey data. This did not only occur where the cadastral points had been directly linked to fenceline data. Regions with no direct links were also in better alignment due to the adjustment propagation to neighbouring points (Fig. 5).

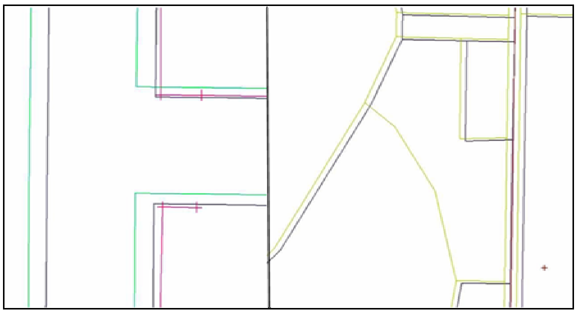

Fig. 5: Adjustment (blue) aligns better to fencelines (red) than original cadastre (green).

The adjustment software provided statistical measures of the precision of each upgraded cadastral coordinate. The positional accuracy of the initial cadastre was reported as 90% of points being within 25m of their true position. The adjusted cadastral coordinates had a mean precision of 1.49m, which, assuming a normal circular error distribution, relates to 90% of points being within 3.2m of their true position. However, more detailed accuracy parameters are available from the adjustment and would enable the spatial variation in positional accuracy of the upgraded cadastre to also be portrayed.

The results reported by the adjustment software are currently being verified in an independent ground survey of the study region. A random sample of 150 points from the adjustment area has been generated. It was ensured that the sample includes points on parcel boundaries that are not road-frontages, as these points are furthest from the direct links made to the survey data. The points are being surveyed using real time kinematic GNSS and these positions will then be used to validate the adjustment results.

CONCLUSION

Positioning technologies such as those based on GNSS enable the rapid collection of data points with high absolute positional accuracy. This data can be used to improve the positional accuracy of existing spatial databases. Since higher accuracy data is typically only available in localised pockets, the relative geometry of the existing database needs to be preserved whilst the data points of higher absolute accuracy are integrated. This leads to an over-determined system, hence an adjustment problem. Additional constraints that preserve geometric or spatial relationships can be incorporated into such an adjustment, thereby maintaining the logical consistency of the data. The result is an optimal positioning solution that takes into account all of the available information, including the associated accuracies. Furthermore, updated measures of the positional accuracy of the resultant dataset are provided at the point level. A case study presented here demonstrates the effectiveness of developed techniques to adjust a subset of the Victorian cadastre using survey data. The reported positions and associated accuracies of the upgraded cadastral points are currently being validated using a ground survey.

Further development of these techniques should enable optimal positioning solutions to be determined whenever direct links, such as corresponding features, can be identified between multiple datasets. As well as enabling positional accuracy improvement of existing datasets, this would improve on current methods for overlaying spatial data and therefore lead to better results in a range of spatial analysis applications. Techniques to extend the adjustment methodology from cadastral upgrade to more general datasets are currently under development. This includes means of matching across point sets, such as linear features that do not display one-to-one vertex correspondence, and the incorporation of additional geometric and topological constraints in the adjustment.

REFERENCES

DSE 2003. *Product Description: Vicmap Property v5.2*. Land information Group, Land Victoria, Department of Sustainability and Environment.

Elfick, M. 2001. Managing the Records which Underpin the Land Tenure System. *International Symposium on SDI*, Melbourne, Australia, 19-20 November.

Gielsdorf, F. 1997. Nachbarschaftstreue Anpassung auf der Basis des Membranmodells. *ZfV Zeitschrift für Vermessungswesen*, 5/1997.

Gielsdorf, F., Gruendig, L. and Aschoff, B. 2004. Positional accuracy improvement - A necessary tool for updating and integration of GIS data, *FIG Working Week,* Athens, Greece, 22-27 May.

Merrit, R. and Masters, E. 1999. Digital cadastral upgrades - A progress report. *Proc. First International Symposium on SDQ*, Hong Kong, 18-20 July, 180-188.

Rowe, G. 2004. The Survey Conversion Project – Making a survey-accurate digital cadastre for New Zealand a reality. www.surveyors.org.nz/user/user_DocView.asp?DocumentID=531&Site_Area=News+And+Events (accessed 2 December, 2006)

Saalfeld, A. 1988. Conflation: Automated map compilation. *International Journal of Geographical Information Systems, 2*(3), 217-228.

Samal, A., Seth, S. and Cueto, K. 2004. A feature-based approach to conflation of geospatial sources. *International Journal of Geographical Information Science, 18*(5), 459-489.

Tong, X.H., Shi, W.Z. and Liu, D.J. 2005. A least squares-based method for adjusting the boundaries of area objects. *Photogrammetric Engineering and Remote Sensing, 71*(2), 189-195.

West, G. and Sarib, R. 2001. Cadastral reform in the Northern Territory 2001 update. *42nd Australian Surveyors Congress*, Brisbane, Australia, 25-28 September.

Williamson, I.P. and Hunter, G.J. 1996. *A scoping study for the establishment of a coordinated cadastre for Victoria*. Report for the Office of Surveyor General and the

Office of Geographic Data Coordination, Department of Treasury and Finance. Department of Geomatics, The University of Melbourne.