DGK Deutsche Geodätische Kommission
bei der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 673

David Nyangau Siriba

Conflation of Provisional Cadastral and Topographical Datasets

München 2011

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

ISSN 0065-5325 ISBN 978-3-7696-5085-3

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 298, Hannover 2011

Deutsche Geodätische Kommission

bei der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 673

Conflation of Provisional Cadastral and Topographical Datasets

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation

von

M.Sc. David Nyangau Siriba

geboren
in Kisii, Kenya

München 2011

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0065-5325 ISBN 978-3-7696-5085-3

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 298, Hannover 2011

Adresse der Deutschen Geodätischen Kommission:

DGK

Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail hornik@dgfi.badw.de ● http://www.dgk.badw.de

Prüfungskommission

Vorsitzender: Prof. Dr.-Ing. Steffen Schön
Referentin: Prof. Dr.-Ing. Monika Sester
Korreferenten: Prof. Dr.-Ing. Winrich Voß
Prof. Dr. Yerach Doytsher

Tag der mündlichen Prüfung: 21.12.2011

ISSN 0065-5325 ISBN 978-3-7696-5085-3

Summary

This dissertation investigated the use of a data conflation approach to improve the geometric quality of provisional cadastral datasets using existing higher quality topographical datasets. The main idea is that if there are similar features in both the provisional cadastral datasets and topographical datasets, then geometric quality improvement of provisional cadastres is possible through data conflation techniques. A data conflation approach was developed that consists of four main steps: preprocessing, derivation of a road network usually represented implicitly in cadastral datasets, road network matching, and local geometric transformation of the cadastral dataset.

The derivation of road network was based on the assumption that the gaps between land parcels in a cadastral dataset represent road reserves. These gaps are extracted as road polygons from which the road centrelines are created. The extraction of the road polygons was achieved as a geometric difference between the parcel polygons in the cadastral dataset and their concave hull. A concave hull algorithm was developed that takes polygon features as inputs. Thiessen polygons and the concave hull of the input polygons are first created. The polygon points contained in the Thiessen polygons that intersect with the convex hull are sampled and constituted into a concave hull through linear referencing. The algorithm ensures a unique and predictable concave hull if there are no extreme concavities in the input polygons. Feature matching and geometric transformation were based on techniques that tolerate local geometric distortions because the provisional cadastral datasets were considered to be locally distorted. The algorithms that were developed for these two steps were based on relaxation labelling and a piece-wise affine transformation respectively.

The approach was experimented with two case studies involving cadastral and topographical datasets from Kenya. Although both case studies are from peri-urban environments, one is typically urban while the other is typically rural. Low matching recalls were obtained for the cadastral datasets from the typically rural case study. This is because there was low feature correspondence between the cadastral datasets and the more accurate and up-to-date corresponding topographical datasets. The cadastral datasets in such areas are generally approximate and are usually not up-to-date. However, the geometric quality of provisional cadastral dataset from the typically rural environment improved more than the one from the typically urban environment. In both cases, survey accuracies could however not be obtained because of the limited geometric accuracy of the reference topographical datasets.

Contrary to the assumption, some gaps between the land parcels were found to represent riparian reserves. Therefore, river networks and other features like fence lines and building footprints could be used with the same approach to establish correspondences between cadastral datasets and topographical datasets. The proposed data conflation approach could be considered by jurisdictions with similar cadastral arrangements as Kenya to enhance the geometric quality of provisional cadastral datasets. However, unlike ordinary topographical objects whose geometric area can change due to geometric transformations, the change in geometric area of land parcels, particularly those that have been registered should be treated with a lot of caution because of possible legal implications.

The Kenya cadastre was described as a basis for understanding the background of the experimental datasets and to avoid any misconception about the nature of cadastre in Kenya. The combined cadastral coverage in Kenya was established to be just about 25%, of which 75% is of approximate positional quality. Different solutions were suggested to improve both the coverage and the geometric quality of the existing cadastral and topographical datasets, among them the data conflation approach presented in this thesis.

Keywords: modern land administration, future cadastres, Kenyan cadastre, provisional cadastres, geometric quality improvement, data conflation, concave hull, linear referencing, relaxation labelling, local geometric transformation

Zusammenfassung

In dieser Arbeit wird eine Methode zur Datenintegration vorgestellt, die zum Ziel hat, die geometrische Qualität von provisorischen, durch grafische Methoden erworbenen Katasterdaten zu verbessern. Die Hauptidee besteht darin, dass es sowohl in den provisorischen Katasterdaten als auch in den topografischen Daten ähnliche Objekte gibt, die eine geometrische Qualitätsverbesserung ermöglichen. Es wurde ein Ansatz zur Datenintegration entwickelt, der in vier Schritte unterteilt werden kann: Aufbereitung, die Extraktion des nur implizit in den Katasterdaten vorhandenen Straßennetzes; die Anpassung dieses Straßennetzes an das Straßennetz eines topographischen Datensatzes höherer Qualität und daraus resultierend, eine Anpassung der Katasterdaten durch eine lokale Transformation.

Die Extraktion des Straßennetzes beruht auf der Annahme, dass die Lücken zwischen den Landparzellen in den kenianischen Katasterdaten Straßen repräsentieren. Diese Flächen werden extrahiert und anhand ihrer Mittellinie eine Straßenachse bestimmt. Um die Lücken zu bestimmen, wurde der Unterschied zwischen den Parzellen und ihrer konkaven Hülle bestimmt. Dazu wurde ein Algorithmus entwickelt, welcher die konkave Hülle von Polygonen ermittelt. Zunächst werden dazu Thiessenpolygone als auch die konvexe Hülle der gegebenen Polygone erstellt. Anschließend werden die Punkte ermittelt, welche in den Thiessenpolygonen enthalten sind die die konvexe Hülle schneiden. Abschließend wird aus diesen Punkten mittels linearer Referenzierung iterativ die konkave Hülle gebildet. Mit diesem Algorithmus lassen sich eindeutige und voraussagbare konkave Hüllen bestimmen, solange es keine extrem großen konkaven Stellen in den gegebenen Polygonen gibt. Die paarweise Anpassung der Objekte und die geometrische Transformation beruhen auf Techniken, die lokale geometrische Verzerrungen tolerieren. Die Algorithmen, diese zwei Schritte basieren auf *Relaxation Labelling* und einer stückweisen affinen Transformation.

Der Anpassungsprozess wurde mit den Datensätzen zweier Regionen aus Kenia getestet. Obwohl beide Regionen peri-urbane Gebiete umfassen, ist die eine eher städtisch während die andere eher ländlich ist. Dabei wurde für das eher ländliche Gebiet eine niedrige Trennfähigkeit (Recall) festgestellt. Das liegt darin begründet, dass die Ähnlichkeit zwischen dem Kataster- und dem topografischen Datensatz eher gering ist. Die Katasterdaten in solchen Gebieten werden normalerweise nur genähert erfasst sind meist nicht mehr aktuell. Besser steht es um die geometrische Qualität der provisorischen Katasterdaten in eher städtischem Gebiet. Auf Grund der begrenzten Genauigkeit des topographischen Referenzdatensatzes können in beiden Fällen jedoch keine Vermessungsgenauigkeiten ermittelt werden.

Allerdings gibt es hier, entgegen unserer Annahme, Flächen ohne Parzellen, die nicht Straßen sind. Es sind parzellenfreie Flächen, die als Überschwemmungsgebiet dienen. Deshalb sollte man in einer weitergehenden Untersuchung Flussnetze, Zäune und Objekte mit geradlinigen Eigenschaften berücksichtigen. Die vorgeschlagene Methode kann für Kenia und Länder mit vergleichbarer Rechtslage im Katasterbereich verwendet werden, um die geometrische Qualität des provisorischen Katasters zu erhöhen. Im Gegensatz zu gewöhnlichen topografischen Objekten, müssen wegen möglicher rechtlicher Auswirkungen jedoch Änderungen im Kataster mit größerer Sorgfalt betrachtet werden.

Das kenianische Katastersystem wird beschrieben, um die Eigenschaften der Testdaten zu verstehen. Das kenianische Kataster umfasst nur etwa 25% der gesamten Landesfläche, wovon etwa 75% nur genäherte Koordinaten besitzen. Es wurden verschiedene Lösungen vorgeschlagen, um die Abdeckung und die geometrische Qualität der vorhandenen Katasterdaten zu erhöhen. Darunter auch die in dieser Arbeit vorgeschlagene Methode zur Datenintegration.

Schlagworte: modernes Landmanagement, zukünftiger Kataster, kenianischer Kataster, provisorische Katasterdaten, geometrische Qualitätsverbesserung, Datenintegration, konkave Hülle, linearen Referenzierung, Relaxation Labelling, stückweise affine Transformation

Contents

Summary **3**

Zusammenfassung **4**

List of Figures **8**

List of Tables **10**

1 Introduction **11**
- 1.1 Background and motivation . . . 11
  - 1.1.1 Modern land management paradigm and the "Butterfly" diagram . . . 11
  - 1.1.2 SDI and multipurpose cadastres . . . 12
  - 1.1.3 Survey-accurate cadastres . . . 12
- 1.2 Geometric quality improvement of cadastres . . . 12
- 1.3 Overview of the proposed approach . . . 14
- 1.4 Objectives and focus . . . 14
- 1.5 Outline of the thesis . . . 15

2 Land Management and the Cadastral Concept **17**
- 2.1 The modern land management paradigm . . . 17
- 2.2 The cadastral concept . . . 18
  - 2.2.1 Hierarchy of concepts related to cadastre . . . 18
  - 2.2.2 Land registration and land recording . . . 19
  - 2.2.3 Categories of cadastres . . . 21
- 2.3 Cadastral boundaries . . . 21
  - 2.3.1 Land parcel boundaries and their documentation . . . 21
  - 2.3.2 Hierarchy of boundary evidence . . . 22
- 2.4 Homogeneous cadastral framework . . . 23
  - 2.4.1 Cadastral surveying . . . 24
  - 2.4.2 Cadastral mapping and geometric accuracies . . . 27
- 2.5 Evolution of the cadastral concept . . . 28
  - 2.5.1 Evolution of applications . . . 28
  - 2.5.2 The changing concept . . . 29
  - 2.5.3 The changing scope . . . 30
  - 2.5.4 From mapping to data modelling . . . 31
  - 2.5.5 Characteristics of future cadastres . . . 33

3 The Kenyan Cadastre and Modern Land Administration **35**
- 3.1 Historical outline of the Kenyan cadastre . . . 35
  - 3.1.1 Before 1902 . . . 35
  - 3.1.2 Between 1902 and 1945 . . . 35
  - 3.1.3 Between 1945 and 1963 . . . 36
  - 3.1.4 1963 to present . . . 36
- 3.2 Land tenure and land tenure patterns . . . 38
- 3.3 The Kenyan cadastre . . . 39
  - 3.3.1 Cadastral maps . . . 39
  - 3.3.2 Common map scales and typical positional accuracies . . . 44
  - 3.3.3 Cadastral coverage . . . 44
- 3.4 Nationwide topographical mapping . . . 44
  - 3.4.1 Large scale base topographical maps . . . 45
  - 3.4.2 Basic topographical mapping . . . 45

3.5 The Kenyan cadastre and modern land administration . . . 46
3.5.1 Survey-accurate cadastres . . . 46
3.5.2 Object-oriented cadastres . . . 47
3.5.3 3D and Time-enabled cadastres . . . 47
3.5.4 Real-time cadastres . . . 47
3.5.5 Regional and global cadastres . . . 48
3.5.6 Organic cadastres . . . 48
3.6 Summary of challenges and suggested solutions . . . 48

4 Quality Improvement and Spatial Data Integration 49
4.1 Spatial data quality and uncertainty . . . 49
4.2 Geometric quality improvement of different cadastral datasets . . . 49
4.2.1 Numerical cadastres . . . 50
4.2.2 Graphical cadastres . . . 51
4.2.3 Conveyance deeds . . . 51
4.2.4 Metes and bounds . . . 52
4.3 Quality improvement of provisional graphical cadastres . . . 52
4.4 Interoperability and data integration . . . 53
4.5 Basic concepts in spatial data integration . . . 54
4.6 Heterogeneities in spatial data . . . 55
4.6.1 Semantic heterogeneity . . . 55
4.6.2 Schematic heterogeneity . . . 56
4.6.3 Structural and syntactic heterogeneity . . . 57
4.7 Spatial data integration workflow . . . 57
4.7.1 Pre-integration . . . 57
4.7.2 Correspondence investigation . . . 58
4.7.3 Integration . . . 59

5 Conflation of Provisional Cadastral Datasets 61
5.1 Data conflation of graphical cadastres . . . 61
5.2 Pre-processing . . . 61
5.3 Derivation of a road network from a cadastral dataset . . . 62
5.3.1 Road polygons from cadastral datasets . . . 63
5.3.2 Concave hull and linear referencing . . . 63
5.3.3 Concave hull through linear referencing . . . 65
5.3.4 Creation of road centerline network through straight skeletons . . . 69
5.4 Feature matching . . . 70
5.4.1 Point feature matching using relaxation labelling . . . 71
5.4.2 Linear network matching using path matching algorithm . . . 76
5.4.3 Finding corresponding vertices of two similar linear features . . . 78
5.5 Geometric transformation . . . 79
5.6 Piece-wise affine transformation . . . 80
5.6.1 Delaunay triangulation . . . 80
5.6.2 Formulation . . . 81
5.7 Assessment of quality improvement . . . 82

6 Experimentation and Analysis 83
6.1 Experimental data . . . 83
6.2 Goal and organization of the analysis . . . 84
6.3 Derivation of the road network . . . 84
6.3.1 Extraction of road polygons . . . 85
6.3.2 Administrative boundary delineation . . . 86
6.3.3 Creation of a network of road centrelines . . . 87

6.4 Road network matching 89
6.4.1 Point feature matching 89
6.4.2 Linear feature matching 92
6.5 Geometric transformation and alignment 94
6.5.1 Linear map accuracy 94
6.5.2 Improvement in positional alignment 95
6.5.3 Improvement in geometric area 98
6.5.4 Homogenization and adjustment 100

7 Conclusions **103**
7.1 Discussion 103
7.1.1 Approach and objectives 103
7.1.2 Significance of the findings 104
7.2 Outlook 104
7.2.1 Limitations and possible improvements 104
7.2.2 Implications 105

Bibliography **107**

Appendix A **115**

Appendix B **116**

Acknowledgements **117**

Curriculum Vitae **118**

List of Figures

Fig. 1.1: The “butterfly” diagram showing the significance of the cadastre 11

Fig. 2.1: The land management paradigm 17
Fig. 2.2: Hierarchy of concepts related to the cadastre 19
Fig. 2.3: Hierarchy of boundary evidence 23
Fig. 2.4: Cadastral survey options 24
Fig. 2.5: Photogrammetric mapping techniques in land parcel boundary mapping 26
Fig. 2.6: Basic classes of the Land Administration Data Model (LADM) 32

Fig. 3.1: A survey plan and a deed plan 40
Fig. 3.2: A Registry Index Map (RIM) 41
Fig. 3.3: A Demarcation Map 42
Fig. 3.4: Preliminary Index Diagram 43

Fig. 4.1: Different approaches for geometric quality improvement of land parcel boundaries contained in different records and maps 50
Fig. 4.2: Data heterogeneities associated with different levels of data modelling 56
Fig. 4.3: Generic framework for spatial data integration 58

Fig. 5.1: Process workflow for geometric quality enhancement of provisonal cadastral datasets 62
Fig. 5.2: Initial approximation of the concave hull by the convex hull 66
Fig. 5.3: Points identified for the initial concave hull 67
Fig. 5.4: Initial and final concave hulls 68
Fig. 5.5: Effect of uneven distribution of vertices on the resulting concave hull 68
Fig. 5.6: Effect of extreme polygon indentation on the resulting concave hull 69
Fig. 5.7: Reconstruction of the road junctions 70
Fig. 5.8: Straight skeletons representing road centrelines 70
Fig. 5.9: Reconstructed road junctions 71
Fig. 5.10: Matching of junctions (points) from subject and reference network datasets based on relaxation labelling 72
Fig. 5.11: Correspondences based on the differences in direction between the subject and reference incident arcs at intersections 73
Fig. 5.12: Subject (grey) and reference (black) network datasets 74
Fig. 5.13: Illustration of results of path matching 77
Fig. 5.14: Possible path matching problem 78
Fig. 5.15: Corresponding subject and reference polylines 78
Fig. 5.16: Direction change at the vertices for the subject and reference polylines 79
Fig. 5.17: Simultaneous Delaunay triangulation of subject and reference datasets 81
Fig. 5.18: Quasi-Delaunay triangulation of the reference dataset with respect to the subject dataset 81

Fig. 6.1: Experimental datasets 84
Fig. 6.2: Resulting concave hull after segmentation 85
Fig. 6.3: Road polygons obtained as a geometry difference between a concave hull and input parcel polygons 86
Fig. 6.4: Comparison of boundaries delineated using the concave hull algorithm with the reference boundary 87
Fig. 6.5: Derived road network for Case 1 88
Fig. 6.6: Derived road network for Case 2 88
Fig. 6.7: Influenced of the vertices in the input road polygons on the derived road centrelines 89
Fig. 6.8: Distribution of corresponding subject and reference points for Case 1 90
Fig. 6.9: Distribution of corresponding subject and reference points for Case 2 90

Fig. 6.10: Possible dual point correspondences 92
Fig. 6.11: Corresponding paths established through the path matching algorithm for Case 1 . 93
Fig. 6.12: Corresponding paths established through the path matching algorithm for Case 2 . 93
Fig. 6.13: Visualization of linear map accuracy of the cadastral dataset for Case 1 95
Fig. 6.14: Visualization of linear map accuracy of the cadastral dataset for Case 2 96
Fig. 6.15: cadatral dataset after local geometric transformation 96
Fig. 6.16: Qualitative improvement of alignment of the cadastral dataset for Case 2 97
Fig. 6.17: Relative positional accuracy before and after transformation 98
Fig. 6.18: Visualization of land parcel before and after geoemtric quality improvement for the first case study 99
Fig. 6.19: Visualization of land parcel before and after geometric quality improvement for the second case study 99
Fig. 6.20: Histograms showing the number of parcel groups exhibiting positive and negative improvement in geometric area 100
Fig. 6.21: Rectangularity of features is affected during piece-wise affine transformation 101

List of Tables

Table 2.1: Evolution of the applications of the cadastre 28

Table 3.1: Cadastral maps and their common scales and positional accuracies 44
Table 3.2: Cadastral coverage 45
Table 3.3: Digital topographical coverage 46

Table 5.1: Matching probabilities based on incident angles at the intersections 75
Table 5.2: Matching probabilities based on a distance threshold of 50 m 75
Table 5.3: Initial P Matrix as a Hadamard product 75
Table 5.4: Corresponding points after 3 iterations 76
Table 5.5: Correspondence of vertices between two similar curves 79

Table 6.1: Evaluation of point matching based on relaxation labelling 91
Table 6.2: Dataset similarity based on corresponding paths 94
Table 6.3: Summary statistics on identical point (IP) correspondences 95
Table 6.4: Positional accuracy improvement of linear features 98

1. Introduction

1.1 Background and motivation

1.1.1 Modern land management paradigm and the "Butterfly" diagram

The cadastre was recognized for the first time as an essential tool that enables good land management by providing reliable and usable land information at a meeting in Bogor, Indonesia (FIG, 1996). Enemark and others (2005) relied on the relationship between the cadastre and land management to identify the modern land management paradigm. The paradigm provides a theoretical framework for relating a country's historical circumstances to its land policy decisions, land administration functions (land tenure, land value, land use and land development), land information infrastructures and its sustainable development goals.

The land management paradigm depicts the cadastre as the engine of land administration systems, underpinning a country's capacity to deliver sustainable development. The "butterfly" diagram (Williamson et al., 2010a) in Figure 1.1 also describes the cadastre as the engine of land administration systems and as the means to implement the land management paradigm. The diagram presents a consolidated description of all the concepts and theories related to the cadastre through essential connections.

The diagram is a virtual butterfly, in which one wing represents the cadastral processes, and the other the outcome of using the processes to implement the land management paradigm. Once the cadastral data (cadastral or legal parcels, properties, parcels identifiers, buildings, legal roads, etc.) is integrated within the Spatial Data Infrastructure (SDI), the full multifunctional nature of a land administration system, which is essential for sustainable development, can be achieved. The "cadastral engines" in the diagram can be composed of tax-driven cadastres, title or deed tenure style cadastres or multipurpose cadastres.

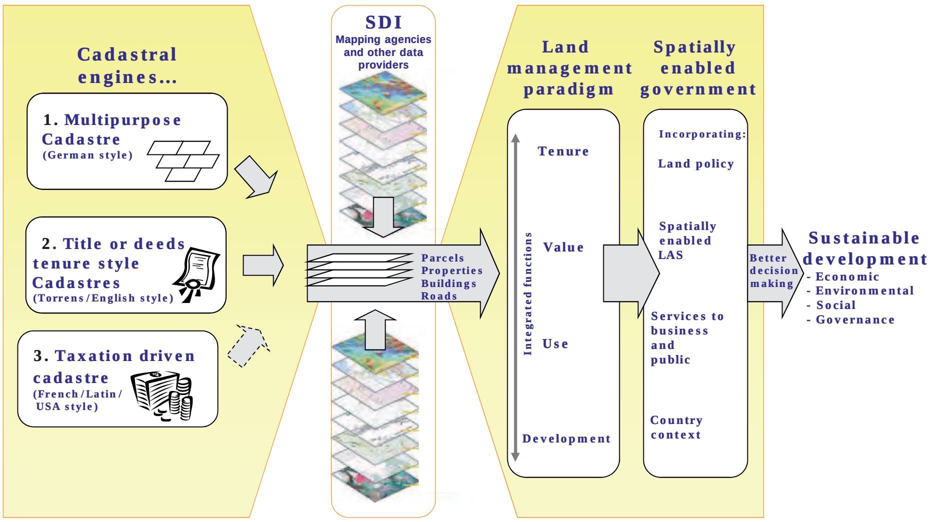


Fig. 1.1: The "butterfly" diagram showing the significance of the cadastre

1.1.2 SDI and multipurpose cadastres

The body of the "butterfly" diagram is the SDI, with the core cadastral framework acting as the connecting mechanism. Like any framework data, cadastral framework should meet the basic requirements specified in its thematic standard. The common elements of framework data specification include accuracy and content. Most data standards specify the accuracy and content limits that a dataset should meet to be included in the framework. Multipurpose cadastres therefore provide the ideal foundation for creating a jurisdiction-wide cadastral framework (or fabric).

The concept of a multipurpose cadastre which emerged in the 1970s and 1980 (NRC, 1980) has been widely promoted as the "best practice". A multipurpose cadastre consists of legal (e.g., property ownership), physical (e.g., topography,), and cultural (e.g., land use) information in a common and accurate reference framework. Multipurpose cadastres are often promoted as "best practice" because they support more applications than the traditional fiscal and legal cadastres. They support sustainable development by guaranteeing ownership and security of tenure, providing security for credit, facilitating improved land use planning, assisting in the management of environment, and as the basis for many other government functions.

1.1.3 Survey-accurate cadastres

Multipurpose cadastres that are also survey-accurate (i.e., 0.02 m - 0.03 m) are most preferable because they could be suitable for many applications. Therefore, jurisdiction-wide survey-accurate and up-to-date cadastral frameworks are aspirations of many cadastral agencies (Bennet et al., 2010).

A survey-accurate cadastral framework is built by combining land parcels in cadastral plans that are plotted to scale and kept up-to-date. In many countries especially in Europe, cadastral frameworks have been created from complete cadastral maps normally encompassing individual jurisdictions. A cadastral framework of homogeneous quality is guaranteed if the geometric scales and quality of the individual cadastral maps do not differ significantly. Otherwise, a homogeneous cadastre is not definite.

In peri-urban environments where rural and urban land uses often overlap, cadastres are usually of different geometric quality. This is because of the different accuracy needs for rural and urban applications. In Kenya, for instance, cadastres in the rural neighbourhood of almost all urban areas are captured by graphical means and are used provisionally because of their poor geometric quality. As demand for land in these environments increases due to urbanization and possible urban sprawl, better spatial planning is required, which consequently requires a homogeneous and high quality cadastral framework.

1.2 Geometric quality improvement of cadastres

One option to realise a homogeneous cadastral framework in a peri-urban environment is to replace existing poor quality cadastral datasets with new survey-accurate datasets captured through field measurements. However, survey-accurate cadastres require lengthy legal procedures and huge financial costs to create. An alternative would be, first to geometrically improve the quality of the existing poor quality cadastral datasets prior to their inclusion into the cadastral framework. Quality improvement can be achieved by exploiting the widespread availability of more accurate topographical datasets through data conflation techniques.

Once land parcel-based cadastres have been created, they are usually updated on a transactional basis. It is rare to find multiple cadastral databases available for a specific area like topographical databases, unless they have been captured with the intention of replacing the existing databases. This is because legal provisions usually regulate the creation of cadastral databases unlike topographical databases.

Different methods can be used to improve the geometric quality of existing cadastral datasets. The methods depend on the category and nature of cadastral data involved. Cadastres may be classified in many ways (FIG, 1995), e.g. by: primary function (e.g. supporting taxation, conveyance, land distribution, or multipurpose land management activities); the types of rights recorded (e.g. private ownership, use rights, mineral leases); location and jurisdiction (e.g. urban and rural Cadastres; centralized and decentralized cadastres); and ways in which information about the parcels is collected (e.g. ground surveys tied to geodetic control, uncoordinated ground surveys and measurements, aerial photography, digitising existing historical records, etc.). Cadastres can also be classified either as numerical (computational) or graphical (i.e., no coordinates exist for the parcel boundaries) depending on the approach used to measure parcel boundaries (Williamson, 1985).

Numeric cadastres would need to be improved because of existing planar geometric differences with reference to more accurate field measurements obtained using more accurate measurement techniques such as differential GPS (Global Positioning System) (Hope et al. 2008). Accuracy improvement of numeric cadastres might also be necessary following a policy to change from an old coordinate system to a new one, say GPS-based coordinate system (Jarroush and Even-Tzur, 2006; Felus, 2007). The common approach would entail adjusting the coordinates of the objects in the cadastre by using accurate measurements captured directly in the field.

The improvement of the geometric quality of graphical cadastres is necessary when there is a policy that requires an implementation of legal coordinate-based cadastres from graphical cadastral maps. The vectorization of a numeric cadastral map, which is graphically represented results in a graphical cadastre. This is because the boundary coordinates, which can be extracted from the cadastre, would not be of the same accuracy as their counterpart coordinates computed using the survey-accurate measurements made in the field. Thus, accuracy enhancement can be achieved by processing existing survey-accurate data contained in field books. Morgenstern (1989) described a procedure to remove geometric discrepancies from a graphical cadastre, which maintains the positional conditions contained in the graphical structure of such maps.

If no field measurements exist at all, new survey-accurate field measurements can be used either to replace or to improve the geometric accuracy of graphical cadastres. Since carrying out new survey-accurate field measurements would be a non-realistic solution for low-income countries (Arvanitis and Koukopouluo, 1999), a data conflation approach, which might be more feasible, would be required.

The creation of a survey-accurate and up-to-date cadastral database will inevitably involve reference to legacy cadastral databases if they already exist. Using aerial orthophotos and rectified satellite images to improve the geometric quality of existing poor quality graphical cadastres seem to be an excellent option and an accuracy ranging from 0.3 to 0.5 m is possible (Boljen, 2009). However, this can be considered as an intermediate solution because a survey-accuracy is not guaranteed. This is partly because of the limited spatial resolution of orthophotos and partly because of the possible presence of vegetation in the imagery that makes it difficult to identify land parcel boundaries. Moreover, establishing the geometric relation between the legacy cadastre and the orthophotos would involve the digitization and identification of similar features in the orthophoto and the legacy cadastral database. This is because the orthophoto exists at an iconic level (i.e., raster form) while the cadastral dataset is already in a symbolic level (vector form). Therefore, using topographical databases in the context of data conflation would be a better solution, because both datasets are in symbolic level and therefore easier to automatically identify similar features like roads and building footprints in the datasets.

Although the geospatial research community has addressed the integration of data from multiple sources long enough, some fundamental challenges remain, in particular, when administrative and parcel (cadastral) boundary datasets are involved. This is because administrative and land cadastral boundaries, which have much in common but in a broader sense, have a legal basis, are created from data that is acquired over a long period and sometimes in a sporadic manner. Besides, cadastres, which often represent land parcel boundaries on maps, do not necessarily have corresponding ground signatures as topographical datasets do. The proposed solution to these challenges exploits the topological relationship that exists between parcel-based cadastres and road network datasets. The approach takes advantage of the widespread availability of more accurate topographical datasets to improve the geometric quality of poor quality graphical cadastres through data conflation techniques.

1.3 Overview of the proposed approach

If topographical datasets of higher geometric accuracy than the provisional cadastral datasets are available, corresponding features between the topographical and the cadastral datasets can be established through feature matching techniques. Similar features in both datasets could include, for example, road centrelines and building footprints.

Sometimes however, cadastral datasets unlike topographical datasets do not necessarily have corresponding ground signatures. Moreover, most cadastres contain only the legal land parcels because of the traditional focus to register and map only land parcels. Thus, considering such cadastres which contain only land parcel boundaries, there is an apparent lack of corresponding features between the topographical dataset and the cadastral dataset. The apparent absence of corresponding features is established through the road network implicitly represented in cadastral dataset and the one from the topographical dataset. The implicit road network in the cadastral dataset is first derived from cadastral dataset, which together with the road network from the topographical dataset forms the basis for feature matching.

Coordinates of points are obtained from corresponding junctions, terminations and edges of the road networks during the matching process. These are used to establish a local geometric transformation between the cadastral and the topographical datasets. A local geometric transformation is used because provisional cadastral datasets are considered to contain local geometric distortions. The quality of the cadastral dataset is expected to have improved after carrying out local geometric transformation.

1.4 Objectives and focus

Considering that cadastres differ from one country to another, it becomes necessary to identify and describe the cadastre of a given country for the purpose of avoiding possible misconceptions about the cadastral concept. Moreover, the description is important for identifying possible challenges that might be encountered during the creation of a seamless countywide cadastre. Thus, the first main objective of this dissertation is to review the cadastral and other related concepts, and to outline and evaluate as a case study, the Kenyan cadastre.

The second main objective of this dissertation is to develop an approach for geometric quality improvement of the provisional graphical cadastres through data conflation techniques. The specific objectives include:

- Establish correspondence between provisional cadastral dataset and the topographical datasets considering the apparent lack of corresponding features;

- Establish feature matching and geometric transformation approach that considers local geometric distortions in the provisional cadastral dataset;

- Assess the extent of quality improvement possible with the data conflation approach.

The improvement of the geometric quality of poor quality graphical cadastres before inclusion in a cadastral framework will ensure that the jurisdiction-wide cadastre is homogeneous. This is necessary for a SDI dedicated to land administration. This thesis focuses on the two-dimensional geometric accuracy improvement of poor geometric quality graphical cadastres prior to their inclusion into a cadastral framework.

The overall aim of this thesis is to contribute to the research on geometric quality improvement of poor quality graphical cadastres through spatial data conflation techniques. The rationale of the work is to exploit already existing more accurate topographical datasets to improve the quality of poor quality graphical cadastres without the need for field surveys.

1.5 Outline of the thesis

This dissertation consists of two logical parts with regard to the two main objectives. The first part consisting of the second and the third chapters, deal with concepts related to land management and the cadastre. In particular, the second chapter presents a review of the concepts related to cadastre. The chapter focuses on the evolution and the recent efforts in cadastral and land administration domain data modelling. An outline of the design elements that will characterize future cadastres concludes the chapter. The third chapter describes the Kenyan cadastre with the aim of identifying some of its shortcomings. The Kenyan cadastre is then evaluated using the design elements of future cadastres outlined in the second chapter.

The second logical part deals with the second main objective of this dissertation and it consists of the fourth, fifth and the sixth chapters. The fourth chapter presents an overview of the state of the art in spatial data integration. This is followed by the identification of challenges that are still outstanding in data integration when cadastral and administrative boundary datasets are involved. With reference to the theme of this thesis, different possible solutions for the quality improvement involving different types of cadastral data are presented. The chapter concludes with a specific focus on data conflation, when provisional graphical datasets are involved. The fifth chapter describes the proposed data conflation approach for geometric quality improvement of poor quality graphical cadastres. The approach is developed in the context of conflating topographical datasets with graphical cadastral datasets.

The sixth chapter contains a description of a systematic experimentation, analysis of the results obtained using the data conflation approach with datasets from the Kenyan cadastre. Conclusions drawn from this study and an outlook for further work are outlined in the seventh, the last chapter.

2. Land Management and the Cadastral Concept

2.1 The modern land management paradigm

The modern land management paradigm (Enemark et al., 2005), represents the latest effort to provide a better understanding of the concept of cadastre within the large context of land management. The paradigm is considered as the cornerstone of modern land administration theory and it is pervasive in modern societies. It involves a holistic treatment of the functions of land administration (land tenure, value, use and development).

The paradigm emerged first, because of the need to address land management issues in systematic manner. Secondly, because of the need to have land administration systems designed as an enabling land information infrastructure for implementing land policies and land management strategies to support sustainable development.

In theoretical terms, the paradigm identifies the principles and processes that define land management. According to Figure 2.1 (ibid.), which depicts the paradigm, land management activities within a country may be described by three components: land policies, land information infrastructure and land administration functions that support sustainable development.

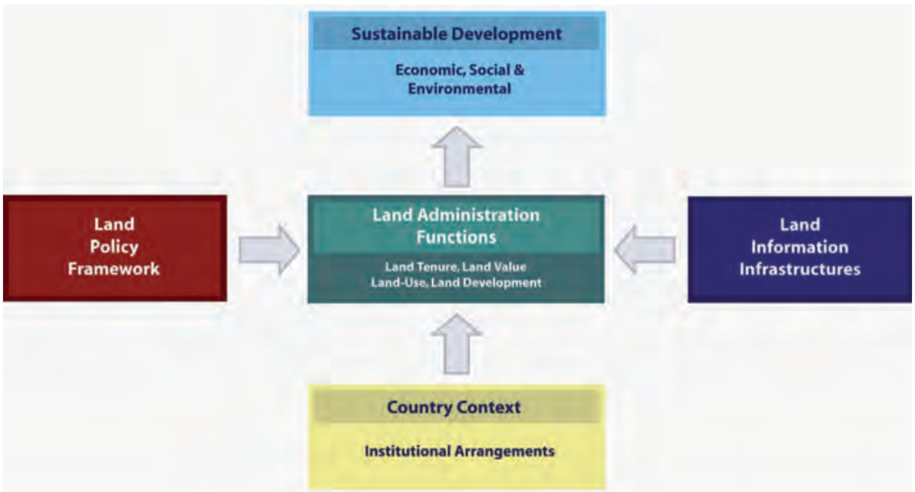


Fig. 2.1: In the land management paradigm, land management is described by three components: land policy, land information infrastructure, and land administration functions in support of sustainable development

Within a given country context, land management covers all activities associated with the management of land and natural resources that are required to fulfil political objectives and achieve sustainable development. It is broader than land administration. Land administration depends on land policies and the land information infrastructure to achieve sustainable economic, social and environmental development.

2.2 The cadastral concept

Though the land management paradigm is neutral irrespective of a country's land administration practice, the cadastre is the core engine of land administration. The "butterfly" diagram in Figure 1.1 described the significance of the usefulness of the cadastre in land management. The diagram depicted the cadastre as a key component within the Spatial Data Infrastructure (SDI) as it supports each of the four functions of land administration. Cadastre is however, a concept that takes different meanings and implemented differently in different jurisdictions depending on their cultures, colonial histories and practice of land tenure.

Although many definitions of the term cadastre abound, it is almost impossible to give a terse and comprehensive definition that captures the essence and practice of the concept in different jurisdictions. The concept takes on different meanings in every country (Dale, 2006) because of each country's historical development, its laws and customs, and largely its form of conveyance and methods of by which land registration was introduced.

It is generally understood that the word "cadastre" arose from a Latin word designating territorial taxation of the Roman province. In this case it is a land recording system used as a mechanism for collecting real estate tax and is referred to as a fiscal cadastre since it relates to the state treasury (Holl et al., 2010). Over the years, cadastres have been introduced to keep track of the technical attributes of territorial land inventories. This is in addition to keeping track of the legal Rights, Responsibilities and Restrictions (RRR) of both private and public land. Thus, alternative interpretations of the concept include land registration that concerns only the system of registration of title (English influence); land registration, which also covers the cadastral system; and where the cadastre includes land registration (e.g., "legal cadastre" or "multipurpose cadastre") (Henssen and Williamson, 1990).

The International Federation of Surveyors (FIG, 1995) defined a cadastre as:

"a parcel based and up-to-date land information system containing a record of interests in land (e.g. rights, restrictions and responsibilities). It usually includes a geometric description of land parcels linked to other records describing the nature of the interests, ownership or control of those interests, and often the value of the parcel and its improvements. It may be established for fiscal purposes (valuation and taxation), legal purposes (conveyance), to assist in the management of land and land-use control (planning and administration), and enables sustainable development and environmental protection".

The boundaries of the properties and the parcels identifier are normally shown on large-scale maps. Besides the maps, there exists also a descriptive part of the cadastre, a sort of register, which contains physical attributes of the parcel, i.e., identifier, location, area, kind of use and abstract attributes like data for land tax such as land value (ibid). The parcel register, which is part of the cadastre is different from the land register. The land register is a file that records legal facts (deeds) or legal consequences (title) and other physical or abstract attributes concerning the parcels depicted on the cadastre. The cadastre and the land register are administered by either the same agency or different agencies depending on the prevailing public administration arrangements.

2.2.1 Hierarchy of concepts related to cadastre

A hierarchy of concepts related to the cadastre and other land related issues is illustrated in Figure 2.2 as adopted from Williamson and others (2010a). The land policy and the land parcel are at the top and bottom of the hierarchy respectively. Land policy is considered an abstraction of reality. It consists of a group of coherent decisions taken by authorities to achieve the objectives related to land. Land policies do not render themselves to effective implementation; instead, they are implemented through a set of laws and administration regulations. As such, land policy forms the foundation for land law and land management.

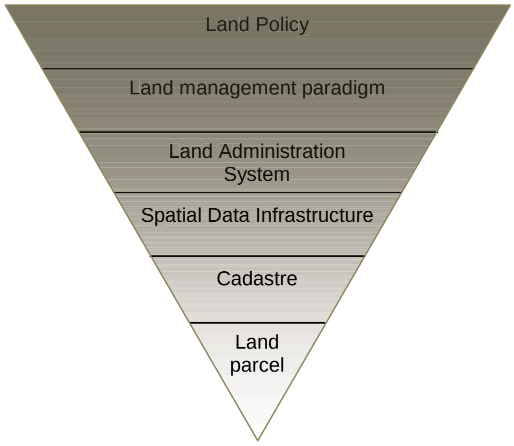


Fig. 2.2: Hierarchy of concepts related to the cadastre

Land management involves the establishment of goals and mechanisms that influence land use in order to achieve desired policy objectives. In modern literature (ibid.), the modern paradigm in land management is characterized by a holistic approach by land administration systems to force their land administration processes to contribute to sustainable development. Traditionally, land management was directed individually to either social, environmental or economic objectives.

The operational component of the land management paradigm includes the range of land administration functions carried out to ensure proper management of rights, restrictions and responsibilities in relation to property, land and natural resources. Land administration is considered to include processes of determining, recording and disseminating information about tenure, value, use and development when implementing land management policies (FIG, 1999).

Spatial Data Infrastructure (SDI) plays a significant role in facilitating a country's land information infrastructure. It ensures a uniform approach for maximum integration and security of data, in particular, cadastral data; effective resource use and development of comprehensive and interoperable land information systems. Cadastre is the core of information in a land information infrastructure. It provides the spatial integrity and unique identification of every land parcel through cadastral maps.

Land parcel is the foundation of the hierarchy and it reflects the way people use land. It is the key object for identification of land rights and administration of restrictions and responsibilities in the use of land. It is increasingly becoming common to find cadastres that contain other land objects in addition to the land parcels.

2.2.2 Land registration and land recording

The possession and ownership of land (or property) is more difficult to prove compared to other physical objects like cars. This is because of the possible existence of different rights over the land, in which the rights can be separated and held by different parties. Title is the legal term commonly used to refer collectively to the bundle of rights in a piece of property in which a party may own either a legal interest (actual ownership) or equitable interest (actual enjoyment). Title in land must be transferred when the land is sold and must be cleared (i.e., free of encumbrances) in order for the transfer to take place. This is only possible if there is some form of land registration or recording in existence. The cadastre or cadastral system provides the basis for establishing property ownership by creating a connection between land and individual persons (van Oosterom et al., 2006).

Depending on the legal traditions, colonial histories and the nature and extent to which the government is involved in land conveyance, two systems exist of land registration: the deeds and the title systems. The key difference is found in whether only the transaction is recorded (the deeds system) or the title itself is recorded and secured (the title system). The deeds system is a register of owners focusing on "who owns what" while the title system is a register of properties presenting "what is owned by whom".

In the deeds registration system, the deed itself being a document, which describes an isolated transaction, is registered. This deed is evidence that a particular transaction (deed) took place, but in principle it is not in itself proof of the legal rights of the involved parties and, consequently, it is not evidence of its legality. Thus, before any dealing can be safely effectuated, the ostensible owner must trace his ownership back to a good root of title (right). Deed registration is usually applied in countries, which are based on the Roman law (in Europe: France, Spain, Italy, Belgium, and the Netherlands). The system can also be found in countries that were influenced by the former ones in earlier times (South-America, parts of North America, and some African and Asian countries) (Williamson et al., 2010a).

In title registration, the legal consequence of that transaction i.e. the right itself (= title) is registered. Therefore, the right together with the name of the rightful claimant and the object of that right with its restrictions and charges are registered. With this registration, the title (or right) is created.

Three principles are upheld in title registration namely: the mirror principle, the curtain principle and insurance or guarantee principle (Henssen and Williamson, 1990). The mirror principle means that the register is supposed to reflect the correct legal situation. The curtain principle means that no further (historical) investigation or search beyond the register is necessary. The insurance or guarantee principle means that what is registered is true for third parties in good faith and that bona fide rightful claimant who is contradicted by the register is reimbursed from an insurance fund of the state.

Based on the two systems of land registration, three basic approaches or styles are practiced worldwide: the French (or Latin) approach; the German approach and English with alternative Torrens approach are identified and are described in (Williamson et al., 2010a).

The French approach is based on deeds system. In this system, the transaction is registered, no titles are guaranteed and the interests in the deed is described using metes and bound or sometimes with a sketch map, which is not necessarily the same as the cadastre. Cadastral surveying and mapping is normally a follow-up process (if at all) after land registration. Deeds systems can be converted to title registration systems.

The German style is a title registration system in which a land book (*Grundbuch*) is maintained at local district courts. The titles are guaranteed by the state and are based on parcel identification. However, neither the boundaries nor the areas are guaranteed (ibid.). Cadastral surveying and mapping is based on fixed boundaries and is normally carried out prior to land registration.

In the English and the Torrens approach, title registration is used. The land records are maintained at the land registration office, the registered titles are guaranteed by the state, and however neither the boundaries nor the areas are guaranteed. Cadastral registration is normally integrated in the land registration process. Whereas fixed boundaries are used in the Torrens system, general boundaries identified in large-scale maps are used in the English system.

Irrespective of the differences in the system of land registration, four basic principles are upheld: the booking principle, the consent principle, the principle of publicity, and the principle of speciality (Henssen and Williamson, 1990). The book principle implies that a change in real rights on an immovable property especially by transfer is not legally effected until the change or the expected right is booked or registered in the land register. The consent principle means that the real entitled person who is booked as such in the register must give consent for a change of the inscription in the land register. The principle of publicity implies that legal registers are open for public inspection. Finally, the speciality principle implies that in land registration, and consequently in the documents submitted for registration, the proprietor(s) and the real property (land parcel and buildings) must be unambiguously identified.

Land registration and the cadastre usually complement each other. Land registration puts emphasis on the relation between the right and the claimant (owner), the cadastre puts emphasis on the relation between the right and the object. Because land registration and the cadastre complement each other, the terms "land recording" or "land records" are used to refer collectively to these two components together.

2.2.3 Categories of cadastres

Cadastres can be classified into two categories depending on the method of survey used to determine the land parcel boundaries and the achievable positional accuracy. They include numerical (or computational) and graphical cadastres (Williamson, 1985). A numerical cadastre refers to a cadastre where the bearings and distances of the individual land parcels are determined from computations based on actual land survey. With reference to monumented landmarks, the dimensions of individual parcels are measured and referenced to either local or national coordinate systems. Maps may represent numerical cadastres in addition to the computed coordinates for the parcel boundaries graphically. Depending on the survey regulations, numerical cadastres are generally considered to be of the best positional quality. In addition, they are considerably more expensive to produce than the graphical cadastres.

Graphical cadastres are cadastres created by scaling bearings and distances either from photomaps or by other graphical methods. Cadastres are considered graphical because they are produced by graphical survey techniques, for instance, the traditional plane table surveying and aerial photogrammetry. This description comes from the fact that the survey of the boundaries does not involve any geometric measurement of coordinates, instead the boundaries are plotted directly in the field or from aerial photographs. In some developing countries, e.g., Nepal and Bhutan, plane tabling, a graphical surveying technique is still used to prepare cadastral maps in the field (Tuladhar, 1996).

2.3 Cadastral boundaries

2.3.1 Land parcel boundaries and their documentation

Land parcel boundary is one of the key elements of a cadastre. A land parcel boundary, which is alternatively referred to as cadastral boundary, defines the spatial extent where homogeneous rights, responsibilities and restrictions exist. Land parcel boundaries are required to be unambiguously defined both on the ground and on the map. The legality and nature of the parcel boundaries depend on whether they have been registered and whether they are accurately defined. If the boundaries are only explicitly marked on the ground using walls, fences and hedges, they are considered general.

Otherwise, the boundaries are fixed and legal if they are marked on the ground by permanent monuments, which are surveyed by ground methods with corresponding measurements displayed on the map. Parcel boundaries can also be implicitly defined, for example, the right of way.

Because boundaries are legal entities, their documentation and official registration is important. The description and documentation of boundaries can be done in various ways with varying degree of technical complexity and financial costs. The approaches include textual description, graphical representation or representation by coordinates.

Textual description (also called metes and bounds) uses landmarks as points of reference to describe the route for a land parcel boundary. The description entails identifying the landmarks along the boundary. The landmarks must be salient features that can be easily recognised even after several years. These landmarks together with the metes, i.e., the measurement of straight distance and direction along the boundary constitute the description of the boundary. The description also referred to as bounds, entails boundary description with reference to some typical landmarks such as buildings, roads and road intersections, rivers and sometimes even prominent trees. The main advantage of textual description of boundaries is that even lay people can create and understand them and that it can be easily used in courts as it can be treated like any other text document (Navratil, 2011). A possible problem with metes and bounds is if there is a change in topology of the referenced landmarks.

Graphical representation of boundaries is achieved by using either a scaled image of each parcel or a map showing all parcels within a specific area. The graphical representation can be used as a basis for mapping the contents of the parcel. Graphical representation usually simplifies land use planning for the landowner. This is because graphical representation provides a starting point for the planning procedure, and particularly if the graphical representations of all parcels are collected and integrated in a set of maps. It has to be ensured that no gaps or overlaps exist between such maps. Graphical representations are typically interpreted locally and scale representations stipulate its accuracy and precision. For example, a mapping scale of 1:1,000 results in the definition of a positional accuracy of 10 cm at best (ibid). Since graphical representation only guarantees that the boundary between two adjacent parcels is coincident, to find the actual boundary may provide a serious challenge.

Cadastral Boundaries can also be described by use of coordinates. This method entails a mathematical description of the boundaries i.e., neighbouring points of the boundary connected by straight-line segments and the resulting figure constitutes the boundary of the parcel. The mathematical description ensures a high positional accuracy of boundary definition and therefore the required coordinate accuracy can be stipulated precisely.

2.3.2 Hierarchy of boundary evidence

In case of a boundary dispute, existing boundary documentation can be used in the resolution of the dispute. However, not every type of document will necessarily be relevant to every case. Generally, the hierarchy of boundary evidence developed from Swensson (2010) and Srebo and Shoshani (2007) is illustrated in Figure 2.3, although country specific precedence and preferences exist. According to the hierarchy, the evidence of the highest in the resolution of boundary disputes entails a precise and detailed documentation. Existing boundary pillars come next in priority. In the absence of the two, dispute resolution can be based on cadastral plans. Finally, graphical maps that do not necessarily contain boundary measurements can be used when neither of first three documents exist. In the absence of all else, the resolution of boundary disputes could be based on possession and other circumstances, for example, verbal evidence.

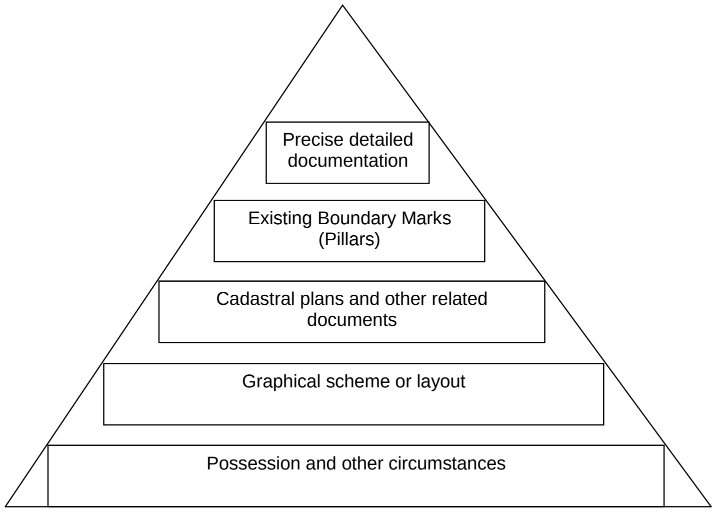


Fig. 2.3: Hierarchy of boundary evidence

In Kenya, for instance, the state guarantees the title but not the boundary and areas. Section 22 of the Registered Land Act (GoK, 1963) stipulates that it is necessary to note in the land register that the boundaries of a parcel have been fixed. Otherwise, the registry map and any field plans are considered to indicate the approximate boundaries and the approximate situation only of the parcel. In this situation, courts should not entertain any action or other proceedings relating to a dispute as to the boundaries of registered land unless the boundaries have been fixed. Thus, plans and maps based on general boundary surveys are considered to have less priority in the hierarchy of boundary evidence than those based on fixed boundaries.

A fixed-boundary is an invisible line defined geometrically through an accurate survey. The boundary lines are generally more accurate and legal. On the other hand, a general (or approximate) boundary is one in which the precise line is undetermined in relation to the physical features which demarcates it. That is, it is not settled whether the boundary runs along the centre of a fence, wall. However, it is clear on the ground where the parcel is situated and where the boundaries are.

Because land adjudication has been the main process through which land has been registered in Kenya, land parcel boundaries are left vague and therefore based on general boundaries. The Land Registrar indicates on a field plan or otherwise in the register, the precise position of the boundaries of a parcel, or any parts thereof or a group of boundaries. This is usually initiated either by the Registrar's discretion or through the application by any of the interested parties.

2.4 Homogeneous cadastral framework

For better land management within any jurisdiction, the cadastre should be complete and of consistent quality. In many jurisdictions, the cadastre consists of cadastral maps of different positional qualities usually captured through different techniques. Digital technology offers the possibility to integrate these maps to create a jurisdiction-wide integrated cadastral framework.

A cadastral framework or cadastral fabric (Konecny et al., 2010) is a single, continuous, integrated and consistent digital map showing the location, extent and nature of all (public, private and community) land parcels. The framework represents the complete survey record within a jurisdiction. The framework, usually created from both existing survey records and new field measurements represents the current ground state in continuous and non-overlapping surface land parcels, which are defined by different surveys. The accuracy of the cadastral fabric is directly related to the precision of survey of the input data. Thus, a cadastral framework with regions of different qualities is possible.

Different options to carry out measurements of land parcels are available. Figure 2.4 illustrates a modified topology of the options as found in Konecny and others (ibid.). The options are given with regard to the method and techniques of measurement, the quality of the measurements of land parcels and the type of cadastre.

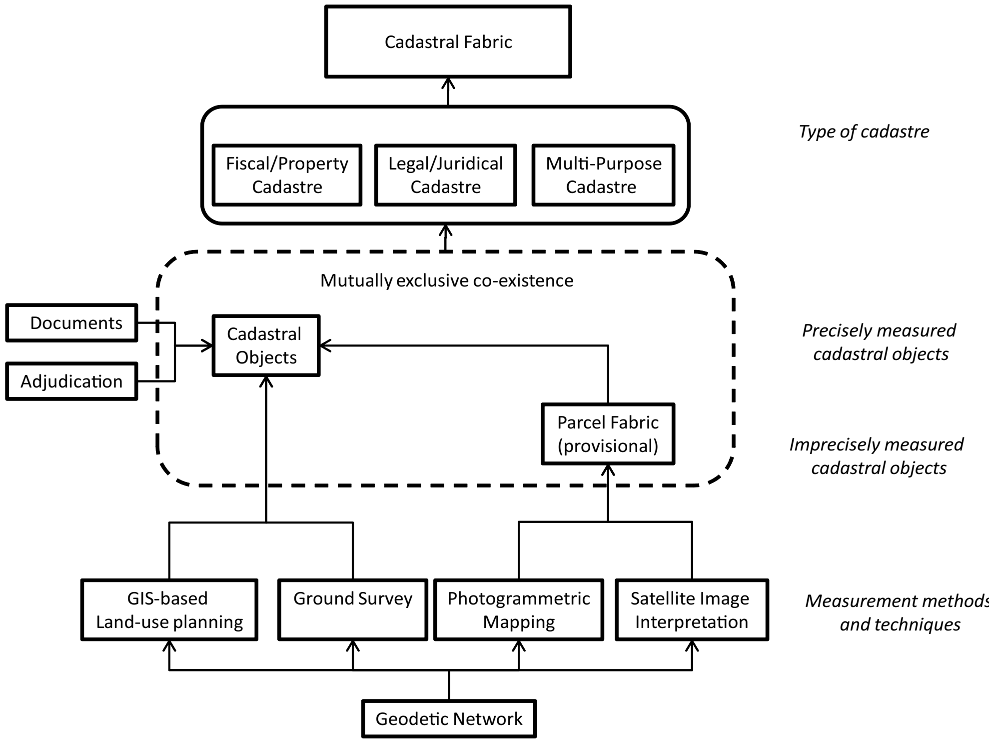


Fig. 2.4: Cadastral survey options

Cadastral framework is at the top indicating that it can be constituted from either one or different types of cadastres used within the same jurisdiction. Individual cadastres could consist of approximately or precisely measured cadastral objects (land parcels and buildings). Parts of the parcel fabric that are approximate can be improved either on a transactional basis by more accurate surveys using modern technology or systematically on a region-by-region basis or through data conflation approaches suggested in this thesis.

In continental Europe, for instance, where an accurate geodetic network exists, and where the measurement of cadastral objects is based on ground surveys, the cadastre is updated on a transactional basis by more accurate field measurements. The situation is different in the USA, where digital mapping based on an accurate geodetic network is first carried out. Then the parcel fabric is generated from the digital topographical map and verified using existing documents, and if necessary, by adjudication to create a property cadastre. The process is still going on (ibid).

2.4.1 Cadastral surveying

Two-dimensional cadastres (i.e., land parcel boundaries) are established by using various surveying and mapping techniques. There are two basic methods, which may be adopted: ground surveying (together with GIS based land use planning) and photogrammetric mapping (together with the associated image interpretation) (Figure 2.4).

Ground survey

Ground survey techniques are the basis for most cadastral survey systems. They suit the sporadic and isolated survey approach that is usually a component of land registration, conveyance or alienation systems based mainly on the "user pays" principle (Williamson, 1983).

Numerical cadastres, which rely on fixed boundaries, are created using ground survey techniques. There are many countries where ground survey methods such as compass or optical square and tape measurements, compass and angle instrument, theodolite and stave, and electronic distance measurement and theodolite are employed. Nowadays, the GPS is also a popular data capture instrument for cadastral surveying.

Photogrammetric mapping

A much more rapid and cost effective procedure than the ground surveys is to use aerial photographs or satellite images together with ground survey techniques. Photogrammetric techniques are the main techniques that utilise aerial images and satellite images. They are only suitable for systematic adjudication and survey of large-scale projects. All cadastral surveys, which utilise photogrammetric techniques, require at least some field completion. Photogrammetric approach can only be used where the physical boundary or some point indicating the boundary is evident on the aerial photograph. There is no guarantee however that the images permit the actual legal boundaries unless they are legally established (Konecny et al., 2010).

Photogrammetric techniques are not only used for cadastral mapping, but also for updating of the existing cadastral maps. This is due to the development of automated photogrammetric techniques and the increase in spatial resolution of satellite imagery. For example, projects are reported that have been carried out to determine and exploit that potential (Vassilopoulou, 2002).

If photogrammetric techniques are to be used for parcel boundary mapping, the technical accuracy requirements of the final map should be considered. Prior to boundary mapping, it is important to eliminate the geometric displacements and distortions from the images. Based on achievable accuracy and manner of use, five techniques for photogrammetric mapping are identified by (Dale, 1979) and are illustrated in Figure 2.5.

The first (1) technique entails extraction of precise ground coordinates for points on the air photographs using digital or analytical stereo plotters or comparators. The precise ground coordinates of boundary corners can be calculated or measured from the point coordinates identified on the photographs. This approach was used in Switzerland (Weissman, 1971). It involved the signalization of ground marks by painting the tops of 140 mm by 140 mm boundary corner stones, followed by the determination of their grid coordinates by transforming stereo model coordinates instead of scaling them from the maps. The results of course depend on the type of camera used, the topography and the scale of photography. For a camera with a focal length of 170 mm and scale of photography of 1:7,000, a mean point error of 0.05 m was obtained. This is comparable to those obtained by ground-based methods. The main concern for this approach however is the lack of existing traverse stations to occupy when it is necessary to carry out a revision necessitated by land transactions.

The second (2) technique entails first, the preparation of base maps by photogrammetric stereo-compilation. The base maps show physical features, which coincide with the legal boundaries. The maps are used to locate points of detail that can be used as control points for simple ground surveys. The cadastral maps are then compiled from a combination of photogrammetric plotting of boundary points and lines that are visible from the air with simple graphical methods of survey to locate specific land parcel boundaries. Similar to the first method, the results depend on the type

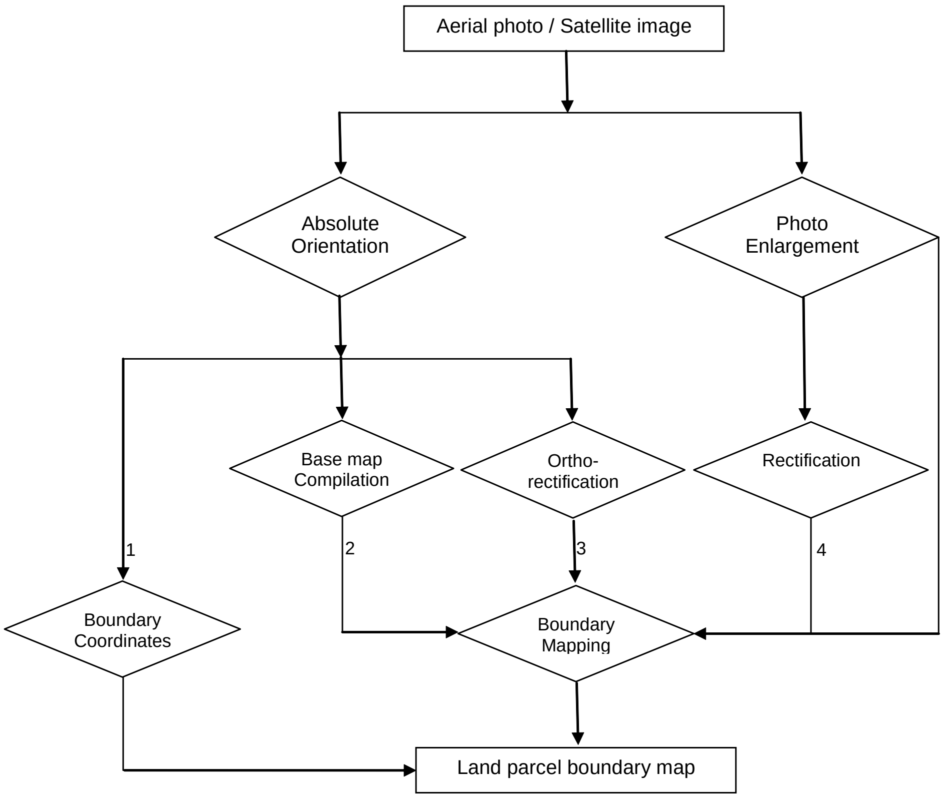


Fig. 2.5: Photogrammetric mapping techniques in land parcel boundary mapping

of camera used, the topography and the scale of photography. This method has been used in the United States of America (Karns, 1981), Kenya and Uganda (Dale, 1979).

In the third (3) technique, instead of compiling the boundary lines and points using stereo-photogrammetry, orthorectified (i.e., differentially rectified) photographs are used. Orthorectified photographs are photographs that have been corrected for tilts and displacement due to relief, are first produced and then the parcel boundaries are identified and scaled off monoscopically from the orthophotos, and therefore sufficient for cadastral purposes (Konecny, 2008). This approach has been used successfully for land registration in Palestine (Mikkonen and Corker, 2000), Australia and Canada (Dale, 1979) and is currently being considered for the renewal of cadastral maps in a number of countries, such as Jordan (Al-Ruzouq and Dimitrova, 2006).

The fourth (4) technique which has been used in Thailand (UN, 1990), Botswana, England and Wales (Dale, 1979) involves the use of simply enlarged but rectified photographs to identify land parcel boundaries. The enlargements are reproduced in the form of photomaps. These are then used as a backdrop for graphical compilation of parcel boundary outlines. This approach however is only appropriate for relatively flat terrain for reasonably accurate measurements of distance and area.

The fifth (5) technique entails the identification and plotting of land parcel boundaries using simply enlarged and un-rectified photographs and has been used in Kenya to prepare the so called Preliminary Index Diagrams (PIDs), which are only approximate (Mwenda, 2001). This approach results in the lowest positional accuracy, because no due care has been expended to correct for effects due to tilt and displacement due to relief. The positional accuracy of the resultant cadastral maps obtained using this technique is quite variable but can be estimated if photographic information and knowledge about the terrain is available.

2.4.2 Cadastral mapping and geometric accuracies

The common scales for cadastral maps range from 1:500 to 1:10,000. The large-scale maps commonly used for urban areas show more precise parcel dimensions. They are often prepared based on precise ground surveys or aerial photography. Smaller-scale maps, which are commonly used for rural areas contain neither the measured nor the dimensions of the parcel boundaries. They serve only as a graphical presentation of the parcel layout. The positional accuracy of graphical cadastral maps (i.e., the coordinates of the boundary points only represent the boundary in the graphic map but not on the ground), is influenced by the cartographic scale used. For numerical cadastral maps, the accuracy depends on the accuracy of the measurement made on the ground before mapping. It is therefore important that the choice of a given mapping scale be guided by accuracy specifications.

For traditional reasons, accuracy specifications for cadastral surveys are expressed in terms of precision ratios by establishing limits of traverse linear or angular misclosure, for example, a linear misclosure of 1: 20,000 would mean an accuracy of at least 1 millimetre per 20 metres. Error ellipses (or root mean square) could also be used for accuracy specifications. When using the error ellipses, every point position as determined on the ground from any of the control points should be within a given tolerance.

In different countries, the standards of accuracy documented in manuals for cadastral surveys specify different orders of survey accuracy for different "classes of surveys" or zones (for example, say Zone 1: 0.03 m and Zone 2: 0.07 m). The rationale of the accuracy specification is to ensure that surveys meet the standard of accuracy. The class of survey addresses issues such as the difficulty of terrain and the value of the land being surveyed and this influences the expected standard of accuracy.

For published topographical maps, the common accuracy standard is the ASPRS "Accuracy Standards for Large-Scale Maps" (ASPRS, 1990). The standard provides accuracy tolerances for maps at 1:20,000-scale or larger "prepared for special purposes or engineering applications". The standard is linearly dependent on the target horizontal scale or target contour interval. Thus, it is only applicable to mapped features that are compiled using a consistent type of data acquisition process (e.g., photogrammetry) where all spatial objects receive approximately the same accuracy. RMSE is the statistic used by the ASPRS standards.

A major feature of these ASPRS standards is that they indicate accuracy at ground scale. And the accuracy is reported as Class 1, Class 2, or Class 3. Class 2 and 3 accuracy applies to maps compiled within limiting RMSE's twice and thrice those allowed for Class 1 maps respectively. For Class 1, the planimetric accuracy or the limiting RMSE (meters) for a map at scale of 1:50, 1:1000, 1:5:000 and 1:20,000 are respectively 0.0125 m, 0.25 m, 1.25 m and 5.00 m.

For vertical accuracy, ASPRS Accuracy Standards, specifies the limiting RMSE for Class 1 maps as one third of the contour interval. Accordingly, spot elevations shall be shown on the map with a limiting RMSE of one-sixth the contour interval or less.

The consideration of aerial photography for particular mapping applications depends on the scale of photography. Traditionally, the positional and elevation accuracies achievable using photogrammetric methods and the scale of the final map are determined based on mapping requirements and practice. The accuracy and photo-to-map scale relationships mainly depend on the scale and resolution of aerial photography, the flying height, the base-height ratio and the accuracy of the stereo plotting. The choice of the photographic scales based on the mapping scale and contour interval as practiced (as a rough yardstick) by the British air survey companies is given in (Petrie, 1990).

Although scale is a concept, which relates to the level of generalization of digital geographic data, it is used mainly with regard to analogue presentation of spatial information. In digital representations, precision or resolution is used instead. Therefore, when capturing digital data from a scanned aerial photograph or satellite imagery, the desired ground resolution (geometric accuracy) of the details, which is independent of image and map scale, should be considered.

2.5 Evolution of the cadastral concept

Cadastre as a concept has changed in its role, scope, implementation and nature over time. This section explores these changes and ends with an outline of the characteristics of future cadastres.

2.5.1 Evolution of applications

The human kind to land relationship is dynamic and is changing over time as a response to general trends in societal development. Consequently, the concept and role of the cadastral systems has been changing. Although the cadastral concept differs from one jurisdiction to the other, its role is however acknowledged to have evolved in four chronological stages as shown in Table 2.1. Chronologically, the cadastre has been used as fiscal tool, legal tool, planning and land management tool, and as a tool for multiple applications (Williamson et al., 2010a).

Table 2.1: Evolution of cadastral applications (developed from Ting and Williamson, 1999)

	Feudalism ← 1800	**Industrial Revolution** 1800 - 1950	**Post-war Reconstruction** 1950 - 1980	**Information Revolution** 1980 →
Human to land relationship	Land as wealth	Land as a commodity	Land as a scarce resource	Land as a scarce community resource
Cadastral applications	Land valuation and taxation	Land-market applications	Land use planning	Sustainable development
	Fiscal cadastre	*Legal cadastre*	*Planning and land management cadastre*	*Multipurpose cadastre*

In the 18th century, when the term cadastre was first used, the main role of the cadastre was to support taxation. As a fiscal tool at that time, the cadastre provided durable land information as a basis for taxing the nobility in parts of Italia and Austro-Hungarian Empire.

The global trend towards private individual land tenure in the 1900s was due to the popularity of land as the principal means of distribution and management of wealth. The individual ownership of land was to be formalized through the registration of interests in land using the cadastre as the basis. This practice saw the emergence of the legal role of the cadastre. For land registration systems based on the title, the legal cadastre is the core component while jurisdictions with well-established fiscal cadastres developed separate deeds registration programme.

The human population boom after the Second World War made it necessary to carry out better spatial planning, particularly in urban areas to accommodate the rising urban population. Large scale topographical maps together with the cadastre, as a record of land parcels and the registry of ownership became a useful tool for urban planning and for the provision of vital services such as electricity, water, sewerage and so on. In this way, the role of the cadastre was extended to include spatial planning.

In 1980s the realization of the scarcity of land and need for better resource management, better environmental management and sustainable development created a demand for more land information. This was achieved by the integration of the cadastre and other land information for better decision-making, thus making the cadastre applicable to multiple tasks in addition to the initial land administration role.

2.5.2 The changing concept

The revolution in information technology since 1990s has largely contributed to the continued evolution of land administration domain and in the concept and scope of the cadastre. In this respect, a number of international efforts in form of declarations and conferences have taken place with a common goal to facilitate the understanding of the evolving concept of land administration. The first international effort was the FIG statement on the cadastre (FIG, 1995). The statement proposed definition of cadastre, which is still relevant. The definition as already stated in section 2.2 embraced the basic functions of land administration served by cadastres.

Bogor declaration

The United Nations Conference on the Environment and Development (UNCED) in Rio de Janeiro, Brazil in 1992 addressed the problems of land management and environmental management. The recommendations of the conference constitute the Global Plan of Action for HABITAT II. This plan of action, popularly known as Agenda 21 (UN, 1992), consists of a comprehensive blueprint of action to be taken globally, nationally and locally by organizations of the UN, governments, and major groups. It concerns almost every area in which humans directly affect the environment.

An inter-regional meeting of experts on cadastre held in Bogor, Indonesia in 1996 was meant to develop an active response to the problems of land management and environmental management addressed in Agenda 21. The outcome of the meeting in Indonesia commonly referred to as the Bogor declaration (FIG, 1996) was the official recognition for the first time of the cadastre as a core infrastructure supporting a sustainable environmental and natural resource management.

Bathurst declaration

As a follow up to the cooperation between the International Federation of Surveyors (FIG) and the United Nations organization, a conference similar to the one held in Bogor, Indonesia, was held in Bathurst, Australia in 1999. The outcome of the meeting commonly referred to as the Bathurst declaration on Land Tenure and Cadastral Infrastructure for Sustainable Development established a powerful link between good land administration and sustainable development (FIG, 1999).

Cadastre 2014

The FIG under its Commission 7 on Cadastre and Land Management initiated a study to explore cadastral reforms in developed countries and develop a vision of the future cadastral systems. This was motivated by the expanded role of the cadastre to multiple applications and in response to new technology. The outcome of that study is the popularly known publication, Cadastre 2014 (Kaufmann and Steudler, 1998). The publication gives a new definition of cadastre based on the expected nature and characteristics of cadastres by the year 2014 as:

"a methodologically arranged public inventory of data concerning all legal land objects in a certain country based on the survey of their boundaries".

This definition extended the definition of the cadastre as was known until then by introducing the idea of a land object. A land object according to the definition does not only include the traditional land parcel based on ownership rights only, but also other land objects defined based on restrictions and responsibilities. Either public or private law defines these land objects. The law may define rights, restrictions, responsibilities or phenomena, which are related to a fixed area on the surface of the earth. Examples of land objects include the traditional land parcels, buildings, encumbrances, easements, land use zones, environmentally protected areas among others.

2.5.3 The changing scope

From 2D to 3D to time-enabled cadastres

Cadastres in most countries consist of land parcel boundaries fixed in two-dimensional spaces. The extension of ownership of rights to the third dimension has come about because of the intensive use of land space particularly in urban areas and because of improved building technology. It is now possible for different individuals to own units or apartments, which are above or below and within the same land parcel. Technically, it involves the vertical subdivision of the land parcel. This concept referred to as 3D cadastre is variously recognized in different jurisdictions as strata titles (i.e., because of stratification of land ownership), sectional properties, condominiums among others. The representation of the third dimension has proved to be especially relevant for apartment units and for physical objects that lie above or below land parcels, such as tunnels, underground shopping malls and utility networks. 3D cadastre issues first received attention in 2003 in a workshop organized by the FIG (Lemmen and van Oosterom, 2003) to consider the merging with regard to this concept. Currently, there are efforts to identify the various conceptual models related to 3D cadastre in different countries.

One of the major concerns in realizing 3D cadastres is how to attach the third dimension to already existing 2D cadastres. Filin and others (2005) proposed an algorithmic approach for extracting and attaching height information from aerial laser scanned data to already exiting 2D cadastres. The approach is similar in concept to the reconstruction of high resolution height models from airborne laser scans by incorporating existing ground plans developed earlier by Brenner (2000).

The Light Detection And Ranging (LIDAR) technology has become an essential tool for surveying and processing of geospatial data. The resulting dense and accurate 3D point cloud of the scanned area provides a possible basis for establishing a 3D cadastre. The non-explicit representation of the point clouds however requires automatic or semi-automatic techniques to differentiate between ground and non-ground features such as buildings.

The time dimension is another addition to the cadastral concept. It arose out of the need to record how the legal and geometric status of a land parcel is changing over time. The idea of including the time in the cadastres extends the concept further from 3D to 4D cadastres (or simply time-enabled cadastres). Various aspects of 4D cadastre are presented in Döner and others (2010), in particular the approaches for modelling the time dimension. In most cadastral registers, time stamps that indicate the creation and revision of objects contained in the cadastre already contain the time dimension.

Land to marine cadastres

The responsibilities and opportunities of governments for land and resource management extend beyond land to the marine environments. The management of the numerous activities that take place in the coastal and marine environments has created the need for a marine cadastre. The common land use activities include, for example, tourism and recreation, mineral and energy exploration, cable and pipeline, shipping and fishing (Binns, 2003).

The concept of a marine cadastre has been addressed in a number of meetings and many of the components of a cadastre in land have parallel conditions in a marine environment. A marine cadastre is defined as a marine information system, encompassing both the nature and spatial extent of the interests and property rights, with respect to ownership and various *rights, responsibilities and restrictions* in the marine jurisdiction. Thus, the concept of the cadastre yet again extended in scope to include not only rights, responsibilities and restrictions in marine environments but also the consideration of their third and four dimensional aspects (Ng'ang'a et al. 2004).

2.5.4 From mapping to data modelling

The use of technology for coordinated and efficient land administration services like e-conveyance, provision of online access to survey plans and digital processing of title transactions require an effective cadastral data model. A cadastral data model is an integrated set of concepts for describing and manipulating cadastral data, the relationships between them and the constraints on the cadastral data. An effective cadastral model must describe what is fundamental to the business and not simply what appears as data (Kalantari, 2006).

The differences in the cadastral concept in different countries are also reflected in cadastral data modelling. For example, in countries with a legal land registration background, land information is organized with land parcels as the basic building blocks. In countries where land administration systems have a fiscal land registration background, land information systems are organised by properties. In most land registries, interests on land include only rights rather than responsibilities and restrictions. Restrictions are usually described with land use and imposed by planning and development agencies and do not appear in cadastral databases (ibid). These examples illustrate the challenge in cadastral data modelling both within one country and internationally. Data modelling can be carried out either at the cadastral level or generally at the land administrating level.

Cadastral data modelling

The two particular cadastral data modelling efforts that are of international nature include Cadastre 2014 and the Core Cadastral Domain Model. Cadastre 2014 (Kaufmann and Steudler, 1998) is based on three objects: the legal land object (instead of only the parcel), the person and the right or restriction. The Core Cadastral Domain Data Model (CCDM) (van Oosterom et al., 2006) is based on the principle of the relationship between the basic entities of a cadastral system namely: real estate object, right or restriction, the person and their relationships.

Comparatively, Cadastre 2014 gives an excellent modelling start but it is a generic or abstract set of guidelines, which must be refined further into a more specific model. The initial proposal of the CCDM was based on Cadastre 2014. In fact, the two models compare with the abstract and the implementation level of the Open Geospatial Consortium (OGC) specifications (OGC, 2002). The main difference between the two models is that Cadastre 2014 stipulates a very new cadastral system that is land object-based, while the CCDM follows the principle of the traditional "parcel-centric" approach. Generally, the two models reveal the problem of lack of shared concepts and terminology about the cadastre. Moreover, cadastral systems are only part of the large land administration systems.

The Land Administration Domain Model(LADM)

International standardization of concepts and development of ontologies in both the cadastral and land administration domains would facilitate meaningful exchange of cadastral information between organizations or effective component based system development through applying standardized models. The current paradigm in ISO TC/211 involves the development of domain specific standards, for example, Land Administration Domain Model (LADM) (ISO, 2011).

The LADM is the successor of CCDM. The model attempts to achieve standardization in the area of cadastral data, provide common definitions for land information and facilitate effective use, understanding and automation of land related data, thereby enhancing data sharing. The development of the model was based on the following principles. To cover the common aspects of land administration all over the world, consider the conceptual framework of Cadastre 2014, to follow ISO standardization, and to create a model that is simple enough to be useful in practice.

The current edition of the LADM as a Draft International Standard (DIS) (ibid.) is expected to provide an abstract, conceptual schema with four basic classes (see Figure 2.6). The classes (or packages) include LA_Party (Party) includes people and organizations, LA_RRR (Rights, Responsibilities and Restrictions) entails for example the ownership rights, LA_SpatialUnit includes for example, the parcels, buildings and networks, and LA_BAUnit (Basic Administrative Unit) are spatial units against which a unique and homogeneous interests are associated to the whole entity.

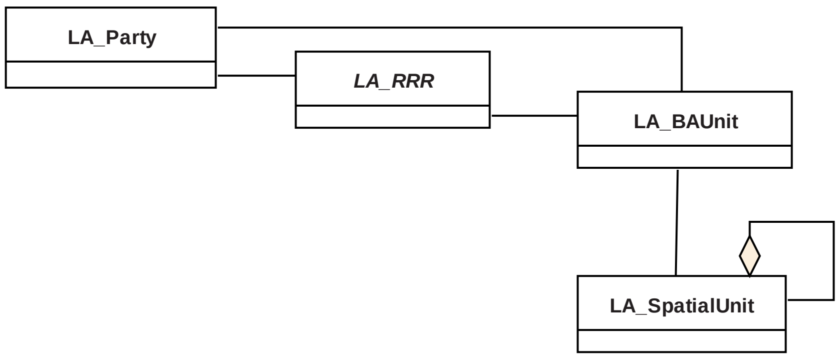


Fig. 2.6: Basic classes of the Land Administration Data Model (LADM)

The support for packages facilitates the maintenance of different data by different organizations. The complete model might be implemented through a distributed set of geo-information systems, each supporting data maintenance activities and the provision of elements of the model. The model might also be implemented by one or more maintenance organizations operating at national, regional or local level. This underlines the relevance of the model: different organizations have their own responsibilities in data maintenance and supply, but can communicate based on standardized administrative and technical update processes.

The Social Tenure Domain Model(STDM)

Based on the observation that existing Land Administration Systems (LAS) are not designed to accommodate informal and customary tenures, a Social Tenure Domain Model (STDM) was suggested (FIG, 2010). The model allows for recording of all possible types of tenures.

The STDM considered as a specialisation of the LADM and is integrated in the standardisation process of the LADM. The LADM is considered to be designed for areas with formal cadastre and land registry systems, but also contains the functionality for the STDM under different terminology.

In the STDM, in principle, the same classes and structures are used as in LADM, but with different terminology. For example, a RegisterObject is named SpatialUnit and a RightRestrictionResponsibility is called SocialTenureRelationship. This makes the two approaches compatible (Lemmen et al., 2009). The LADM enables the combination of land administration information from different sources in a coherent manner. The LADM can include informal and customary rights. It is expected that there could be no interference with (national) land administration laws (ISO, 2011).

2.5.5 Characteristics of future cadastres

Cadastre 2014 has been an influential contribution to the role and concept of cadastre. It is however becoming clear that as the time closes in, the objectives of Cadastre 2014 will continue to be the target for most jurisdictions. In this respect, Bennet and others (2010) have isolated and described six design elements that will characterize future cadastres. They include:

- Moving from approximate boundary representations to survey-accurate boundary representation;
- Shift of focus from purely parcel-based towards systems of several layers of land objects;
- Expansion from merely 2D to include the third (height) and the fourth (time) dimensions;
- Updating and accessing of cadastral information in real time;
- Making national and state-based cadastre interoperable at regional and global levels;
- Inclusion of modelled organic natural environments that allow fuzzy and dynamic boundary definitions that are not necessarily designed around strict bearings and distances or Cartesian coordinates.

3. The Kenyan Cadastre and Modern Land Administration

Cadastral datasets form the basis for SDI dedicated to land administration. Their usefulness is enhanced if they are consistent with other topographical datasets. This is because cadastral and topographical datasets are the most important datasets in any country. These datasets provide the foundation for the concept of spatially-enabled society, in which location, place and other spatially information are available to governments, citizens and businesses as a means of organizing their activities and information (Williamson, 2010b).

A cadastral system is not just about the concept of the cadastre or the individual activities of title registration or cadastral surveying (Williamson, 1996). Instead, it is embedded in the various land administration processes, which depend on a country's political, historical and cultural systems. As such, cadastral systems differ from one country to another.

The aim of this chapter is to describe the Kenyan cadastre and the status of nationwide topographical mapping. The cadastre is then evaluated against the design elements of future cadastres identified in the previous chapter. The evaluation will identify the particular aspects of the cadastre with regard to the creation of an integrated cadastral framework from existing cadastral datasets.

3.1 Historical outline of the Kenyan cadastre

In Kenya, the cadastre is related to the land registration regimes because the cadastral processes are integrated in the land registration process. In this respect, the cadastre has developed through four chronological periods: before 1902, between 1902 and 1945, between 1945 and 1963, and since 1963.

3.1.1 Before 1902

Prior to 1902 when the Crown Lands Ordinance was enacted vesting all land in what is Kenya today to the British Crown (all land in the colony including the land occupied by the indigenous people), land was held under customary tenure. Under the customary tenure then, land could neither be considered as a commodity for sale, alienation or transfer. During this period, the cadastre was simply non-existent.

3.1.2 Between 1902 and 1945

After 1902, the British in the crown land introduced the concept of individual ownership. Individual ownership of land had to be registered according to a cadastral plan. Cadastral activities were effectively introduced in 1903 when a survey section was established and a chief surveyor appointed to superintend the demarcation and survey of parcels of land that had been alienated in Nairobi (Njuki 2001). This was realised through the enactment of four ordinances namely: the Land Titles Ordinance of 1908, the Crown Lands Ordinance of 1915, the Registration of Titles Ordinance of

1918, and the Land Surveyors Ordinance of 1923. These Ordinances guided the land tenure policies for almost fifty years.

Before the Second World War, cadastral survey activities were confined mainly to the land alienation programme at the Kenya Coast and the British settlements in Kenya's temperate highlands occupied by farmers of European origin. These regions subsequently became known as the "White Highlands". The objective of the programme was to alienate crown land to the British community according to the provisions of the Land Surveyors Ordinance of 1923 and registered under the provisions of the Registration of Titles Ordinance. The two ordinances required only fixed-boundaries whereby land was demarcated by permanent survey marks and the position of the survey marks accurately determined by mathematical computations. All Crown grants were to be compulsorily surveyed before they could be registered.

3.1.3 Between 1945 and 1963

After the Second World War, cadastral activities were mainly carried out to support the land tenure reform policy of transforming land in trust land areas from the customary land tenure to the statutory freehold individual ownership. The Trust Lands Act of 1959 was enacted to provide for the registration of land in the trust lands. Land registration in the trust lands was mainly carried out under the land consolidation programme.

The Land consolidation programme involved the adjudication (formalization of land ownership), exchange and gathering of fragments of small sizes of plots that were in common ownership. This programme was carried out to create more economically viable units for each owner. The process of land consolidation as was carried out in Kenya is equivalent to the accelerated land consolidation procedure used in Germany (Thomas 2004). In both cases, the objective is to consolidated scattered and uneconomically shaped parcels without necessarily creating new road systems or water resources projects. Usually, the procedure is initiated by a state authority.

The areas where land consolidation programme has been carried out in Kenya are referred to as consolidation areas. The core of the land consolidation programme was in the Central Province of Kenya carried out under the provisions of the Land Consolidation Act of 1959. Neighbouring farmers were compulsorily displaced in an effort to create room for larger units. The compulsory relocation of people caused discomfort, thereby making the programme unpopular. Now, the land consolidation programme is restricted to the former Meru district. Interestingly, this programme, which began in 1966 in this area, is yet to be completed.

3.1.4 1963 to present

After independence, the main land reform programmes, i.e., programmes that involve the changing of laws, regulations or customs regarding land ownership were land adjudication, land redistribution and land allocation. These programmes influenced the current nature of the Kenyan cadastre.

i) Land adjudication

Since the land consolidation programme proved to be unpopular and slow because the desired targets could not be achieved, the government appointed a mission on Land Consolidation and Registration (1965-66) to find ways and means of accelerating the process. The mission recommended the process of land adjudication (Lawrence et al. 1966) mainly in agriculturally high potential trust lands as the basic process for the ascertainment of land rights. In addition, the requirement to compulsorily survey land before it is registered was changed after the government realized that a lot of development was being hampered by the slow pace of settlements caused by backlog in land surveys.

Land adjudication programme was the main land tenure reform carried out after independence. The programme was meant to make individual land titles available to the indigenous people by formalizing their customary rights where they were living. The areas where land adjudication has been carried out without consolidation are called enclosure areas, named after the enclosure movement of the 1700s. The movement was a controversial process of taking common lands for traditional purposes such as communal farming, grazing, hunting and access to timber and other resources, and fencing and placing them in private ownership (Williamson et al., 2010a).

Land adjudication is also carried out in the rangeland areas (also called group ranches) for nomadic pastoral communities, where group ownership is preferred to individual ownership. The parcels of land in these areas are registered in the name of the group representatives in trust on behalf of the rest of the group members under the provisions of the Land (Group Representatives) Act of 1968. A group in this context refers to a tribe, a clan, a family or any other group of persons whose land, recognized under customary law, belongs communally (undivided) to more than five persons who are members of that group. Each group selects about ten of its members to be registered as trustees of the land by the Government. These trustees can allocate portions to the group. This model of land tenure was introduced in Kenya as part of the African Land Development (ALDEV). It aim was to improve on the carrying capacity of the land, the productivity of cattle, and to control the ecological imbalance usually associated with such fragile ecosystems (Wayumba 2004). By far, land adjudication has been the largest programme for individualization of tenure ever undertaken in Kenya and in a systematic manner.

ii) Land redistribution

Individualization of tenure has also been realised through land redistribution and through allocation of government land. The redistribution programme based on a "willing buyer, willing seller" arrangement, owes its origin to the Swynnerton Plan of 1954 (Swynnerton, 1954). The main aim of the plan was to intensify the development of African agriculture in the colony of Kenya through the redistribution of land to more efficient farmers through land markets. The programme was effected mainly through government and private initiatives. After independence, through a government initiative, the government would buy land from the British settlers and redistribute it to the indigenous people through the Land Settlement programme. Through private initiatives, companies and cooperative societies would buy large farms previously owned by settlers, subdivide and allocate them to a large number of farmers. The areas where this programme has been carried out are called settlement schemes.

iii) Land allocation

Although the allocation or alienation was the first tenure process introduced through which the government could lease land to individual owners, the process was reinvigorated after independence. Land allocation was and still is carried on a need basis, but especially for residential, commercial and industrial development in urban areas. This programme has been going on for over 100 years and land parcels of varying sizes are involved.

3.2 Land tenure and land tenure patterns

Land tenure entails the terms and conditions under which land is held, used and transacted and is one of the principal factors determining the way in which resources are managed and used and the manner in which benefits are distributed (UNECA, 2003). It is multi-dimensional, bringing into play social, technical, economic, institutional, legal and political aspects into account. Land tenure relationships may be well-defined and enforceable in a formal court of law or through customary structures in a community.

Land policies and legislations in Kenya have over the years given rise to three types of land tenure systems. These are private, customary and public tenure system.

Private tenure confers on the individual or corporate entity an indefeasible and exclusive title to a specified estate in land. This includes all land held on freehold or leasehold by individuals, companies, co-operative societies, religious organizations, public bodies, and legal bodies. Private land may be as a result of alienation of government land, adjudication of trust land.

Customary land tenure is mainly found in areas that have not yet been transformed though adjudication, consolidation and registration. Under customary tenure, land belongs to a clan, ethnic group or a community as a whole. Areas under customary tenure system are designated as trust land

Public tenure establishes control over forests, national parks, open water, townships and other urban centres as well as alienated and un-alienated government land. In effect, this tenure arrangement designates the government as a private landowner.

The most extensive tenure system is trust land which takes up about 78.4% of the total land surface. Most of this land is available for smallholder registration. Government land takes about 20% while freehold land is only 1.5% (Odhiambo and Nyangito, 2003). The proportions are not because of any policy requirement but as a result of the natural registration and land transfer processes. These categories have been renamed as community, public and private lands respectively in the national land policy and in the recently promulgated constitution (GoK 2010).

According to the new constitution, public land is land vested in and held by the government in trust for the people of Kenya. The Constitution and the National Land Policy contain clauses that specify what is considered public land, for example, land held, used or occupied by a state organ, national parks, water catchment areas, all rivers, lakes, and specially protected areas. Recognizing the non-existence of a system for registering public land, the policy requires mechanisms to be established for registering public institutional land and to establish mechanisms for repossessing any public land acquired illegally or irregularly.

Private land is land held by individual persons or legal persons like private companies and co-operative societies after alienation from government land or adjudication from trust lands. The land is held under either freehold tenure or leasehold tenure. The policy requires the land to be held on terms that are clearly sub-ordinate to the doctrines of compulsory acquisition and development control.

Community land is the land vested in and held by communities identified based on ethnicity, culture or similar interest. Community land include land lawfully registered in the name of group representatives under the provisions of any law, ancestral lands and lands traditionally occupied by hunter-gatherer communities, and land held lawfully as trust land by county governments. This includes, for example, rural markets, rural public schools etc. Any unregistered community land shall be held in trust by county governments on behalf of the communities for which it is held. Concerning community land, the policy recommends the documentation and mapping of existing forms of communal tenure in consultation with the affected groups and incorporates them into broad principles that facilitate the orderly evolution of community land laws. This does not matter whether the tenure is customary or contemporary; in the rural or urban area.

Clearly, the largest category of land in Kenya is the community land. This is because much of communally owned land, which mainly lies in the rural areas, has not been registered and mapped. Although a greater percentage of community lands are in the rural areas, still some secondary towns in Kenya are predominated by community land tenure. Consequently, there is usually shortage of serviced land for all forms of urban development due to limited availability of public land. In addition, it is generally argued that customary land tenure is less responsive to general urban land demand (Olima and Obala 1998).

3.3 The Kenyan cadastre

The nature of the Kenyan cadastre depends on the efficacy of the different land tenure process to create a homogeneous cadastre and the type of land involved. However, because of the different requirements of these processes, the existing cadastre consists of a patchwork of cadastral maps of different qualities. In this section, the Kenyan cadastre is described in terms of the existing cadastral maps, their common mapping scales, their typical positional accuracies and their combined coverage.

3.3.1 Cadastral maps

Different types of maps are used to support land registration in Kenya. The most famous maps are the survey plans for urban areas and the Registry Index Maps (RIM) for rural areas, which are respectively numeric and graphical. Other maps used to support land registration are graphical and only used provisionally. As an integral component in the land registration process, these maps are created during a first registration of the land and amended afterwards in case of a sub-division of the land.

i) Survey and deed plans

A deed plan, usually produced for each parcel of land in urban areas, is traced from a survey plan based on fixed-boundaries. A survey plan and a deed plan, courtesy of the Survey of Kenya, are depicted in Figure 3.1. A survey plan [Figure 3.1 (left)] typically shows measurements (bearings, distances, areas and coordinates) of the surveyed parcels and the adjoining parcels. A deed plan [Figure 3.1 (right)] which is usually abstracted from the survey plan typically shows the bearings, distances and area of an individual land parcel. The survey and the subsequent deed plan are named so because they are prepared from surveys that are based on a planned layout of land parcels. The preparation of the Part Development Plan (PDP) on which the surveys are based precedes the survey.

Typically, survey plans and the subsequent deed plans are of the highest positional accuracy. This is because survey plans are based on accurate fixed-boundary (coordinated) surveys. Survey plans are based on ground survey methods that ensure that the requirements of fixed-boundary surveys are met. The methods employ precise equipment and techniques to obtain accurate and reliable mathematical data from which the numeric cadastre is created. These methods use the traditional optical survey equipment including theodolites, tachometers, Electro-optical Distance Measurement Equipment (EDM), total stations, and the Global Positioning Systems (GPS). The only problem with fixed-boundary surveys is that they are expensive to carry out.

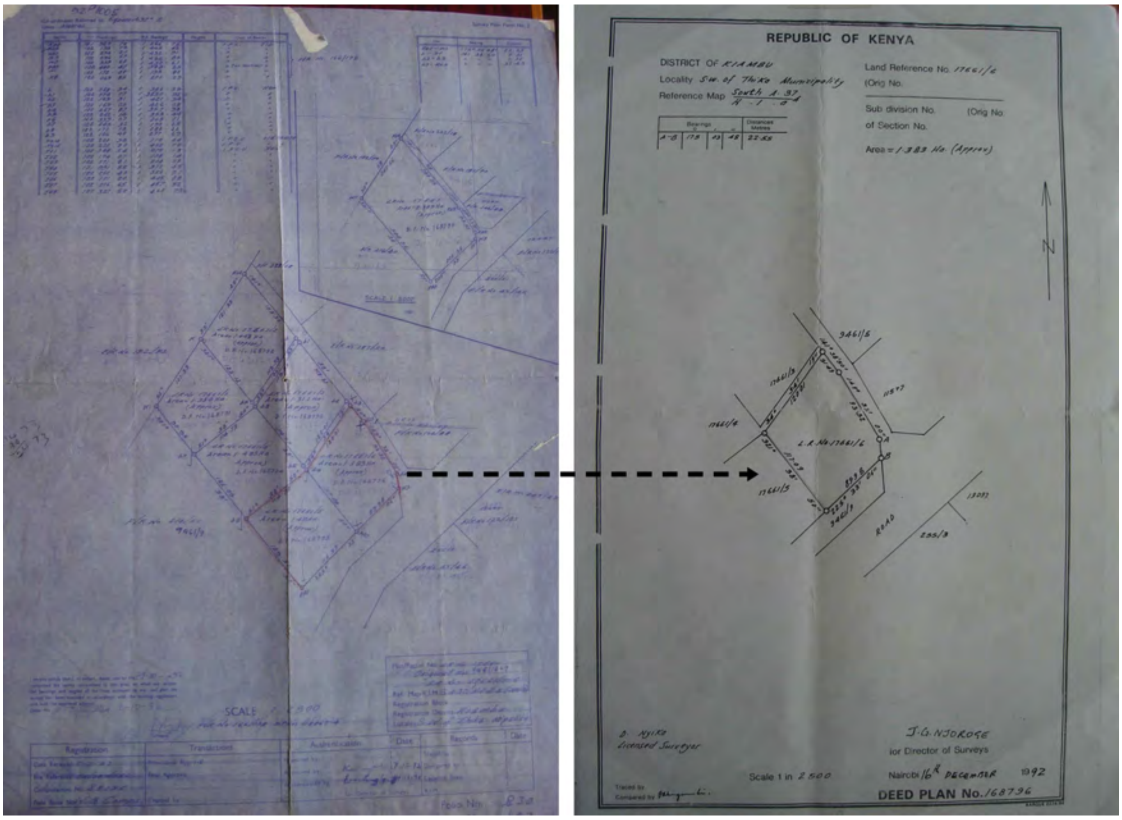


Fig. 3.1: A Survey Plan (left) from which a Deed Plan (right) is extracted

ii) The Registry Index Map (RIM)

According to the provisions of the Registered Land Act of 1963 (GoK, 1963), a Registry Index Map (RIM) is the main cadastral map usually prepared for the first registration of land and amended during subsequent land subdivisions. RIMs are based on "general boundaries". Therefore, RIMs are only used to identify on the ground a plot shown on the register, assist in the re-location of a boundary, should it be lost, enable sub-division to be effected, and for approximate calculation of areas.

The RIMs are prepared for every registration district (equivalent to an administrative district) or a part thereof and is divided into registration sections, which are identified by distinctive names. The registration section may be further divided into blocks, which are identified by either numbers or letters or a combination of both. A RIM (example in Figure 3.2) typically shows the outline of all individual land parcels within a given jurisdiction and contains information about the location, sheet and index number, edition of the sheet, map sheet history (amendments), plot numbers and scale. The scale is very important since no measurements are provided to show the dimensions of the boundaries, and the map user can only rely on the scale ruler to scale off the distances from the map.

RIMs are prepared through various methods depending on the programme and type of legislation used for registering rights to land (Wanyoike, 2001).The programmes include land consolidation, land adjudication and land settlement.

In land consolidation and land adjudication programmes, RIMs are mainly produced through a process known as "refly". The process is carried out after existing rights in any particular parcel of land are authoritatively ascertained and confirmed in an adjudication register. After hedges have been planted and are air "visible", aerial photography carried out at the scale of 1:12,500. From

Fig. 3.2: A Registry Index Map (RIM)

these photographs and after ground control was provided by means of triangulation or trilateration, maps at the scale of 1:2,500 were produced photogrammetically showing the boundaries. Ground survey methods are then used to mark the missing boundaries. The RIM for consolidation areas are prepared at a scale of 1:2,500.

In the land settlement programme, the “refly” process was used in the One Million Acres Schemes in Nyandarua District, drawn at 1:10,000, were produced by this process. In other settlement schemes, ground survey methods are used by the Survey of Kenya and licensed surveyors for production of accurate Registry Index for a number of settlement schemes including Shirika, Sugar Settlement Organization, Ol’Kalou Salient and Magarini settlement schemes. Ground methods are also used for maps in urban areas and in the subdivision of Company and Cooperative farms (Mwenda, 2001).

The obvious limitation of the RIM is the lack of indication of measurements on both the length of the boundaries and of the areas of the individual parcel. Another limitation is that since all amendments are made on the original map sheet whose scale is fixed, the map can eventually get very congested due to continuous changes on consecutive resultant parcels, which may lead to illegibility of the map.

iii) Provisional cadastral maps

The original intention that all demarcation plans would serve temporarily as registry maps and then replaced by more accurate registry maps within a period of 18 months was however not realised. Instead, there was a backlog because of considerable mapping commitments posed by mapping holdings in the settlement schemes and in the former Scheduled area (Lawrence et al., 1966). In this regard, the "refly" process was abandoned in 1967 due to its slow speed and expense and the mapping efforts were directed to the accelerated programmes of land consolidation and registration. In these programmes, demarcation plans (maps) of sufficient accuracy were used as registry maps. Similar temporary demarcation maps called "Preliminary Index Diagram" and provisional Registry Index Maps - Rangeland were introduced.

Demarcation maps

The Registered Land Act of 1963 allows the use of provisional cadastral maps for land registration prior to the preparation of more accurate RIMs. For example, during the land consolidation programme, cadastral surveying and mapping was based on aerial photography at a scale of 1:12,500, from which base maps at a scale of 1:5,000 are prepared. The base maps were subsequently enlarged to scale of 1:2,500 for purposes of area computation and as the basis for the preparation of an "Allocation Plan". The allocation plan is then used for demarcation of the boundaries on the ground as reflected on the allocation plan. The maps used to support registration are then prepared by tracing the allocation plans and are known as demarcation maps, which are provisional RIMs. Figure 3.3 shows an example of a demarcation map (courtesy of the Survey of Kenya). The Department of Physical planning prepares demarcation maps called Development plans for municipalities, towns and settlement schemes.

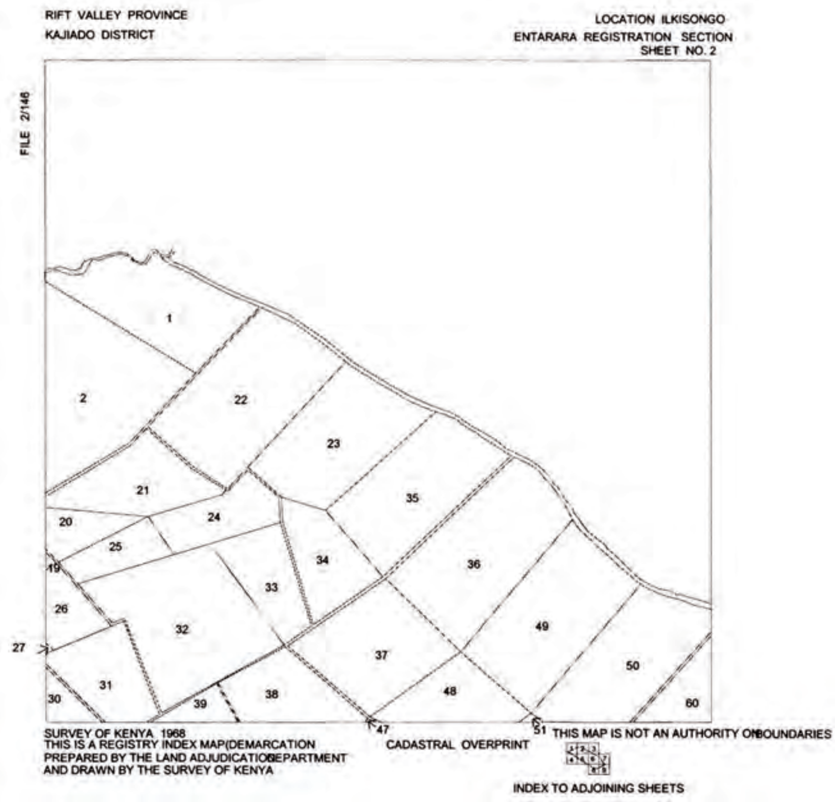


Fig. 3.3: A Demarcation Map for part of a Registration section

In the earlier days of land the consolidation programme, these provisional RIMs were upgraded through the "Refly" process. The main limitation of the demarcation maps is the absence of measurements.

Preliminary Index Diagram (PID)

Cadastral survey and mapping for land adjudication in enclosure areas uses provisional maps called Preliminary Index Diagrams (PID). These maps are created by the land officers using enlarged aerial photographs or by ground methods or both. Only in a few cases in enclosure areas have boundary surveys been done on base maps using plane table surveying. In the majority of cases, enlarged aerial photographs are used. The process entails undertaking aerial photography at a scale of 1:12,500 or 1:25,000. These photographs, which are usually un-rectified and mosaicked, are enlarged to scales of 1:2,500 or 1:5,000. After the identification and marking of parcel boundaries on them by photo interpreters, parcel boundaries are then traced out from the marked photographs, thereby producing an interim map for registration.

PIDs (illustrated in Figure 3.4 (courtesy of the Survey of Kenya)) are prints of maps showing approximate parcel boundaries. Their main weaknesses are the non-uniformity of scale within a particular index map sheet, unreliability of distances and areas calculated using these map sheets, and sometimes distortion of the shapes of the land parcels. This is largely because the PIDs are not geometrically controlled in any way.

Fig. 3.4: Left -marking of parcel boundaries on un-rectified, uncontrolled and mosaicked aerial photographs, Right -resulting PID after tracing from the marked photographs

Registry Index Maps - Rangeland (Provisional)

In Rangelands where trust lands are converted to group ownership, natural features appearing on published 1:50,000 topographical maps are accepted as boundaries. Boundary markers though coordinated to the nearest metre using approximate ground survey methods are marked on the 1:50,000 maps but their coordinates are not often used to compute the areas.

3.3.2 Common map scales and typical positional accuracies

The different maps used to support land registration have different positional accuracies depending on the registration legislation and the method of survey used. For the preparation of survey plans and deed plans, the land parcels are mapped to the highest positional accuracy specification of 0.03m. The RIMs for urban areas, although based on both fixed and general-boundaries, have a nominal positional accuracy of about 0.30m. The Demarcation map, the PID and the RIM Range land (Provisional) have variable geometric accuracies. They are considered as interim RIM subject to more accurate surveys. For PIDs, Mulaku and McLaughlin (1996), for example, showed that deficient inaccuracies of more than 20m in position and more than 50% in area are not unusual. Table 3.1 gives a summary of the cadastral maps used in Kenya based on the common scales and typical positional accuracies. For the survey plans and deed plans, the actual positional accuracy of the survey-accurate field measurements is usually the highest (±0.03m).

Table 3.1: Cadastral maps and their common scales and positional accuracies

Type of map	Common scales	Positional accuracy
Survey Plans/ deed plans	1:2,500 - 1:5,000	± 0.03m
Registry Index Maps	1:2,500 - 1:10,000	± (0.625 - 2.5) m
Demarcation maps	1:2,500 - 1:10,000	± (0.625 - 2.5) m
Preliminary Index Maps	1:2,500 - 1:5,000	± (0.5 - 20) m (variable)
RIM Range (Provisional)	1:50,000	± 10 m (within scale)

3.3.3 Cadastral coverage

Because of the voluntary nature of land registration and integration of cadastral surveying, mapping and registration processes, cadastral coverage is incomplete, in addition to being of inconsistent positional accuracy. To determine the extent of cadastral coverage, knowledge of the extent of land registration is important. Table 3.2 gives a summary of cadastral coverage in terms of the category of the land parcels, the type of map used and the total surface area covered. The values quoted in Table 3.2, were obtained from Departments of Survey and Land Adjudication and Settlement (MoL 2010) and in (Njuki 1999). The values are approximate do not necessarily reflect the current situation. Latest statistics on government land in urban and rural areas in terms of coverage are not readily available. With a total land surface area of about 582,600 square Kilometres, 14.6 million Hectares of registered and mapped land represent a coverage of about 25%. About 75% of the coverage consists of provisional cadastral maps.

3.4 Nationwide topographical mapping

By its very nature cadastral data is captured and represented at large scale and very different from topographical data, which is produced at medium to small scales over large regions using a range of different techniques. Despite these differences, the SDI provides the basis for integrating the built (cadastral) and natural (topographical) spatial data (Rajabifard and Williamson, 2004) as long as the mapping scales are comparable. The rest of this section presents an outline of the status of topographical mapping in Kenya.

Table 3.2: Cadastral coverage

Category of land	Map	Area(Million Ha)	% of the total
Trust land			
Consolidation Areas and Enclosure Areas	Demarcation maps/RIMs /PID	8.0	54.8
Group Ranches	RIM(provisonal)	3.3	23.0
Private Land			
Settlement Schemes	RIM/PIDs	1.012	7.0
Company and Cooperative farms	Deed plans/RIM(provisional)	2.2 [1]	15.1
Government Land			
Urban and rural Areas	Survey plans and RIM	0.101 [2]	0.1
Total		14.6	100.0

[1](Njuki, 1999)
[2]This is an approximate value for 250,000 plots of various sizes in urban areas (a plot size of 1 acre is assumed)

3.4.1 Large scale base topographical maps

During the land settlement and the land consolidation programmes, large scale topographical base maps at scales of 1:2,500 and 1:5,000 were prepared to support these programmes. The maps were used to plan the parcel layouts that were demarcated during the implementation of these programmes. These maps have been prepared from aerial photographs utilizing conventional analogue stereo plotters.

Large scale base topographical maps are prepared at various scales depending on developments within the area of settlement or planned development. In Kenya, most of the large scale maps are for the city of Nairobi, municipalities and other towns. The majority of these maps are prepared at a scale of 1:10,000, but in a few cases where there is more intense development, maps are prepared at large scales. For instance, a number of towns have also been mapped at scales of 1:2,500, 1:5,000 for different road design and construction projects. Approximate analogue spatial coverage at these scales is indicated in the second column of Table 3.3.

3.4.2 Basic topographical mapping

The basic mapping scale for topographical mapping in Kenya is 1:50,000. This is the basic mapping scale from which all other smaller scales maps are derived. About sixth three percent (63%) of the country is covered at this scale. However, most of the maps at this scale have not been updated for a long time. The entire country is covered by maps at both scales of 1:250,000 and 1:1,000,000. Only some areas are also mapped at a scale of 1:100,000 (Njuki, 2000).

There have been efforts to digitize all the basic topographical maps at 1:50,000 under the National Digital Topographical Database (NDTDB) programme. The datasets are basically in the shapefile data format. The digital map coverages that have been undertaken under this programme are given in Table 3.3 according to Mbaria (2002). During the NDTDB programme, the Kenya portion of Global Map was prepared and is already available online.

The Global Map is an initiative of the International Steering Committee for Global Mapping (ISCGM) articulated by the Ministry of Construction of Japan in response to the United Nations Conference on Environment and Development (UNCED) (ISCGM, 2002). The fundamental basis of the initiative was to develop global scale geographical information through international cooperation. That is to develop and provide easy and open access to Global Digital Geographic Information at a scale of 1:1 million with specific standards for interoperability.

The Global Map consists of both vector and raster data. The vector data consists of four layers: transportation (roads, railways and airports), boundaries (international, provincial, district, etc), hydrological (rivers, streams, lakes), and population centres (towns, municipalities, etc). The raster data consist of four layers: elevation, land cover, land use and vegetation.

Table 3.3: Digital topographical coverage

Reference scale	Analogue coverage	Digital coverage
1:1,000,000	100%	100% (entire country)
1:250,000	100%	30%
1:50,000	63%	10%
1:20,000	0.12%	0.12%
1:10,000	0.45	0.05%
1:5,000	0.26%	0.05%
1:2,500	≈ 4%	≈ 4%(various towns in Kenya

3.5 The Kenyan cadastre and modern land administration

Multipurpose cadastre, cadastre 2014, the modern land management paradigm and the Spatial Data Infrastructure have all altered the understandings and the role of the cadastre. In their quest to provide an outline of the role and nature of future cadastres, Bennet and others (2010) came up with six design elements, which will characterize future cadastres. These elements together with those used in Siriba and others (2011) are used as the basis to evaluate the current Kenya a cadastre with the aim of identifying the main aspects that will be addressed when creating a nationwide cadastral fabric.

3.5.1 Survey-accurate cadastres

Cadastres in many jurisdictions were mainly created to support land administration, with little consideration for other applications. Many applications nowadays however require and depend on survey-accurate cadastres. This includes for example, building management, utility administration and so on. Therefore, only survey-accuracy will enable the complex layering of property interests to be accurately understood.

Whereas earlier analogue cadastral maps in some countries exhibited survey-accuracy (i.e., centimetre accuracy), the crude digitization of these maps resulted in large errors being introduced in the corresponding digital cadastral databases. Therefore, even in such jurisdictions, survey-accurate cadastres will be an on-going aspiration as the cost and expertise required for its implementation reduce. This is because any survey-accurate cadastre will definitely fit all purposes.

The Kenya cadastre is still analogue with a nationwide coverage of about 25%. Moreover, this coverage is not continuous; instead, it consists of maps of different positional accuracies as shown in Table 3.1. In addition, the maps are based on different coordinate systems (UTM, Cassini-Soldner and local coordinates) with a number of them, particularly the Preliminary Index Diagrams (PID) not based on any coordinate system at all.

The only public initiated effort to improve provisional cadastral maps was the “refly” process used during the land consolidation programme to map the land parcels more accurately. Otherwise, the other process is the “fixing of general boundaries” in the areas that mapped with general boundaries.

An individual mainly initiates this process and since it is not systematic, achieving nation coverage in not feasible.

A nationwide survey-accurate cadastre can be created either by carrying out new field surveys or by integrating the existing cadastral maps. The first alternative is expensive considering that, the creation of a cadastre depends on some lengthy and costly legal procedures. In this regard, the creation of a nationwide survey-accurate cadastre can be achieved by the integration of the existing cadastral maps at the same time paying attention to special characteristics of these maps. For instance, the existing numeric cadastres do not have to be digitized but be recreated from the existing field measurements. The inaccurate graphic cadastres could first be digitized. They can then be geometrically enhanced through data conflation techniques such as the one proposed in this thesis.

3.5.2 Object-oriented cadastres

Traditionally, a cadastre has been defined to include land parcels and additionally buildings in some cases. However, now more interests other than the ownership rights are recognised and registered. Land objects other than the conventional land parcel represent these new interests (rights, restrictions and responsibilities). Consequently, the focus is shifting from land parcels to land objects, which require a data model that ensures a common understanding of the relationship between different land objects within a land administration system. Currently, there is no formal data model that exists for either the cadastre or the entire land administration system. The Land Administration Domain Model (LADM) using Unified Modelling language (UML), an object-oriented language would be a starting point for the implementation of nationwide distributed land administration.

3.5.3 3D and Time-enabled cadastres

The proliferation of property interests and sustainable analysis require modelling and visualising of the third and fourth dimensions. As such, the incorporation of height and time dimension into the traditional 2D cadastral frameworks will be essential. The third dimension (height) is already being considered in Kenya in the registration of sectional properties, while time is always included in the cadastral maps and registers and the land registers. Whereas this will continue to be the trend, the inclusion and implementation of these two dimensions in the national land administration model would be very necessary.

3.5.4 Real-time cadastres

The administrations of utilities achieve real-time updates across their networks if there is a connection between computers and GPS units in the field. By the same token, the cadastres that these applications depend on should be up-to date as well. The process of updating cadastre and ownership information though takes a long time, because cadastre and the land register are often not synchronized. A distributed land administration system based on a standardized data model would make it possible for Kenya and other jurisdictions to achieve a real-time update of their cadastres.

3.5.5 Regional and global cadastres

Because of the increased demand for cadastral information and its subsequent integration with other information such as environmental information, there is a trend toward making national and state-based cadastre interoperable. Interoperable cadastral systems appear to offer a platform for integrating and better understanding the relationships between the land markets. Again, the Land Administration Domain Model (LADM) provides a potential foundation for a technical solution. There are currently no efforts in Kenya aimed at creating a cadastral system that is interoperable with its regional neighbours.

3.5.6 Organic cadastres

The traditional land parcels are based on strict bearings and distances or Cartesian coordinates. However, many new property interests are designed around natural phenomena. For example, many interests in the marine environment exhibit fuzzy and changeable boundaries. This generally means that the boundaries can be accurately captured by graphic means without having to be numerically computed. For the Kenyan cadastre, only the positional accuracy of the provisional maps needs to be improved without necessarily having to fix them, as is currently the practice.

3.6 Summary of challenges and suggested solutions

The main challenges that the Kenyan cadastre presents with respect to modern land administration can be summarized as follows.

- The cadastral maps, which constitute the cadastre are still analogue and do not provide complete countrywide coverage.
- The maps are inhomogeneous in terms of coordinate systems and geometric quality; besides, a great percentage of the maps are provisional pending replacement.
- The maps show only land parcel boundaries and roads are not explicitly represented; outlines of buildings are not included in the maps as well.

To create an integrated nationwide digital cadastre that supports as many applications as possible under the SDI framework, the following actions are necessary.

- Ensure complete coverage by extending cadastral mapping and registration to un-mapped areas using appropriate techniques that yield accurate results;
- Digitize the existing analogue maps to create a cadastre that is built on an agreed data model;
- Harmonize the existing maps to a common national coordinate system;
- Improve the geometric quality of the provisional maps using appropriate means.

Some of the suggested solutions might involve significant scientific research to develop an appropriate solution. This dissertation is about geometric quality improvement of provisional graphical cadastral maps. The proposed approach for quality improvement through data integration techniques is presented in the fifth chapter after a review of spatial data integration concepts in the next chapter.

4. Quality Improvement and Spatial Data Integration

4.1 Spatial data quality and uncertainty

Data quality improvement is the underlying goal of any data integration process. Data quality has been defined as a measure of the difference between data and the reality that they represent. It becomes poorer as the data and the corresponding reality diverge. Thus if the data are of poor quality and tell us little about the geographic reality, then they have little value (Devillers and Jeansoulin, 2007).

All geospatial data are often at different levels of detail, imprecise, inaccurate, out-of-date, incomplete, etc. These differences between reality and representation may however be deemed acceptable within certain applications. Data quality is therefore fundamentally about the fitness for use for a particular application. As such data quality is a concept of relativities.

When data quality was first used as a concept, the focused was only on positional uncertainty (accuracy) and was specified in terms of a distance threshold on the printed map, based on the cartographic scale. At the time, the common measures of quality included relative positional accuracy, covariance or correlation.

The concept of data quality has evolved and now quality is reported at a dataset level in terms of the joint quality of features. Quality also varies from one type of feature to another and from one geographic area to another in a given dataset. There is now a general agreement among most standardization bodies about the characteristics describing internal quality of spatial datasets. The common elements include: completeness, logical consistency, spatial accuracy, thematic accuracy and temporal quality and usability (ISO, 2010). The first five elements are sometimes considered as the "famous five" and are well documented (Devillers and Jeansoulin, 2007). The sixth element as included in the ISO standard on data quality describes the suitability of a datasets for a particular application or requirement.

This chapter addresses the problem of quality improvement of cadastral datasets. Although different approaches for quality improvement of different cadastral datasets are identified, data integration is considered as a viable means for improving the geometric quality of approximate cadastres created through graphical means. After highlighting possible data integration challenges when provisional cadastral datasets are involved, the chapter presents an overview of the state of the art in data integration. The chapter concludes with a discussion of the general workflow in spatial data integration.

4.2 Geometric quality improvement of different cadastral datasets

Within many jurisdictions, it is common to find cadastral datasets of different geometric qualities. Therefore, when creating a nationwide cadastral framework referenced to a national coordinate system by integrating the existing cadastral datasets, the nature and quality of these datasets should be considered. The datasets might include numerical cadastres, graphical cadastres, conveyance deeds, and cadastres described by metes and bounds. The possible approaches for geometric quality improvement of land parcel boundaries contained in these different cadastral datasets are presented in Figure 4.1 modified after Blixt (2008). The approaches are described in the subsequent subsections referred to in the figure.

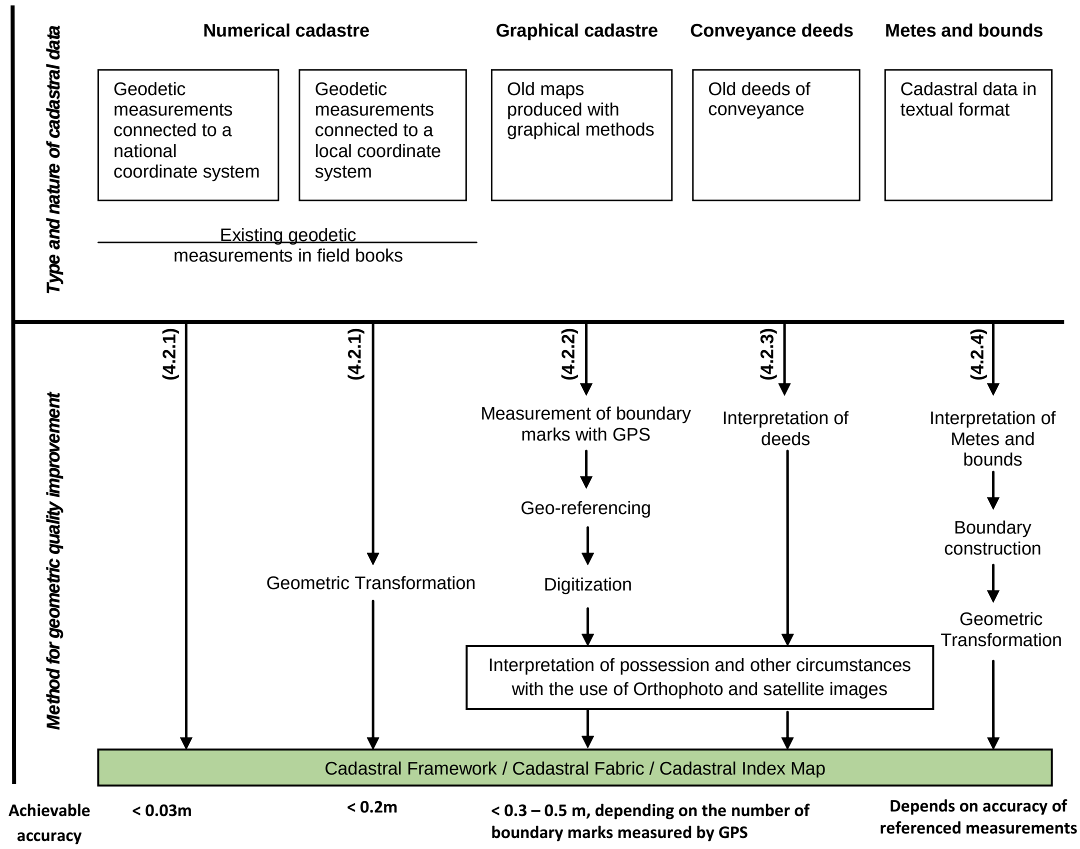


Fig. 4.1: Different approaches for geometric quality improvement of land parcel boundaries contained in different records and maps

4.2.1 Numerical cadastres

The main reason for improving the accuracy of numeric cadastres is if there exist geometric differences with more accurate field measurements obtained using more accurate measuring techniques such as differential GPS. Sometimes, if out of policy requirements, it becomes necessary to replace a local coordinate system or an old coordinate system with a national GPS-based coordinate system, then accuracy improvement of the cadastre would be required. The approach often involves field measurement of boundary marks followed by mathematical adjustment using least squares.

Croitoru and Doytsher (2003) presented an approach for identifying positional inhomogeneities in a numeric cadastral dataset that was graphically represented using survey-accurate field measurements contained in field books. Jarroush and Even-Tzur (2006) and Felus (2007) have used this approach to improve the geometric accuracy of the existing numeric cadastre. If survey-accurate measurements exist in field books, the creation of a cadastral framework dataset will only involve the processing of the measurements followed by mathematical adjustments. However, if field measurements and the maps are spatially referenced a local or a national coordinate system, a necessary process will involve geometric transformation. Felus (2007) employs a local rubber sheeting transformation with linear features. The transformation is based on the affine transformation modelled through Delaunay triangulation. The boundaries are subsequently adjusted incorporating geometric area constraints. The common achievable accuracy is 0.02 - 0.03 m (Boljen, 2009).

4.2.2 Graphical cadastres

Graphical datasets involve not only cadastres captured by graphic means, but also numeric cadastres that are captured from their graphical representation instead of using original field measurements. Although coordinates of the parcel boundaries can be extracted from digitized graphical cadastral datasets, the output datasets cannot however be considered to be numerical. This is because the coordinates of boundaries are affected by various processes including plotting, scanning, geometric transformation and vectorization. Under these circumstances, the geometric quality improvement is a necessary process.

With reference to Figure 4.1, field measurement of boundary marks in the maps and on the ground are made, which are then used for geometric registration. If only such maps referenced to the same coordinate system are involved, then the process simply involves geometric transformations. However, if they are based on different coordinate system the process will additionally involve a geometric transformation between the coordinate systems and the subsequent mathematical adjustments. Accuracies ranging from 0.3 to 0.5 m are achievable (Mikkonen and Corker, 2000; Boljen, 2009).

In countries where graphical cadastral maps have been converted to coordinated cadastre, the process involves carrying out new field measurements and followed by mathematical adjustments to maintain some mathematical conditions typically present in cadastral maps like parallelism and orthogonality. A procedure to remove geometric discrepancies from a graphical cadastre at the same time maintaining the geometric conditions contained in the graphical structures of such maps is presented in Morgenstern and others (1989).

The modernization of a graphical cadastre to a numerical cadastre usually relys on some data integration processes. Steinberg (2000) and Klebanov and Forrai (2010) describe an interesting procedure used in Israel to modernize the national cadastre by implementing a coordinate-based cadastre. In particular, the procedures used in three categories of street blocks were defined according to the background of the cadastral blocks involved (ibid.). These categories of the blocks include blocks with solid cadastral basis, blocks lacking solid cadastral basis, and blocks with mixed cadastral basis. Considering the varying positional accuracy across the original cadastral map, the adjustment of the boundaries is realised through a rubber sheeting transformation. The transformation employs triangulation to subdivide the area into small units and the actual transformation is calculated from certain constraints.

4.2.3 Conveyance deeds

A conveyance deed is a document by which the sale of a parcel of unregistered land is effected. When registered land is involved, the term transfer deed is used instead. There is often a clause in a conveyance deed about the boundaries of the parcel of land involved. The information in the clause refers to a plan that is usually attached to the conveyance deed laying more weight to either the conveyance deed or the plan itself.

Boundary description in the plan attached to a conveyance deed has a number of shortcomings among them: the dimensions of the land parcel are not always quoted; the source of the dimensions is usually never stated. If the dimensions are given at all, neither the directions nor the angles is given. This inadequate description of the boundary makes the interpretation of the conveyance plan difficult.

The common method used to convert a conveyance deed to a map involves its interpretation with the help of a topographical map or an aerial orthophoto or satellite imagery. The positional accuracy of the resulting mapped boundary is however limited by the interpretation and the accuracy of the imagery used. In Sweden, for example, where this has been used, the obtainable accuracy is about 3 m (Blixt, 2008).

4.2.4 Metes and bounds

The description of land parcel boundary by using metes and bounds is still used in countries like Brazil (Mueller, 2008). In metes and bounds, the boundary is described using linear and angular or direction measurements in textual format. Although the description of boundaries by this technique is not based on an actual field survey of the land parcels, reference to registered and surveyed land parcels or their boundaries is used instead to provide a link to spatial reference information. This reference to registered and surveyed land parcels makes it possible to convert the textual description to map representations.

One main problem with metes and bound is that they tend to be lengthy and sometimes the referenced monuments may have disappeared or the registered land parcels referred to in the description may have changed. Moreover, there is possible vagueness in the records (ibid.). The general approach used by survey experts to extract the data and information contained in the metes and bounds involves interpret of the data in addition to field measurements and with reference to other maps. New boundaries are then compiled from coordinates obtained from all these sources. The accuracy of resulting maps interpreted from metes and bounds depends on the measurement technology used at the time of their creation.

4.3 Quality improvement of provisional graphical cadastres

The geometric accuracy of ordinary graphical cadastral maps is comparable to topographical maps at the same geometric scale. However, the same assumption may not hold true for provisional (approximate) graphical cadastral maps even when they are represented at a particular cartographic scale.

Provisional (or approximate) cadastral maps are graphical maps that are produced without fulfilling the positional accuracy requirements of either graphical or numerical cadastres. They are only used on a temporal basis pending their replacement or enhancement. This is similar in concept to what Chrisman (2011) calls the “data effect”, i.e., when the existing data sources are considered the best and influence the kind of analysis and possibly the decision made.

In Kenya, for example, provisional (approximate) maps described in Section 3.3.1 have persisted in usage because of the legacy nature of cadastral maps. However, they are only limited to the originally intended application, i.e., land registration. Another example of the graphical cadastre problem in Turkey is discussed in Demir and Çoruhlu (2008).

The improvement of the geometric quality of these maps is necessary prior to their integration with other geospatial datasets (Gielsdorf et al., 2004). The approaches that can be used to improve the accuracy of provisional graphical maps irrespective of the means used to produce them include replacement with survey-accurate measurements, photogrammetric restitution, digital image processing and data conflation.

Measurements of survey-accurate data of parcel boundaries would provide the most accurate solution. The drawback of this solution is the huge financial and time costs that would be involved. Carrying out survey-accurate field measurements of boundaries in many low income countries is a practically non-realistic solution (Arvanitis and Koukopouluo, 1999).

Accurate photogrammetric restitution techniques can be used to improve the positional quality of poor quality graphical cadastral maps produced by different photogrammetric options discussed in Section 2.4.1. Such a solution is only possible if the original photography used to create the maps is available. If however there are no records of aerial triangulation, the use of such photography could be limited. This is because of the likely lack of corresponding details between the photographs and the cadastral maps due to updates.

An alternative is to use newly acquired aerial photographs or imagery. According to Figure 4.1, one approach would involve superimposing the existing land boundaries on the orthorectified aerial photographs or geo-referenced satellite image and then trace the boundaries afresh. This approach would entail restoring boundaries as-is and measuring the restored boundaries. This is however an impractical solution and would be considered as a new data capture exercise.

Another approach would involve the alignment of the boundaries with respect to the boundary features on the photographs or the imagery using energy minimization techniques such as the snake approach, as used, for example, by Butenuth (2008). Such an approach is best carried out in the context of vector-raster data conflation. Although the approach produces acceptable results where the imagery has homogeneous regions, it could be challenging in imagery that consists of largely inhomogeneous regions. A vector-vector data conflation of provisional graphical cadastral maps with existing higher quality topographical maps provides a viable alternative. An approach for conflation of provisional cadastres to improve their geometric quality, which is the main theme of this dissertation, is presented in the next chapter.

In much of the existing literature about spatial data integration (Saalfeld, 1988; Walter and Fritsch, 1999; Mustière, 2006; and Zhang and Meng, 2007), a significant number of problems involve features in transportation datasets, such as road networks. There is little progress reported with regard to the integration involving administrative and cadastral boundary datasets. This is because cadastral boundary datasets are rarely created more than once as transportation datasets except in special cases (Hope et al., 2008). Another challenge is that administrative and cadastral boundary datasets do not necessarily have corresponding ground signatures in images as transportation datasets. This is further complicated if cadastral maps and records are only approximate and provisional.

The proposed data conflation approach generally adopts the general data conflation workflow presented in the next section with the main focus on geometric aspects. The main departure and one of the contributions of this research is in the pre-integration and the integration processes. These involve the derivation of the road network from the cadastral dataset and the local geometric transformation in the pre-processing and integration steps of the workflow. The integration process considers only the geometric criteria.

4.4 Interoperability and data integration

Sharing and provision of geospatial data and services through the Internet presumes that the services are interoperable and the datasets are integrable. However, because of the large heterogeneities of the computer systems and geospatial datasets, interoperability and data integration are still significant issues in SDIs.

Interoperability is defined as the capability to communicate, execute application programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units (OGC, 2002). Interoperability facilitates the integration of data from different organizations and across applications and industries, resulting in the generation and sharing of information that is more useful. To ensure interoperability during data integration, structural, semantic and geometric differences in the datasets have to be taken into account (Sester et al., 2007). It is because of these differences that data integration still remains a challenge.

4.5 Basic concepts in spatial data integration

Uitermark (2005) describes data integration as a process used to establish relationships among corresponding object instances in different, autonomously produced spatial datasets of the same geographic space. However, data integration can be carried out with different goals in mind. These include enhancing and improving the overall quality and reliability of a dataset (e.g., when updating) (Saalfeld, 1988), transferring attributes from one feature geometry to another one. Other goals include automatically registering one spatial dataset to another through the recognition of common features, and when an integrated analysis is intended (Butenuth et al., 2007). Considering these objectives, spatial data integration can be comprehensively regarded as the process of assuring consistency among various data elements in terms of geometric, topologic and attribute accuracy and precision for the purpose of analysis, sharing and for enhancing the information content and quality of the final output.

Other terms used to describe the same or related operations include data conflation, data fusion, merging, and harmonization. Although these terms are sometimes used interchangeably, there is however, a subtle difference among them depending on type of datasets and operations involved. Data integration or combination is used if the output of the operation is a separable type of output, where individual characteristics of each of the input elements are preserved. An example is the superimposition of vector data over imagery (Savopol and Armenakis, 2002). The terms data conflation, merging or fusion are more appropriate if the result is a single composite dataset obtained from the combined elements and the individual characteristics of the input elements are not necessarily preserved. Data conflation is commonly used with vector data, while data fusion is commonly used with imagery, particularly when combining remotely sensed data with different resolutions, for example, in pan-sharpening (Vijayaraj, 2004).

Data harmonization is yet another term that is used in relation to the integration of data. It is involves developing a common set of specifications for data products in a way that allows datasets to be combined in a coherent way (Fichtinger, 2009). This includes agreements about coordinate reference systems, classification systems, application schemas, etc. Data harmonization emerged because of the need to create consistent framework datasets.

A detailed review of the digital map conflation process and a proposed classification are presented in Ruiz and others (2011). Accordingly, the classification may be based on the matching criteria used (Casado, 2006), the categorization problem (vertical, horizontal or temporal) (Yuan and Tao, 1999), the representation model (vector or raster), as well as the level of automation (manual or automatic).

Data integration may involve only vector or raster datasets or a mixture of both. The integration of raster-to-raster datasets may include, for example, images, raster maps and raster grid (e.g., the integration of Digital Terrain Models (DTM)). The main consideration when integrating different raster datasets is the differences in spatial resolutions.

Integration involving vector datasets should consider the differences in the geometric scales and the relative geometric accuracies. If the scale of output is the same as the scale of one of the input datasets, then the result is data assimilation, otherwise the output of the result is a new fused product.

Integration of vector and raster datasets can simply be achieved through the superimposition of vector datasets on raster data. Tasks that are more complex would usually involve more than one digital image-processing techniques. Some of these techniques include, for example, feature extraction, re-sampling, and model recognition, in addition to geometric transformations. The main challenge in ensuring effective vector-to-raster conflation is establishing the optimal ratio between the scales of the vector data and the spatial resolution of the image and transformation of various data formats. An example of an approach for the integration of vector and raster data is the snake algorithm (Agouris et al., 2001) based on the principle of energy minimization as proposed by Kass and others (1988).

Spatial data integration can also be categorized as either vertical or horizontal. Vertical integration entails elimination of position and attribute-differences by virtue of existing in the same spatial extents. Data conflation is commonly used to refer to vertical data integration (Longley et al., 2005). Horizontal integration aims to eliminate spatial feature position and attribute discrepancies, which exist only in the common area of overlap. This process is also referred to as edge matching.

4.6 Heterogeneities in spatial data

Heterogeneities (differences) between geospatial datasets depend on the applications they have been acquired for, the level of detail (scale), and on the acquisition itself. Different data modelling schemes may lead, for instance, to different aggregation levels of objects and thus different cardinalities between object instances. An example is a set of lakes in one dataset, which could be modelled as an aggregated lake in another dataset. If the datasets originate from different geometric scales, more severe differences between them can occur, for example, large geometric and topologic differences (Sester et al., 2007).

Villa (2007) presents a list of the most common data heterogeneities that necessitate data integration and harmonization. They include differences in data formats, spatial reference systems, and conceptual schemas, classification schemes (e.g. different ways to classify land cover or flood risk-warning levels). Additional heterogeneities include differences in geometric scales or spatial resolutions of vector or raster datasets respectively, levels-of-detail, terminology, and multiple representations of the "same" spatial objects, and spatial inconsistency at borders.

In general, the heterogeneities are introduced at different stages during the data modelling and acquisition process and each particular type depends on the stage and domain of data modelling involved. Figure 4.2 is an attempt to associate the heterogeneities with the corresponding data modelling stages. This illustration is motivated by the frequent classification of the heterogeneities into three common categories related to the general data modelling process (conceptual, logical and physical). They include semantic (differences in the intended meaning of terms in specific contexts), schematic (differences in schemas) and structural and syntactic (i.e., data-level differences and differences in data formats) heterogeneities.

4.6.1 Semantic heterogeneity

The inevitable differences in perceptions during data modelling usually result in different conceptual data models, which are manifested by various semantic heterogeneities. Semantic heterogeneity occurs because of disagreements in meaning, interpretation and intended use (Xu and Lee, 2002). In general, the sources of semantic heterogeneity include conceptualisation differences, differences in formalization, and differences in the assumed context (i.e., contextual heterogeneity).

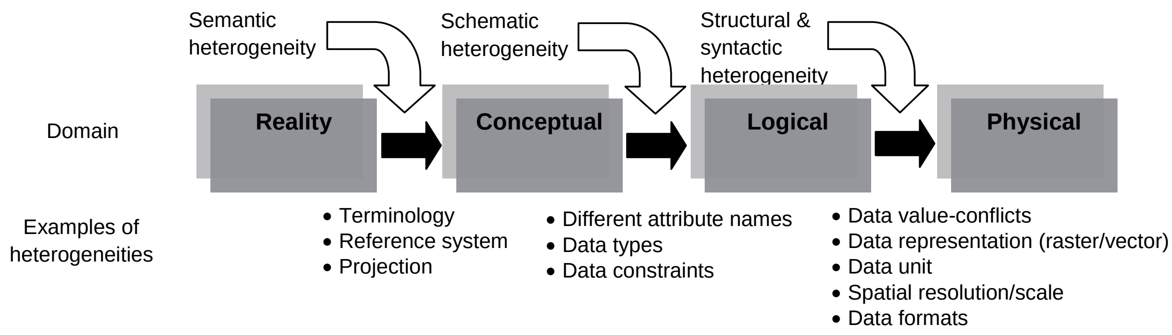


Fig. 4.2: Data heterogeneities associated with different levels of data modelling

Conceptual heterogeneity is due to the fundamental differences in modelling the real world. It means that the data may not be readily suitable for a different application. Formalization heterogeneity results when different data languages are used when formalising a conceptualization. The data languages may be of different semantic richness. For example, in representing and processing geographic information, many GIS software packages with varying sophistication have been developed, all of which have their own data language. Another aspect of formalization heterogeneity is formalizing a concept either in one or more languages. Contextual heterogeneity is due to differences in the implicit assumptions about the conditions or environment upon which data modelling is based. For geospatial data, the measurement unit, scale, reference system and map projection used when creating the data are among the most significant contextual factors.

It is harder to achieve semantic interoperability because usually datasets from different domains and communities are involved. A basic approach, which can be used to solve semantic differences in datasets, is one that uses geometric relations to infer semantic relations. This approach, which has been used by Volz (2005) to link linear data, is based on the idea that if two objects have an identical name or coincide geometrically, then they probably have something common on the semantic level. Semantic heterogeneities are commonly solved using a thesaurus or ontology. A thesaurus comprises of a collection of significant terms and the relationships between those terms in a domain knowledge space. A thesaurus has explicit concepts, relations, terminology and definitions (explicit specification) and reflects a consensual abstracted knowledge model for a domain (Deliiska, 2007).

Ontology is a logical theory accounting for the intended meaning of a formal vocabulary. It is defined as a formal, explicit specification of a shared conceptualisation (Gruber, 1993). It usually describes a domain of interest and a specialization of the meaning of terms used in the vocabulary. It encompasses, for example, sets of terms, classification, database schemas (Vaccari et al., 2009). Hence, ontology is essentially a machine-readable (formal) thesaurus. Therefore, ontology allows sophisticated information processing, such as inferred properties, based on logical reasoning. For this reason, ontology is preferred over a thesauri approach. However, these are not competing techniques as ontology is a richer conceptualization of a thesaurus. Both ontology and thesaurus define a common vocabulary to reduce the semantic problems however, neither in itself guarantees complete interoperability.

4.6.2 Schematic heterogeneity

Schematic heterogeneity is introduced when a conceptual data model is expressed in different ways in a database during logical modelling. Schematic heterogeneities result from the differences in logical structures and inconsistencies in data in the same domain used in different databases. Two basic examples include (i) the use of different logical structures (tables and attributes) for the same information, and (ii) the use of different specifications (e.g., names, data types or constraints, precision) for the same data model (Park, 2001). Schematic heterogeneity is mediated mainly through query and transformation languages.

4.6.3 Structural and syntactic heterogeneity

Structural heterogeneity also referred to as data-level heterogeneity is introduced during physical modelling in which similar schemas are implemented using different data structures. Data-level heterogeneities result from the differences in data domains caused by the multiple representations and interpretations of the semantically same data. Data-level heterogeneities includes data value conflicts, data representation conflicts, data unit conflicts, data precision conflicts including data granularity and spatial resolution, data accuracy conflicts, and spatial domain conflicts (ibid). Structural heterogeneities are mediated through various data-level transformation processes.

Syntactic heterogeneity is introduced if different schemas that use similar data structures are based on different file format specifications. Syntactic heterogeneity is manifested by the use of different data encodings or formats. Syntactic heterogeneities are resolved using format conversion tools or by using standardized data formats (e.g., ISO and OGC) and it is usually the most common type of heterogeneity and common level at which the bulk of data integration is carried out.

Although the data heterogeneities discussed above are of a technical nature, they are associated with non-technical issues categorized as institutional, legal, political and social (Mohammadi, 2009).

- institutional - different coordination and maintenance mechanisms;
- legal - different license conditions and liability regimes;
- political - access policies, use restrictions and pricing models;
- social - aversion against data sharing, integration and silo mentality.

4.7 Spatial data integration workflow

Data quality is achieved by resolving different heterogeneities and inaccuracies present in spatial datasets. This is often realised through various data integration and harmonization mechanisms. The mechanisms are based on either ad-hoc or standard-workflows. Due to the complexity of the problem of data integration, it is nearly impossible to develop a system that solves all the problems. Instead, various methods exist for different data integration problems. Any system would however be similar in some respects to the generic workflow in the data integration by Shereen and others (2004) as presented in Figure 4.3. The spatial workflow in data integration relate to the three classical phases defined for general databases. They include pre-integration, investigation of correspondences and actual integration (Parent and Spaccapietra, 2000, Song et al., 2006). These three phases of the generic framework for data integration can be carried out either at schema- or data-level or both.

4.7.1 Pre-integration

In the pre-integration step, the input schemas and data are arranged in various ways to make them more semantically, geometrically and syntactically homogeneous. The step involves a detailed study

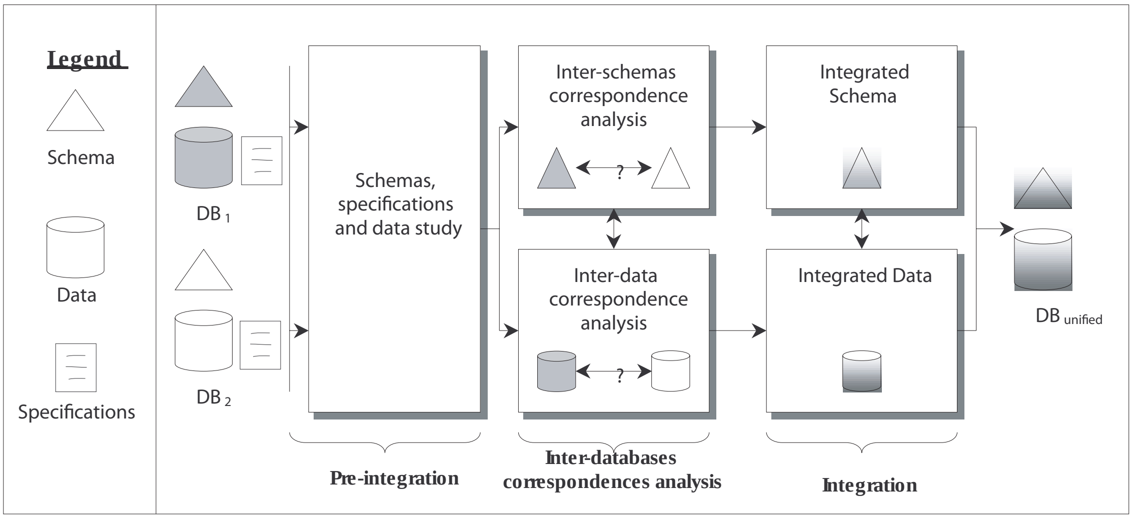


Fig. 4.3: Generic framework for spatial data integration

of specifications, schema and the data itself (Devogele and Spaccapietra, 1998; Sheeren et al., 2004). The tasks in this step consist of the analysis of database specifications at the schema- and data-level. At the schema-level, the schemas are rearranged to make them more semantically and syntactically homogenous. At the data-level, the geometries of the data features are harmonized to bring them to the same geometric dimension, scale and map projection for the subsequent feature matching. In the proposed approach, pre-integration involves preliminary registration and derivation of road network from cadastral datasets.

4.7.2 Correspondence investigation

The basic issue in this step concerns the identification of the correspondence (or matching) between the elements in the datasets. This phase involves the identification of related items in the input schemas and data, and the precise description of these inter-schemas relationships. The investigation of correspondences is carried out either at the schema- or feature-level. At the schema-level, the task involves the identification of related schema elements. Feature-level data correspondence involves the identification and assessment of consistency between the corresponding spatial representations with spatial data matching algorithms. It is necessary to detect and eliminate any errors before the feature matching.

Many feature matching methods developed in the literature yield good results and efficiency on certain data types. The matching algorithms are classified based on three criteria: geometric, topological and attribute (Yuan and Tao, 1999). A combination of these criteria is possible, for example, Mustière (2006) suggests the simultaneous use of semantic, geometric and topologic information in automated matching. The proposed approach employs mainly the geometric and topologic criteria.

Geometric criteria

The corresponding features in two vector datasets is based on the similarity of the features at the data-level and based on the relations between the elements in the datasets. The methods involve scanning the geometric objects from both datasets and comparing them by geometric criteria. Different criteria exist for the different geometric features: point, polyline, polygon and networks. For example, for point matching, the Euclidean distance is suitable, while for line matching, the

Fréchet discrete distance (Devogele, 2002) Hausdorff distance (Hangouët, 1995) is much better (Yuan and Tao, 1999). Other geometric criteria include angular information of linear objects, location relationship, and shape of object features.

Angle between linear objects is used mainly to match linear objects. Location relationships, for example, the point in polygon criteria can be used to match centroids and polygons. The shape of features, which include properties such as length, concavity, perimeter, area, density, Fourier shape descriptor, etc., can be used to match polylines or polygons. These properties are used in a statistical approach developed by Walter and Fritsch (1999). The approach is similar to relational matching method first proposed Vosselman (1992). Relational matching has been widely used in computer vision for the recognition and location of objects in digital images, in which the objects are represented by structural descriptions.

Topological criteria

Topological criteria are seldom used alone, they are used only when topological information is available, and otherwise a topology-building component is needed. Topological information includes, for instance, connectivity between lines, adjacency between polygon and composition relationship such as outlets of nodes, arcs that form a polygon and so on. Sometimes topology can be used as a main criterion, based on some known matched objects, or aided by a geometrical method. Topological criteria are appropriate if there is a high topological similarity between the datasets.

Attribute criteria

The attribute criteria can be used to match features very efficiently if both data sets have a common attribute field and the semantics of both data sets are known, so the relationship of these fields can be determined. The simplest case is when two datasets have the same attribute field and the same definition of the field. An approach for semantic data integration that exploits attributes is presented in Kieler and others (2007). Rule-based feature matching as an attribute-based approach (Cobb et al., 1998) has also been used to achieve semantic data interoperability. The approach is based on shape similarity of lines as well as the semantic similarities of non-spatial attribute values.

4.7.3 Integration

Integration (or alignment) is the final step in the workflow. It entails unifying the corresponding items in an integrated schema and the associated mappings. With the objective of integration in mind, the basic task involves the transformation of the subject data to the reference dataset both semantically, schematically or geometrically.

If the aim of integration is to create a dataset with richer content and improved spatial accuracy, the basic tasks will involve both attributes transfer and geometric alignment. Attribute transfer is trivial if there is a one-to-one correspondence between features in the datasets, such ideal cases are however rare. Instead, many challenging scenarios could exist, which may entail one-to-many, many-to-one or many-to-many feature correspondences.

At the schema level, the main goal is to conform a given schema specification to a particular prescriptive specification. The data instances are either semantically or geometrically transformed, based on a new schema defined for the unified dataset. The semantic or geometric alignment of the two datasets alignment is achieved through various schema and geometric transformations. The proposed approach deals only with geometric transformation.

5. Conflation of Provisional Cadastral Datasets

5.1 Data conflation of graphical cadastres

This chapter presents a vector-based data conflation approach for enhancing the geometric quality of approximate cadastres created by graphical means. The approach consisting of different steps is presented in form of a workflow. All the underlying assumptions, variables and limitations in each step are discussed.

The general idea of the approach is that because reference cadastral datasets rarely exist, topographical datasets would constitute a good reference for accuracy enhancement of approximate cadastral maps. The use of a topographical dataset is justified because the majority of geospatial datasets consists of topographical features and usually have corresponding ground signatures if the datasets are up-to-date.

The main challenge involving the conflation of cadastral maps with topographical maps is the possible absence of corresponding features in the two maps (datasets). This is particularly the case if building footprints and publicly owned land such as roads are not explicitly included in the cadastral maps. Derivation of road network from cadastral datasets will therefore be necessary to obtain features that will correspond with those obtained from the corresponding topographical datasets.

If the original cadastral maps are also not referenced to any coordinate system, geometric registration of the maps might be necessary before carrying out the feature matching process. If in addition the cadastral datasets contain local geometric distortions, a non-rigid transformation would be necessary in the final quality enhancement process.

Figure 5.1 shows the steps of the proposed workflow for geometric accuracy improvement of provisional cadastral maps. The approach is applicable to any approximate graphical cadastres irrespective of whether the data lineage is known. The workflow consists of four steps: pre-processing, derivation of road centrelines from a cadastral map, feature matching, and finally, a non-rigid geometric transformation. Each of these steps is described in greater detail in the successive sections.

5.2 Pre-processing

This step is intended specifically for approximate analogue cadastral maps that do not have a coordinate grid except descriptive location information. Pre-processing therefore entails scanning, approximate geometric registration and vectorization of the land parcel outlines.

Approximate geometric registration can be carried out based on manually identified corresponding positions in the cadastral map (subject) and the reference topographical dataset. A global geometric transformation is just sufficient. Vectorization of the land parcel outlines can also be carried out manually or semi-automatically if possible. It is assumed that from this step onward, the cadastral (i.e., the subject) and the reference topographical datasets are in the same data format, are approximately aligned and in the same coordinate system.

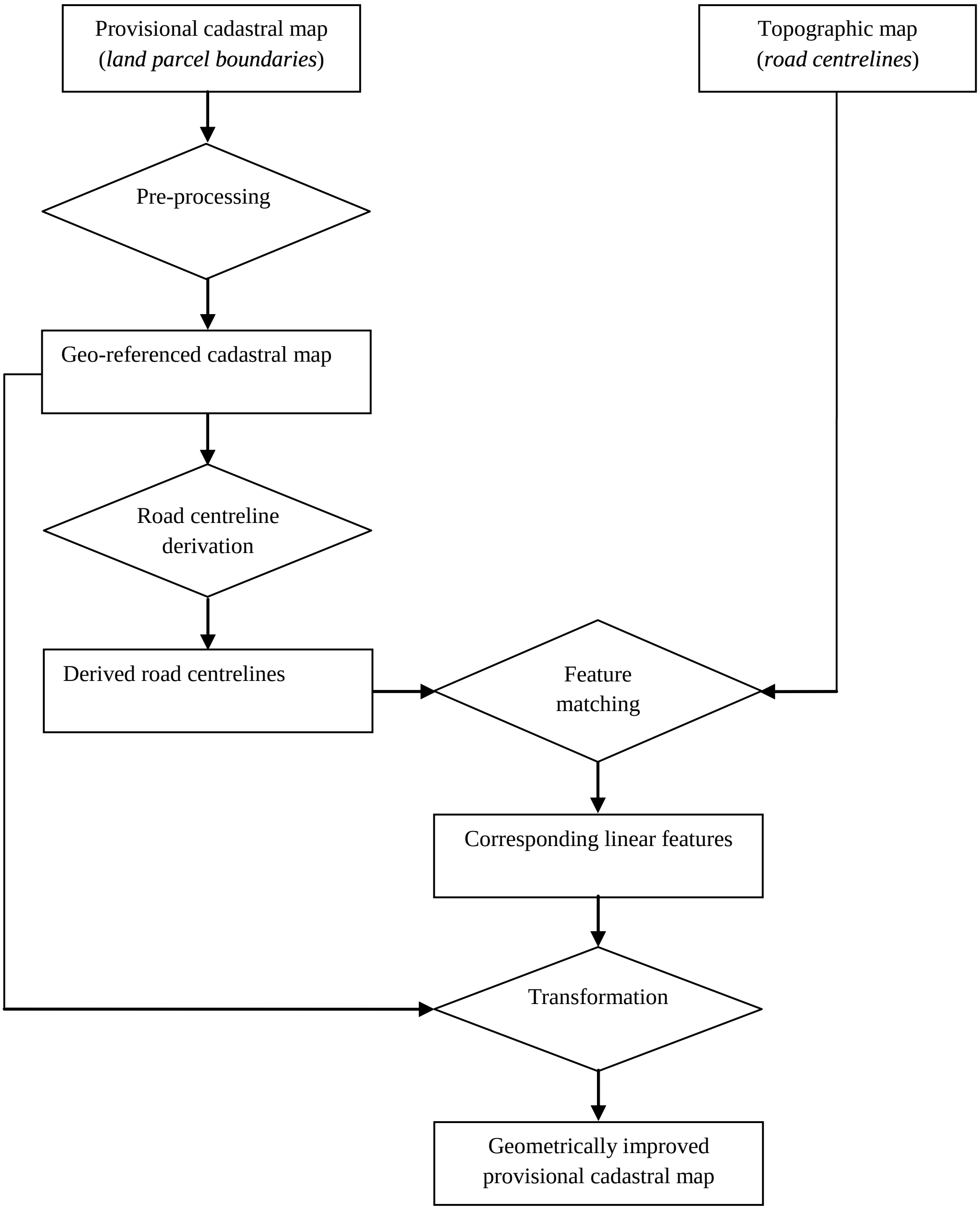


Fig. 5.1: Process workflow for geometric quality enhancement of provisonal cadastral datasets acquired by graphic means

5.3 Derivation of a road network from a cadastral dataset

In countries with either purely fiscal or purely legal cadastres, it is a common practice to find only privately-owned land parcels included in the cadastre. In such cadastres, publicly-owned lands such as road and riparian reserves are only implicitly represented by gaps between the polygons that represent these privately-owned land parcels. Countries with multipurpose cadastres however, have all private and public land included in the cadastre. In fact, road and riparian reserves are explicitly represented as parcel polygons. The cadastral template project (Steudler, 2010) contains sample cadastral maps from different countries, where some do not explicitly contain roads as land parcels.

This step is included in the workflow considering cadastres that do not explicitly contain road and river reserves as land parcels. The main assumption in this step is that the empty spaces between the parcels represent road parcels. The step entails two processes. The first process involves the extraction of the polygons representing road (and also riparian) reserves from the cadastre as road polygons. This process exploits the concepts of the concave hull and linear referencing. The second process involves the derivation of road network centrelines from the road polygons using straight skeletons.

5.3.1 Road polygons from cadastral datasets

Different techniques can be used to derive road polygons from cadastral datasets if they are not explicitly represented. One such technique involves geometric buffering of the input land parcel polygons. The road polygons are then obtained from the geometric difference between positive and the negative geometric buffers of the input parcel polygons. The main shortcoming of this technique is the creation of the unwanted spurious polygons at the periphery of the outline of the input land parcel polygons.

Another technique involves the identification of all the outer vertices of all exterior polygons and re-sampling them to create one combined polygon. The geometric difference between this new polygon and the input polygons would result in the intended road polygons. Identifying all the outer vertices of all the outer polygons is however not trivial. In this thesis, the derivation of the road polygons is achieved through the use of a combination of concave hull and linear referencing concepts. The concave hull is used to identify the outer polygons and the creation of a combined polygon with the same spatial extent as the input polygons. Linear referencing is used to identify the vertices of the identified outer polygons. The next two subsections discuss the concepts of the concave hull and linear referencing.

5.3.2 Concave hull and linear referencing

Geometry Containers

The Minimum Bounding Rectangle (MBR) and the convex hull are perhaps the most basic and common geometry containers that are used in computational geometry. The convex hull has been applied in many fields, for example, business, engineering, science, daily life and so on. The convex hull is used in particular when the only objective is to minimize the outline length. However, if the objective is to minimize the area of the hull, then a non-convex hull (or better a concave hull) would be most appropriate.

Concave hull

A concave hull is a concave polygon that encloses all geometries within a set, but has less area compared to the convex hull. Because of minimizing the area, the concave hull's line length is longer than the corresponding convex hull. A concave hull could be suitable for some real-world problems, for example, finding the boundary of a city based on the amalgamation of the land parcel boundaries.

Computing the concave hull is considered one of the complicated problems in geometry, and as such, there are many variations of it (Sunday, 2006), which can be used depending on the intended application. However, there are currently no algorithmic fundamentals that exist for the creation of a concave hull. This is because the algorithms for concave hulls are much more complicated than convex hulls. Moreover, for any given set of points, there may be lots of different concave hulls. In this regard, we present an approach for the creation of a concave polygonal hull based on the concept of linear referencing. The approach was motivated by the need to extract the road polygons implicitly represented by contiguous spaces in a cadastral dataset. After a review of the few algorithmic efforts to construct a concave hull, some theory on linear referencing is presented followed by a discussion on the algorithmic implementation of the concave hull.

The concave hull approach is a more advanced approach used to capture the exact shape of the surface of a features contained in a dataset. However, producing the concave hull is difficult; this is because of several possible and often conflicting objectives. Little work has focused on concave hull algorithms.

Galton and Duckham (2006) suggested “Swing Arm” algorithm based on gift-wrapping algorithm. In the “Swing Arm” algorithm, the polygon hull is generated by a sequence of swings of a line segment of some constant length, r (the swing arm). The initial line segment is anchored at an external point, and at each subsequent step, the line segment is anchored to the last point added to the hull, and rotated clockwise until it hits another point in the hull. If r is not less than the longest side of the hull perimeter, then the procedure will generate the convex hull; but if r is shorter, the resulting polygon is concave. The Swing Arm Algorithm may produce separated concave hulls instead of single one, a situation that is not desirable especially for the current problem.

Another approach is based on alpha shapes, first described by Edelsbrunner (1981). Alpha shapes are considered as a generalization of the convex hull and a sub-graph of the Delaunay triangulation. They can be used in place of simple convex hulls to create a polygonal boundary containing the geometric objects within it. Mathematically, alpha shapes are defined as a family of shapes that can be derived from the Delaunay triangulation of a given point set with some real parameter, α controlling the desired level of detail. For sufficiently large alpha, the alpha shape is identical to the convex hull, while for sufficiently small alpha, the alpha shape is empty. As such, the resulting shape is neither necessarily convex nor necessarily connected. Alpha shapes maybe good but sometimes they are not flexible enough because the alpha parameter is fixed. The “α-shape” algorithm based on Delaunay triangle suggested by Duckham et al. (2008) is similar to the concept of alpha shapes and has the same weaknesses.

Adriano and Yasmina (2007) suggested a concave hull algorithm based on the k-nearest neighbours approach. The algorithm although fundamentally designed for a set of points, can be used for other geometry primitives. The undesirable feature of the algorithm is that holes are produced in the resulting concave hull even when they are not expected.

In the approach presented in this thesis, the concept of using linear referencing to create a concave hull of a set of polygon features is described in 5.3.3. The process begins with a convex hull of the features. This is then iteratively modified using a set of points. The set of points is iteratively populated by any vertex of the input polygon whose Thiessen polygon intersects the boundary of the modified convex hull. During the iteration, a new boundary (concave hull) is created from the points in the point set, with the previous hull taken as the reference linear feature.

Linear referencing

The term linear referencing, which is also technically referred to as dynamic segmentation, emerged from engineering applications where it was preferable to locate a point relative to a linear feature (often roads). Linear referencing uses a linear feature as a reference instead of the typical geographical coordinate systems. The most familiar illustration of linear referencing is the mileage markers along roads and highways. Linear referencing is also similar to the method of booking measurements in the traditional method of chain surveying. An example of a linear reference address with reference to an engineering station would be 1000 + 56. This simply means that the station is 1000 divisions (with each division being 100 m) plus 56 m i.e., 100,056 m from the start of the reference linear feature.

Linear referencing differs from the traditional geographical coordinate and reference systems because the underlying entity used as a basis for measurement is not the earth, but a linear feature or a set of linear features organized into a network (Curtin et al., 2007). As a consequence of using any linear feature as a reference, there can be as many linear referencing systems as there are coordinate systems.

A linear referencing system (LRS) is a mechanism for finding and stating the location of an unknown point along a network by referencing it to a known point (Vonderohe et al., 1997). The process entails measuring the distance along the path of an underlying linear feature and an offset from the path to the feature to be referenced.

Linear referencing can be used in applications that deal with linear features, such as road-management organizations, transit organizations, oil and gas exploration industries, and water-resources management agencies among others. All these organizations deal with linear features and sometimes need to reference a position or measurement along those features.

One of the benefits of using linear referencing is that locations specified with linear referencing can be readily recovered in the field and are generally more intuitive than locations specified with typical geographical coordinates. Secondly, the requirement to highly segment a linear feature based on differences in attributes values is avoided. As a result of this, a network database with many different attribute events associated with a single and reasonably small set of network features can be maintained thus reducing the redundancy and potential error within the database. In addition, it facilitates multiple cartographic representations of network attribute data (Curtin et al., 2007).

The ability to reference events to linear features was identified as an essential functionality within GIS by Nyerges (1990). Several current GIS software packages offer tools for linear referencing, although each has its specific procedure. Linear referencing as a process consists of a number of steps. The typical steps for a linear referencing are as follows:

- Identifying the underlying linear feature (or the route structure) to which events can be referenced;
- Defining and identifying measurements along the identified route (linear feature);
- Output of linearly referenced events.

5.3.3 Concave hull through linear referencing

To create a concave hull of a given set of polygons, the fundamental concept is that the concave hull is the polygon with the minimum possible area that encloses all the polygons. Assuming that some of the polygons in the set have contiguous empty spaces between them, then the minimum polygon that will enclose the polygons is one that contains the outlines of all the outer polygons plus the empty spaces. We sample these outlines by a combination of the Thiessen polygons and linear referencing. Our approach follows the typical steps for linear referencing presented in the previous sub-section.

a) Identify an underlying linear feature to which events can be referenced

The first step in linear referencing is to define the reference linear feature or network. However, a dataset consisting only of a set of polygons does not consist of a linear network or linear features. The representative linear features from such dataset would include extracting the outlines of all the polygons. In this technique, such an outline, as the initial reference linear feature, is approximated by the convex hull of the polygons. Figure 5.2 illustrates the outline of the convex hull that represents the initial reference linear feature (in black). Its direction, which is important in the referencing process is noted and is depicted by the arrow in the figure.

b) Defining and making measurements along the linear feature

The event data (points) to be referenced and the direction of measurement are identified. The points to be referenced should include all the vertices of all the outer polygons. This is achieved by identifying the points (vertices) whose Thiessen polygon intersects with the reference linear feature, initially approximated by the the convex hull. The Thiessen polygons are used as a means to identify the points that will eventually form the concave hull because each Thiessen polygon technically represents one individual point and for all points of interest, their Thiessen polygons should definitely overlap (intersect) with required concave hull. Moreover, a unique concave hull is possible if there are no extreme concavities in set of input polygons.

Figure 5.2 illustrates the concept of the concave hull using linear referencing. The input polygons representing the land parcels are shaded in grey, the vertices of all the polygons are represented by the grey dots. The outlines of the Thiessen polygons are shown by the dark lines. The initial linear reference feature (and consequently the initial concave hull) of the input polygons is approximated by the convex hull and is represented by the red outline. The direction of the linear reference feature deciphered from the direction of the line string of convex hull is shown by the arrow.

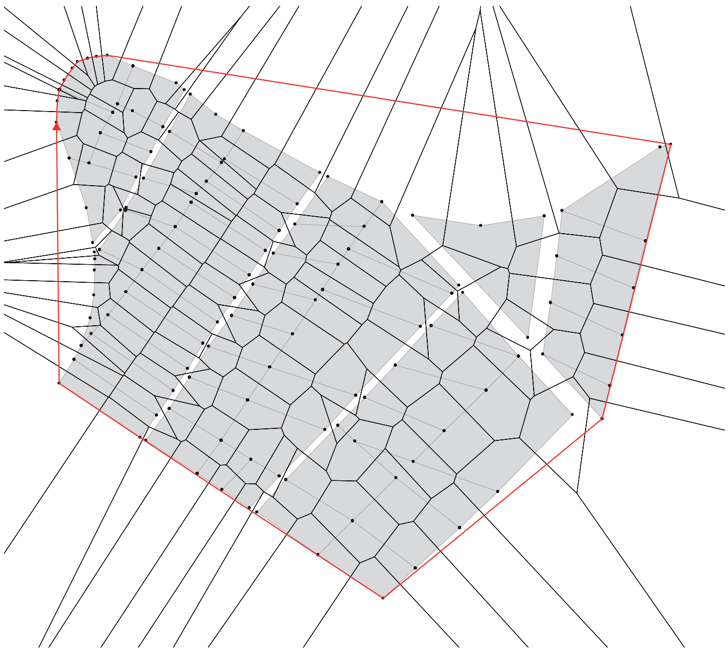


Fig. 5.2: Initial approximation of the concave hull by the convex hull (red outline). The input polygons are shaded in grey, the polygons vertices are represented by the dark dots and the outlines of the corresponding Thiessen polygons are shown by the dark lines

After the first iteration, points whose Thiessen polygons intersect with the convex hull (red outline in Figure 5.2) are identified. These points are then referenced to the reference linear feature, the outline of the convex hull represented by the red outline in the figure. The starting and end point (and therefore its direction) of the reference linear feature are depicted by the red arrow. Linear referencing is done as follows: for each identified point, its distance from the start of the reference linear feature and its offset from the reference linear feature are calculated and this constitutes the reference information. Then, a new reference linear feature is created, from all the identified points (red dots). This new reference feature is used to create the outline of a new concave hull, which effectively replaces the earlier one. This procedure is done iteratively until no more points can be identified and the concave hull cannot be modified any further. Figure 5.3 shows the points (red dots) of the intended concave hull identified after the first iteration.

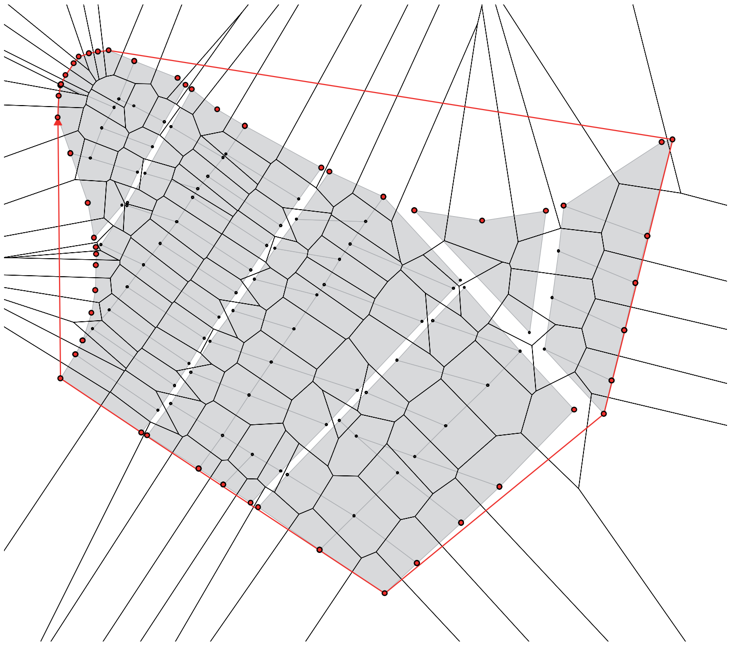


Fig. 5.3: Identified points (red dots) of the concave hull (in Figure 5.4)

c) Output of linearly referenced events

The linearly referenced points are consecutively constituted into a linear ring from which the concave hull is build. Figure 5.4 shows the final concave hull (red outline) created from the points identified after two iterations. The initial concave hull approximated by the convex hull is depicted by the red broken outline. Once the final concave hull has been obtained, no additional iteration could change the result. To ensure that final and unique concave hull is obtained, the number of iterations is set as high as possible. Based on experiments, twenty (20) iterations are just enough. If there are no extreme concavities, even two iterations are enough.

This approach produces a unique concave hull as long the points along the outlines are evenly spaced, and if the expected outlines of the input geometry are not very indented. Figure 5.5 (left) shows an incorrect section of the concave hull (in red) for a case with uneven distribution of vertices along the outlines of the input polygons. The land parcels are illustrated in gray polygons with the black dots representing the vertices. The Thiessen polygons are depicted by grey lines. The initial position for part of the hull is represented by the convex hull (in bold black outline). It is evident in Figure 5.5 (right) that the hull is coincident with the outlines of the outer polygons when the more regularly spaced vertices are introduced. Using constrained Delaunay triangulation

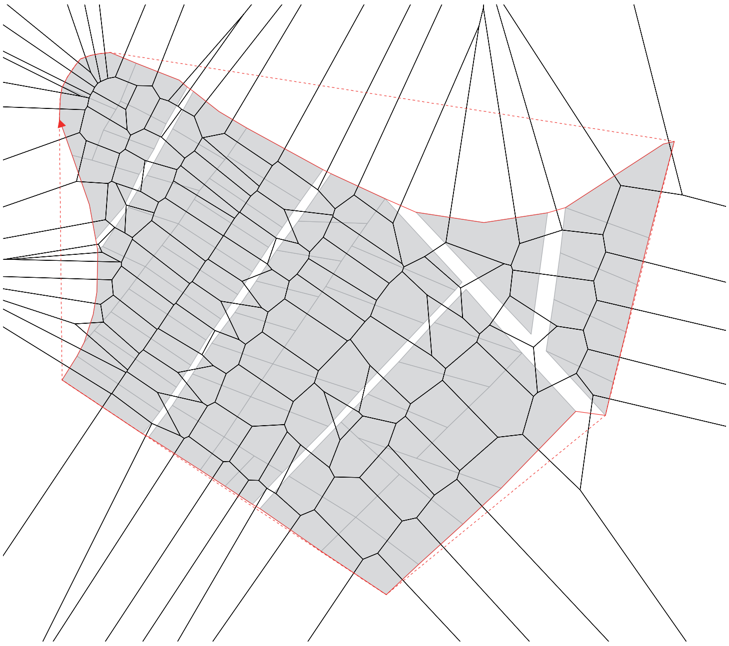


Fig. 5.4: Final concave hull (red outline) and the initial concave hull approximated by the convex hull (in broken red outline)

should overcome the intersection problem of the concave hull and the input polygons as illustrated in Figure 5.5 (right).

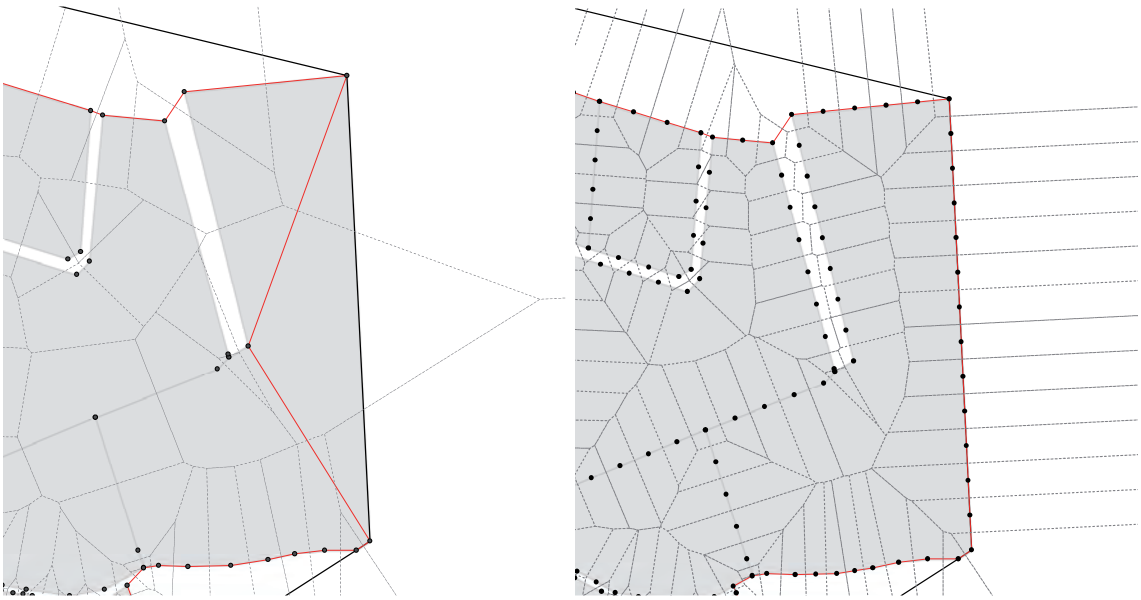


Fig. 5.5: Resulting concave hull (in red) due to uneven distribution of vertices on the left and the hull after introducing additional vertices on the right

Another challenge to the concave hull approach is when the outline of the input polygons is very indented. This is due to the fact that not all Thiessen polygons which extend outside the outer input polygons intersect with the successive evolution of the concave hull. This challenge is remedied sometimes by the introduction of additional vertices. Figure 5.6 (left) shows the situation where the concave hull does not coincide with the outline of the outer polygons due to concavity. Figure 5.6 (right) shows that a perfect coincidence in the lower section while no coincidence is achieved in the upper section. This happens when there is parallelism between a given segment of the concave hull at a particular iteration and any one side of a Thiessen polygon that intersects with the segment.

The figure also shows that, sometimes the addition of more vertices does not completely solve the concavity problem because of complex concavity. Again, the introduction of additional vertices is only possible if the input geometry consists of either polylines or polygons.

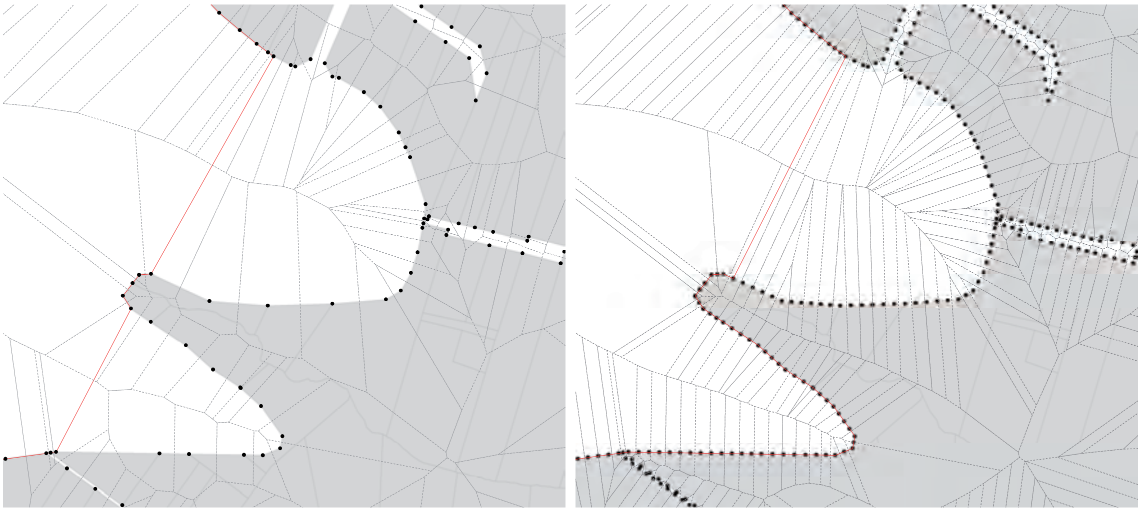


Fig. 5.6: Failure of the concave hull to coincide with the outline of the input polygons due to extreme polygon indention (left) and coincidence is achieved after additional vertices are introduced

The road polygons are then obtained as a geometric difference between the input polygons and the final concave hull. The resulting road polygons should indeed represent road reserves if the input parcel polygons represented the complete parcel coverage. Caution should however be exercised if the input polygons do not have a complete coverage, in which case some of the empty spaces would be false gaps. It should also be determined at the onset if riparian reserves are also mapped as land parcels or left out just like the road reserves.

5.3.4 Creation of road centerline network through straight skeletons

The geometry difference between the polygons representing the land parcels and the concave hull consists of the road polygons, i.e., polygons representing the roads parcels. A road network is then created from the road polygons. The derivation of road centrelines from a road polygon dataset can be achieved through either raster-based or vector-based techniques. In the raster-based derivation, the road polygons are first converted to raster and then the generation of the road centrelines is carried out in raster domain. Because of the potential introduction of distortions during the thinning and the subjective smoothing procedures in the raster approach, the vector-based approach was considered instead. The vector-based road network derivation is based on straight skeletons algorithm. This is an area collapse method based on straight skeletons generated via Delaunay triangulation of the road polygons (Haunert and Sester, 2008). Although the algorithm produces perfect centrelines, the junctions are however problematic.

It is necessary to reconstruct the junctions because sometimes, the resulting centrelines might not be well-behaved at the junctions. The reconstruction of the road centrelines should fulfil the Gestalt principle of continuity (Chang et al., 2001). The principle suggests that people tend to continue shapes beyond their ending points. i.e., the edge of one shape will continue into the space and meet up with other shapes or the edge of the picture plane if they are collinear or follow a similar direction.

The continuity principle is exploited in this study as follows: the triangles of the Delaunay triangulation at the junctions are identified first as depicted by the gray triangles in Figure 5.7. The straight skeletons (illustrated by the dark lines in the figure) that touch the junction triangles are also identified. The general directions of these dark lines are also determined. If any two such axes are almost parallel (for example 1 and 2), based on a specified tolerance in direction, say 20°, they are simply connected by a single line (dotted line). Then the remaining axis (3) is simply extended to intersect the two axes that have already been connected. These intersections (at the right of the figure) are considered as the reconstructed road junctions.

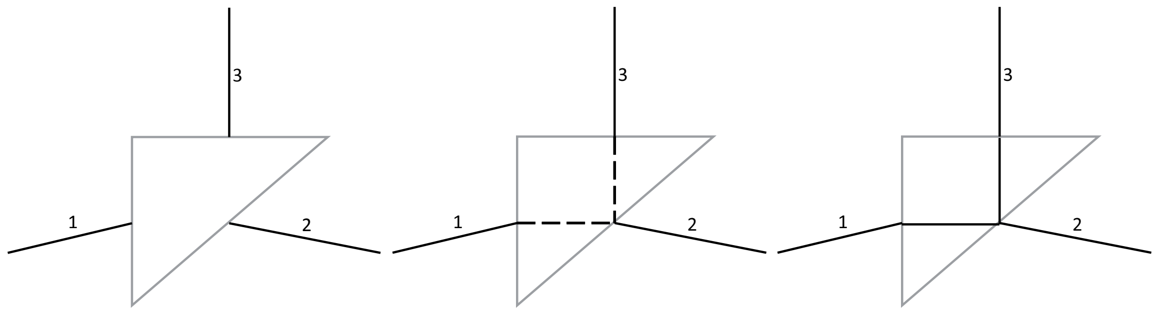


Fig. 5.7: Reconstruction of the road junctions

Figure 5.8 and 5.9 illustrate an example of a situation before and after the road junctions have been reconstructed.

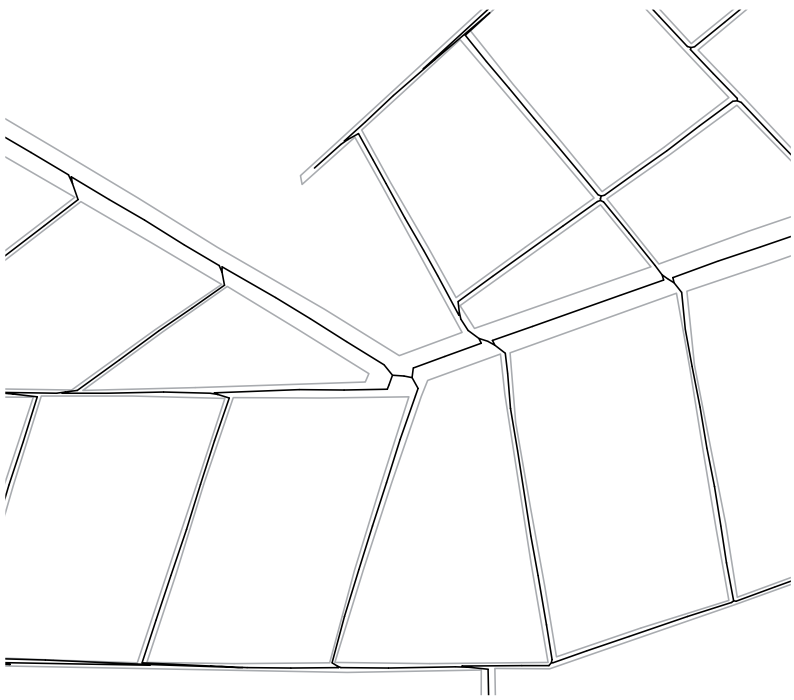


Fig. 5.8: Straight skeletons representing road centrelines

5.4 Feature matching

This step involves the establishment of correspondence between the derived (subject) and the mapped (reference) road networks. In linear network matching, the generic procedure involves, first, a preliminary matching of nodes and arcs (lines) in the network. Based on the results of the pre-matches of the nodes, a final matching of nodes is carried out, which is then followed by a final matching of the lines (Mustière, 2006; Mustière and Devogele, 2008; Kieler et al., 2009).

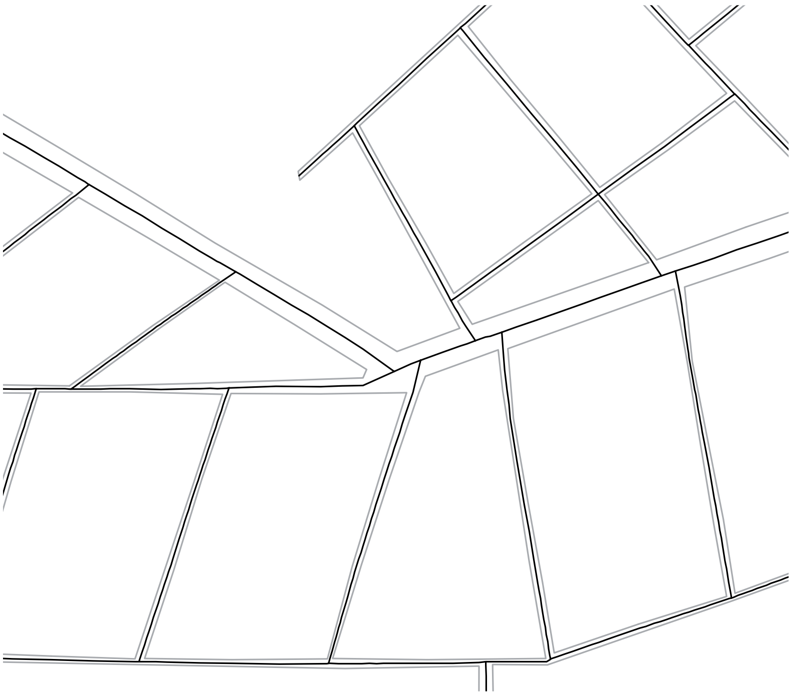


Fig. 5.9: Reconstructed road junctions

The linear network matching algorithm used in this thesis generally adopts this generic approach but leaves out the preliminary matching of nodes and arcs. Instead, the algorithm consists of two main steps: point matching based on relaxation labelling and linear feature matching based on a path matching algorithm. The first step ensures that accurate point correspondences are obtained as a basis for the second step. The second step involves establishing corresponding linear segments from which corresponding vertices (points) are established. The additional corresponding points are combined with the initial corresponding points as the basis for a better local geometric transformation of the subject dataset.

5.4.1 Point feature matching using relaxation labelling

Relaxation labelling (Ranade and Rosenfeld, 1980), an optimization technique, is defined as a formal method that uses iterated local contextual updates during point matching to achieve a globally consistent result. The contextual constraints are often expressed in form of a compatibility function. It is a bottom-up strategy that involves local rating of similarity, which depends on the confidence of label assignments of the neighbours. For road networks, points represent the terminations and junctions, while the topological structure is built from the road network.

Some of the point matching algorithms that consider local geometric distortions have been proposed in literature mainly in the fields of computer vision and pattern recognition. The point matching algorithm, for example, by Chui and Rangarajan (2003) tolerates local distortions. However, it only works on point datasets that contain contour-like structures. The relaxation labelling algorithm is not limited to this condition. The relaxation labelling, which is an inexact point matching algorithm is adopted in this study because it also considers the topological structure that underlie the point datasets.

Formulation

The problem of Relaxation as a labelling is to identify corresponding points (junctions or terminations) from two datasets covering the same geographical extent. For illustration (Figure 5.10), let A= $A_1, A_2 ... A_n$ be a set of network intersections and terminations from the subject network dataset (in grey) and $B = B_1, B_2, ..., B_m$ be a set of network intersections and terminations from the reference network dataset (in black).

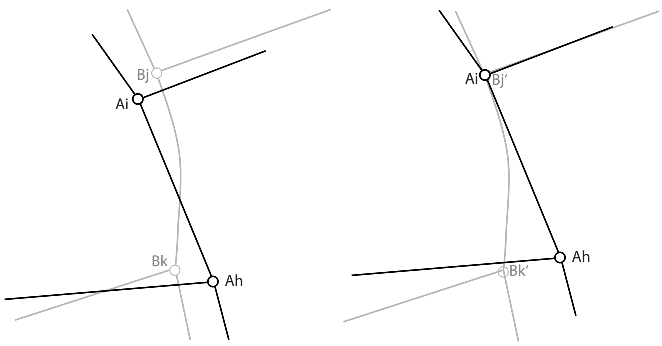


Fig. 5.10: Matching of junctions (points) from subject and reference network datasets based on relaxation labelling

Suppose A_i and B_j are assumed as corresponding points. For any point pair (A_h, B_k), their joint compatibility C(i,j; h,k) is defined as a function of how much the actual position of A_h relative A_i differs from the corresponding position of B_k relative to B_j (Song et al., 2011). If B_j is moved to A_i and becomes $B_{j'}$ and B_k is moved by the same amount to $B_{k'}$. The magnitude of the relative difference δ is the distance $D_{hk'}$ between A_h and $B_{k'}$. Then the compatibility of the pair is defined as $1/(1+\delta^2)$, that is, when $\delta = 0$, compatibility is equal to 1.0; and decreasing to zero as δ becomes large. δ is divided by the distance from A_i to A_h (D_{ih}) to make it a relative rather than an absolute difference (Song et al., 2011). The relative distance is squared to make the compatibility always non-negative.

$$\delta = D_{hk'}/D_{ih} \tag{5.1}$$

$$C(i, j; h, k) = 1/(1 + \delta^2) \tag{5.2}$$

The resulting matrix of point pair compatibility is computed once per datasets pair and is used to modify the confidence assignment in point pair between the two datasets. If P_{ij} represents the confidence of a match between two points A_i and B_j, then at any given iteration, the set of $C(i, j; h, k).P_{hk}$ for all "neighbours" of A_i and B_j is used to obtain a new estimate of P_{ij}. The new estimate defined as the average of the previous estimate and the assignment confidences of the other points. It reasonable to use the maximum rather than the average of the terms $C(i, j; h, k).P_{hk}$, since if any of these terms is large there is a strong support for P_{ij} from A_h, even if all the other terms are small. A possible relaxation formula is then:

$$P_{ij}^{r+1} = \frac{1}{n} \sum_{h=1}^{n} max_{k=1}^{m} \{C(i, j; h, k).P_{hk}^{r}\} \tag{5.3}$$

where r = 0, 1, 2 ... is the number of iterations

The initial estimates of P_s can be made in various ways. $P_{ij}(0)$ can be made with some measure of similarity between A_i and B_j. If a good correspondence exists between some of the points in A and those in B, those $P_{ij}s$ for which A_i corresponds to B_j decreases slowly, since they have support, while the other $P_{ij}s$ decrease more rapidly. The goal of relaxation labelling is not to iterate until convergence but to iterate the assignment of the confidence matrix toward an easily threshold version. The results matrix produces one-to-one correspondences between subsets of points from the two datasets.

Chen et al. (2008) and Song et al. (2011) exploited the similarity of the degree of intersection as the basis for the context information, in which case road intersections of different types are not allowed to match. This however assumes that the datasets are largely similar. In a scenario where the datasets are quite different, more auxiliary information is required. In this study, the angles of incidence of the incident arcs at the junctions and terminations are considered as the basis for the additional information.

The probability of correspondences of the junctions is determined based on the differences in directions of the incident arcs. Figure 5.11 illustrates two road network junctions from the subject (in grey) and the reference (in dark) network datasets. The difference in direction between the subject and each individual reference incident arcs is first established. The reference incident arc (j) with the minimum difference in direction for a particular subject incident arc (i) is considered as the most probable match. For example, in the figure, reference incident arc (j=2) has a high probability of correspondence with subject incident arc (i=1). The probability of correspondence for an individual subject incident arcs is determined as follows:

$$P_i = \frac{30° - min(\|\theta_i - \alpha_j\|)}{30°} \tag{5.4}$$

Where θ_i is the direction at subject incident arc i, α_j is the direction at reference incident arc j. 30° is the maximum allowable difference in direction between the arcs. This was empirically established to be sufficient. To avoid establishing incorrect arc correspondences, a tolerance of the differences in direction should be established. According to equation 5.4, the lower the difference, the higher the probability and vice versa. The overall probability of correspondence between the junctions is determined from the product of the individual incident arc probabilities.

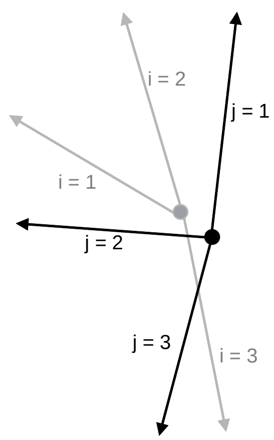


Fig. 5.11: Determination of junction correspondences based on the differences in direction between the subject incident arcs (grey) and the reference incident arcs (dark)

To avoid unnecessary computations, a distance threshold is is set to ensure that only a point pair assignment confidence with assignments and compatibilities of pairs that are relatively close to the chosen pair of points in the two datasets. The value of threshold is not critical as long as it is larger than the maximum displacement (relative positional accuracy) between the two datasets.

Illustration of relaxation labelling

To demonstrate the concept of relaxation labelling, two example network datasets with corresponding junctions and terminations are used. The subject and reference datasets are depicted in grey and dark colours respectively in Figure 5.12.

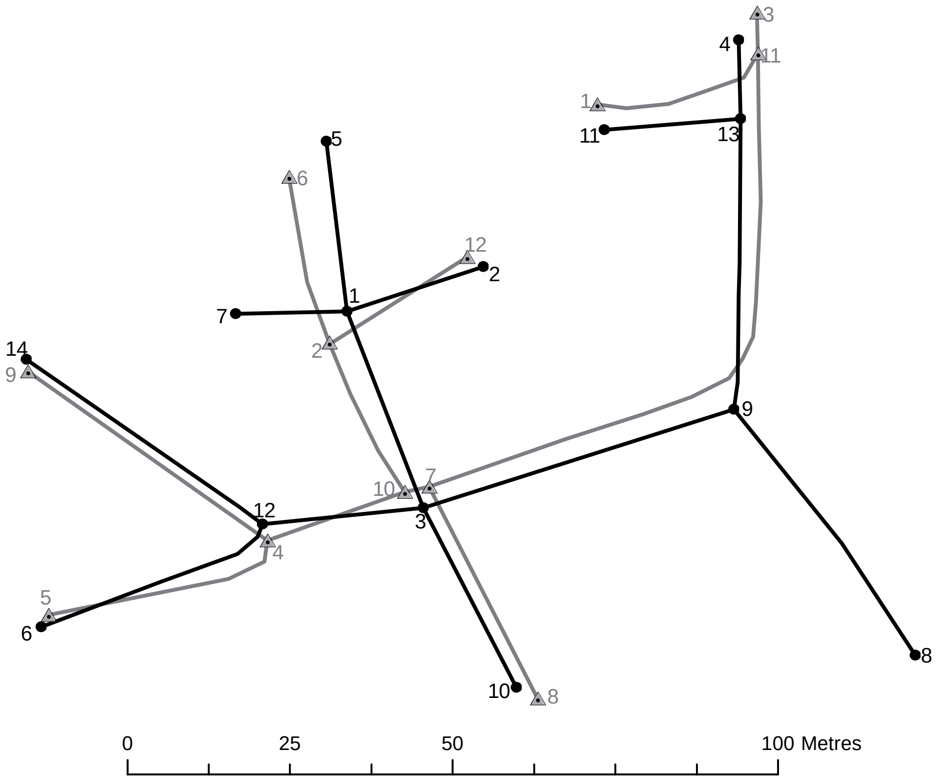


Fig. 5.12: Subject (grey) and reference (black) network datasets

Table 5.1 gives the probabilities based on the correspondence of the incident angles at any given point. The rows show the subject points (1 -12) and the columns show the reference points (1 - 14). The probabilities are calculated based on the sum of the differences of the incident angles at any given junction. The maximum allowed difference specified was 30° for incident arcs considered to be similar. If the sum of the differences is zero, then a probability of 1.0 is expected.

Table 5.2 show the matching probabilities of the same points based on distance between the points. In this example, the threshold distance is 50 metres. Therefore any points in the reference dataset outside of this range with respect to a specific point in the subject dataset are automatically assigned a probability of 0.0. For example, reference points 3 and 10, which are within 50 m of subject point 8, have respective probabilities of 0.316 and 0.926, while the rest have a probability of 0.0.

The initial approximation of the P Matrix in Equation 5.3 is then obtained as the Hadamard product (Marcus and Khan, 1959) of the probabilities of the incident angles and those based on the neighbourhood. Table 5.3 contains the resulting matrix after the multiplication.

After iterating a specified number of times, for each row (i.e., each point in the subject dataset), the column where the maximum value occurs is identified. The reference point at identified column is considered to be the most probable match. Table 5.4 show the resulting match matrix after

Table 5.1: Matching probabilities of reference points with respect to the ditance from points in the subject dataset

0.997	0.003	0.010	0.410	0.441	0.959	0.910	0.395	0.605	0.445	0.929	0.935	0.590	0.709
0.962	0.475	0.958	0.068	0.038	0.562	0.432	0.873	0.917	0.924	0.451	0.770	0.932	0.230
0.892	0.591	0.858	0.998	0.972	0.372	0.503	0.193	0.807	0.142	0.483	0.628	0.983	0.704
0.714	0.005	0.747	0.071	0.065	0.011	0.018	0.218	0.432	0.212	0.016	0.892	0.431	0.000
0.996	0.004	0.993	0.403	0.434	0.967	0.903	0.402	0.612	0.452	0.922	0.928	0.597	0.702
0.958	0.525	0.924	0.932	0.962	0.438	0.568	0.127	0.873	0.076	0.549	0.562	0.917	0.770
0.733	0.084	0.775	0.103	0.115	0.030	0.028	0.019	0.737	0.000	0.019	0.426	0.633	0.114
0.887	0.550	0.967	0.143	0.113	0.487	0.356	0.948	0.842	0.999	0.375	0.845	0.857	0.155
0.814	0.296	0.848	0.704	0.734	0.666	0.797	0.101	0.899	0.152	0.778	0.771	0.689	0.998
0.765	0.010	0.717	0.024	0.032	0.020	0.056	0.155	0.580	0.173	0.048	0.576	0.468	0.087
0.632	0.330	0.617	0.001	0.012	0.004	0.013	0.200	0.790	0.171	0.010	0.571	0.742	0.042
0.799	0.683	0.765	0.910	0.879	0.279	0.410	0.285	0.715	0.235	0.391	0.721	0.925	0.611

Table 5.2: Matching probabilities of reference points with respect to the points in the subject dataset based on a threshold distance. Probabilities values of 0.0 indicate that the points are beyond the specified distance threshold of 50 m

0.002	0.391	0.000	0.522	0.160	0.000	0.000	0.000	0.000	0.000	0.918	0.000	0.557	0.000
0.890	0.472	0.416	0.000	0.382	0.000	0.700	0.000	0.000	0.000	0.000	0.405	0.000	0.067
0.000	0.000	0.000	0.899	0.000	0.000	0.000	0.000	0.000	0.000	0.409	0.000	0.671	0.000
0.254	0.000	0.509	0.000	0.000	0.257	0.296	0.000	0.000	0.111	0.000	0.948	0.000	0.073
0.000	0.000	0.000	0.000	0.000	0.958	0.000	0.000	0.000	0.000	0.000	0.285	0.000	0.211
0.551	0.343	0.000	0.000	0.842	0.000	0.549	0.000	0.000	0.000	0.020	0.000	0.000	0.017
0.405	0.304	0.933	0.000	0.000	0.000	0.204	0.000	0.033	0.326	0.000	0.476	0.000	0.000
0.000	0.000	0.316	0.000	0.000	0.000	0.000	0.000	0.000	0.926	0.000	0.000	0.000	0.000
0.002	0.000	0.000	0.000	0.000	0.214	0.337	0.000	0.000	0.000	0.000	0.139	0.000	0.962
0.417	0.267	0.925	0.000	0.000	0.000	0.246	0.000	0.000	0.308	0.000	0.552	0.000	0.000
0.000	0.000	0.000	0.928	0.000	0.000	0.000	0.000	0.000	0.000	0.473	0.000	0.793	0.000
0.595	0.943	0.219	0.000	0.440	0.000	0.269	0.000	0.056	0.000	0.424	0.000	0.057	0.000

Table 5.3: Initial Compatibility (P) Matrix as a Hadamard product of Table 5.1 and Table 5.2. For each subject point, the reference with the high matching probability can be identified

0.002	0.001	0.000	0.214	0.071	0.000	0.000	0.000	0.000	0.000	0.853	0.000	0.328	0.000
0.856	0.224	0.398	0.000	0.014	0.000	0.302	0.000	0.000	0.000	0.000	0.312	0.000	0.015
0.000	0.000	0.000	0.897	0.000	0.000	0.000	0.000	0.000	0.000	0.198	0.000	0.659	0.000
0.181	0.000	0.381	0.000	0.000	0.003	0.005	0.000	0.000	0.023	0.000	0.845	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.926	0.000	0.000	0.000	0.000	0.000	0.265	0.000	0.148
0.528	0.180	0.000	0.000	0.810	0.000	0.312	0.000	0.000	0.000	0.011	0.000	0.000	0.013
0.297	0.025	0.723	0.000	0.000	0.000	0.006	0.000	0.025	0.000	0.000	0.203	0.000	0.000
0.000	0.000	0.306	0.000	0.000	0.000	0.000	0.000	0.000	0.925	0.000	0.000	0.000	0.000
0.002	0.000	0.000	0.000	0.000	0.143	0.268	0.000	0.000	0.000	0.000	0.107	0.000	0.960
0.319	0.003	0.664	0.000	0.000	0.000	0.014	0.000	0.000	0.053	0.000	0.318	0.000	0.000
0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.588	0.000
0.476	0.644	0.168	0.000	0.387	0.000	0.110	0.000	0.040	0.000	0.166	0.000	0.053	0.000

3 iterations. For example, the highest probability (0.010) for subject point 1 occurs at reference column 11. Similarly, subject point 2 corresponds to reference point 1.

In Figure 5.12, it is important to note that reference point 3 could have both subject points 7 and 10 as possible matches. If we consider the match with the highest compatibility, then the correct match for reference point 3 should be reference point 10. Such cases, which are due to generalization during data capture should therefore, warrant a visual validation after the matching process. The

Table 5.4: Corresponding points after 3 iterations

0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000
0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.0002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
0.000	0.000	0.140	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.0001	0.000
0.000	0.0003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

reference points 7, 8 and 9, which seem not to have any correspondences, are not matched to any subject point (see Table 5.4).

5.4.2 Linear network matching using path matching algorithm

The objective of this step is to ensure that only one-to-one linear feature correspondences are obtained instead of one-to-many or many-to-many feature correspondences. This is achieved by combining individual linear features in the network into paths (i.e., compound linear features) and find corresponding points between them using a path matching algorithm.

The path matching algorithm should find the corresponding paths between a given pair of corresponding nodes obtained during the point matching process. A matching precision of 100% is assured if all the point matches are accurate. To identify the corresponding paths, the shortest paths between any two nodes from one dataset are determined. A potential match is found in case a (shortest) path between the corresponding nodes from the other dataset exists. Determining the shortest path employs the Dijkstra shortest path algorithm, DSPA (Dijkstra, 1959).

The matching proceeds as follows: first, the overall shortest path within the subject network (graph) is established. In case a match is found in the reference network, the paths are removed from the graphs and a new overall shortest path needs to be determined. The removal is necessary in order to increase the matching speed and quality. Depending on the size of a given graph, applying the DSPA many times may become a time consuming task. Since the first shortest paths are likely to consist of a single edge, the path matching step is split into two. The first step incorporates matching only single edges from one dataset to the corresponding shortest path in the other dataset and vice versa. This includes only one DSPA between two points in one graph instead of DSPA between all points in both graphs.

After that step, all one-to-many and many-to-one linear feature matches are removed from both graphs, which are now likely to be split into various unconnected sub-graphs. This improves the DSPA search, because not all nodes are connected and therefore fewer paths exist in the graph. Finally, apart from the paths that have a one-to-one linear feature correspondence, all the other path that consist of one-to-many and many-to-many line correspondences are combined to form compound one-to-one linear features for the purpose of determining corresponding vertices. Since the worst-case running time for the Dijkstra algorithm on a graph with n-nodes and m-edges is $O(n^2)$ (Barbehenn,1998), the time complexity for the path matching algorithm would be $O(n^4)$ considering that two graphs are involves and the whole process involves sub-search process, i.e., one-to-one, one-to-many and many-to-many

Figure 5.13 shows the corresponding paths obtained after applying the path matching algorithm. The matched paths are depicted in corresponding colours, while the unmatched arcs are shown in broken lines. Incident arc labelled 6 from the subject dataset should be matched to incident arcs labelled 7 and 10 from the reference dataset. However, no correspondence is established because the linear feature from the subject dataset does not have a junction (see Figure 5.12).

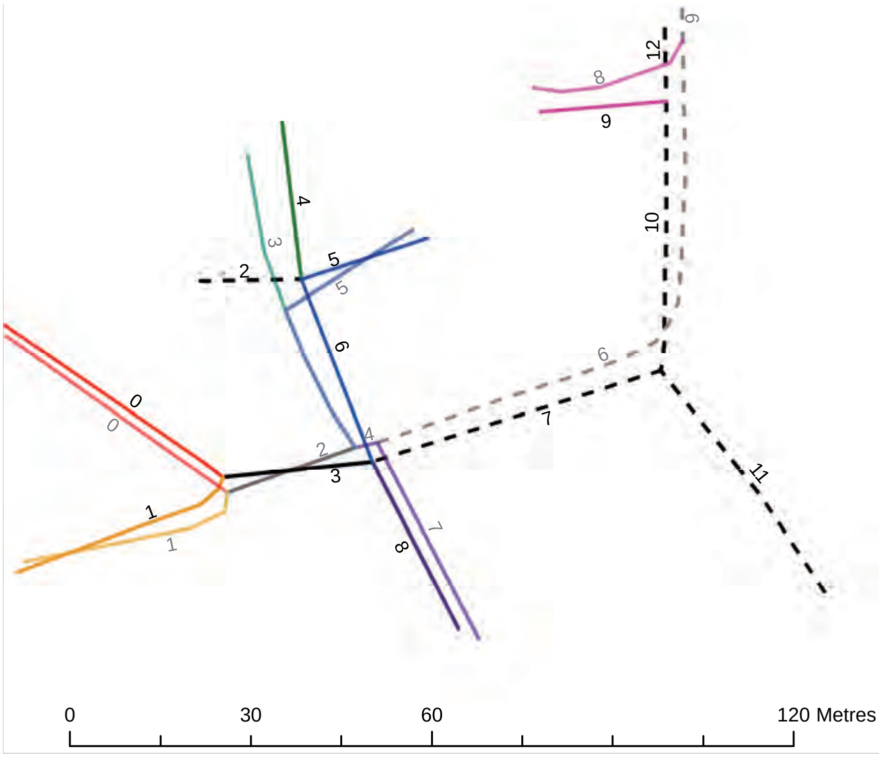


Fig. 5.13: Matching incident arcs (paths) are depicted with corresponding colours; unmatched incident arcs (paths) are shown by broken lines

The weight of an edge is set to the length of the underlying polyline because the DSPA is based on weighted graph edges. The identified match between the two shortest paths still needs to be validated. For this reason, various validation methods can be applied, for example, the lengths of the identified paths could be compared. In case the length ratio is within a given threshold (e.g. 1:2), which can be empirically determined, the path match is then considered valid. This condition may fail if the paths form loops as illustrated in Figure 5.14. In the figure, the subject dataset is represented by the grey lines and the nodes with capital letters (ABCD); the reference dataset is represented in black lines with its nodes in small letters (abc). If a path ABC is established in the subject datasets, there might be two alternative paths (ab and acb) in the reference dataset that fulfil the specified condition. To ensure a robust solution, an additional condition should be specified, for example, requiring that the Hausdorff distance between the corresponding paths is less than the relative positional accuracy of the datasets.

The overall objective of the entire approach is to obtain correspondences between the subject and the reference datasets as the basis for geometric transformation. Because geometric transformations are based on corresponding coordinates, it is necessary to obtained all the corresponding coordinates between the subject and reference datasets. Corresponding coordinates from the corresponding points are already available. Additional coordinate correspondences are obtained from the corresponding vertices of the corresponding linear features obtained through the path matching algorithm. These additional coordinate correspondences should lead to a better geometric alignment of the subject dataset with the reference dataset.

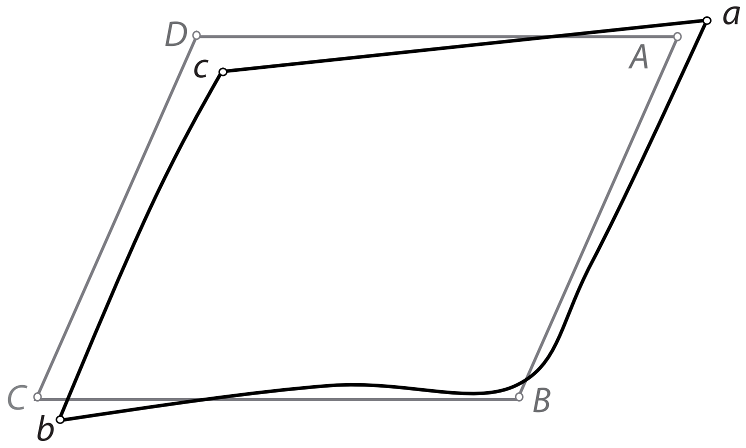


Fig. 5.14: Presence of more than one corresponding paths in the reference dataset to a specified path in the subject dataset

5.4.3 Finding corresponding vertices of two similar linear features

Given similar two linear features, for example, in Figure 5.15, it may be required to identify their corresponding vertices. This study exploits the relative curve orientations paradigm to achieve this. The concept behind the curve orientation is that high curvature points along a linear feature usually possess an anatomical meaning. They are therefore good landmarks to guide the matching process, especially in the absence of a reliable physical or deformable geometric model of the observed structures (Cohen et al., 1992). An algorithm for vertex matching could generally consist of two steps: i) computation of the direction change ii) the identification of the matching vertices.

Direction Change

To compare the direction changes between the vertices of the corresponding curves, normalization of the curves is required. A direction change is calculated at each vertex based on the relative normalized position. The first and the last vertices are assigned a direction change of zero, since no direction change takes place there. Figure 5.15 show an example of a pair of corresponding polylines captured by different means. The subject polyline is represented in grey, while the reference polyline is represented in black line.

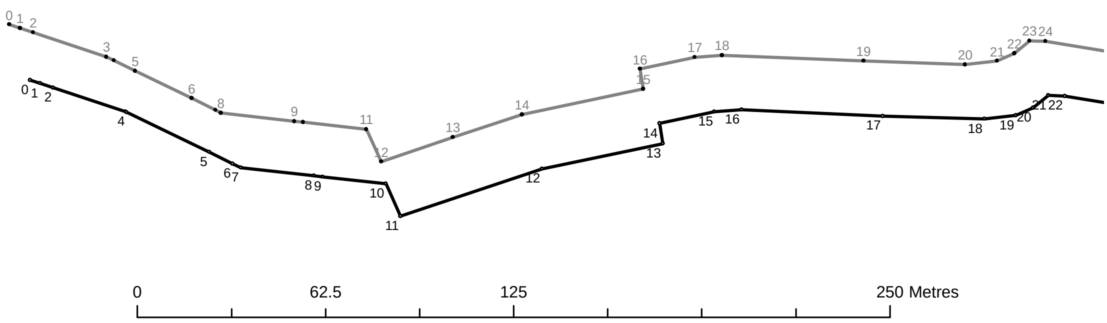


Fig. 5.15: Corresponding subject and reference polylines

A corresponding graph of the direction change (vertical-axis) at each vertex plotted against its relative normalized position (horizontal-axis) is shown in Figure 5.16. The similarity of the example polylines in Figure 5.15 based on the direction change of their points is evident. The basis for identifying the actual vertex-matches is described in the next step.

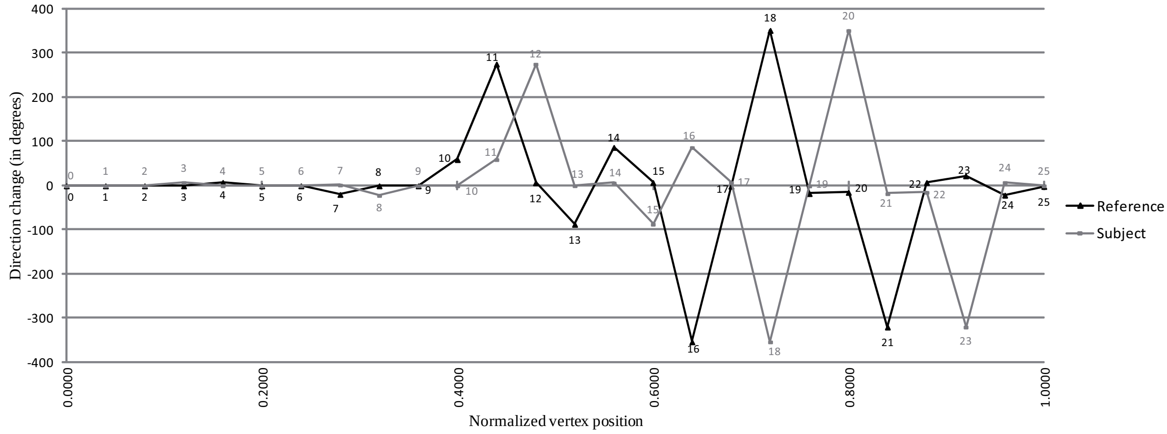


Fig. 5.16: Direction change at the vertices for the subject and reference polylines shown in Fig. 5.15

Matching vertices

Matching of the vertices depends on the distance between the relative normalized positions of the vertices. Matches with the least distance are considered to be the most probable. A further condition is included to ensure that the vertices have comparable amplitude, i.e., the perpendicular distance of the vertices and the straight line connecting the start and end points of the respective polylines. The vertex matches are ascertained further if the curvature (i.e., the derivative of the curve's tangent angle with respect to position on the curve at that point) at the vertices is nearly the same.

The subject curve has got 25 vertices (i.e., 0 - 24) in Figure 5.15 while the reference curve has got 23 vertices (i.e., 0 - 22). If the graph of direction change for the subject curve is slightly shifted to the left, a correspondence with graph of direction change of the reference curve will be achieved. The vertex with correspondences and those without (for example, vertex 14 from the subject curve) can be identified easily. The complete vertex correspondence for the two curves is given in Table 5.5.

Table 5.5: Correspondence of vertices between two similar curves

Subject	0	1	3	8	9	11	12	13	15	16	17	18	19	20	21	22	23	24
Reference	0	1	4	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22

5.5 Geometric transformation

This step entails the geometric quality enhancement of the approximate graphical cadastres through geometric transformation of all the data in the approximate graphical cadastre especially the land parcel boundaries.

Due to the presence of local systematic geometric distortions in initial the graphical cadastre considered in this study, a geometric transformation that considers local distortions should be selected. The common Helmert or affine transformation does not consider correlations between neighbouring points. They are therefore not sufficient for modelling local distortions.

In most studies involving the reconstruction of cadastres, the overarching assumption is that the original field measurements of the digitized cadastre, which is the subject of adjustment still exist in field books. These field books usually contain redundant observations which are exploited in a least squares adjustment. Additional geometric conditions are introduced to the adjustment process using either deterministic or stochastic approaches.

However, usually no field measurements exist for a cadastre obtained via graphical means and for rural areas where land parcels have irregular shapes. The absence of redundant observations and the irregularity of shapes of land parcels in such graphical cadastres mean that the geometric adjustment can neither benefit from redundant measurements nor from any geometric conditions.

To determine the most optimal transformation approach for such graphical cadastres, we have to consider that the observations include only the coordinates of points graphically derived from digitized cadastral maps. The positional differences between corresponding points from the subject and reference maps are not only as a result of random uncertainties, but also because of the distortions of the local geometry. Geometric transformation methods that consider local geometric distortions that might be considered include for example, piece-wise rubber sheeting, which is technically referred to as Piecewise Linear Homeomorphism (PLH) (Gillman, 1985) and homogenization followed by adjustment (Gründig et al., 2007).

Rubber sheeting transformation can be implemented in many different ways. One implementation involves radial basis functions like the Thin Plate Splines (TPS). The downside of these approaches is the high number of point correspondences required which often translates to high computation times. Some of the approximate methods for dealing with a high number of correspondences during the transformation are discussed in (Donato and Belongie, 2002). Piece-wise rubber sheeting based on a polynomial of a higher degree and the Delaunay triangulation will only ensure smooth connection between the triangular sheets, but the positional distortion in the interior of the triangular nets becomes extreme. Piece-wise rubber sheeting transformation is often based on affine transformation and Delaunay triangulation and usually yields a perfect correspondence between the corresponding points although this results in discrepancies between adjacent triangles.

5.6 Piece-wise affine transformation

Piecewise affine transformation was implemented in this reaserch by creating a triangulation of the points in the subject dataset and the corresponding points from the reference dataset. This is followed by a linear transformation between one triangle in the subject dataset and the corresponding triangle in the reference dataset. When the subject dataset has local geometric distortions, piece-wise affine transformation is a better model compared to surface splines. This is because, unlike surface splines, this transformation affects only the regions inside the convex hull of the corresponding points, while the areas outside the convex hull are not affected at all.

5.6.1 Delaunay triangulation

The triangles upon which the affine transformation is built are preferably based on the Delaunay criterion. However, sometimes there might be no similarity between the two sets of triangulations. An example of such a case is illustrated in Figure 5.17, in which dissimilar triangles are created.

To make the triangles similar, either of the triangulations can be manually edited. To avoid this manual editing which is prone to error , a triangulation of the subject point set is carried out first and then a corresponding quasi-Delaunay triangulation is carried out for the reference point set. The points in both datasets are obtained from the point and path matching process. Figure 5.18 show

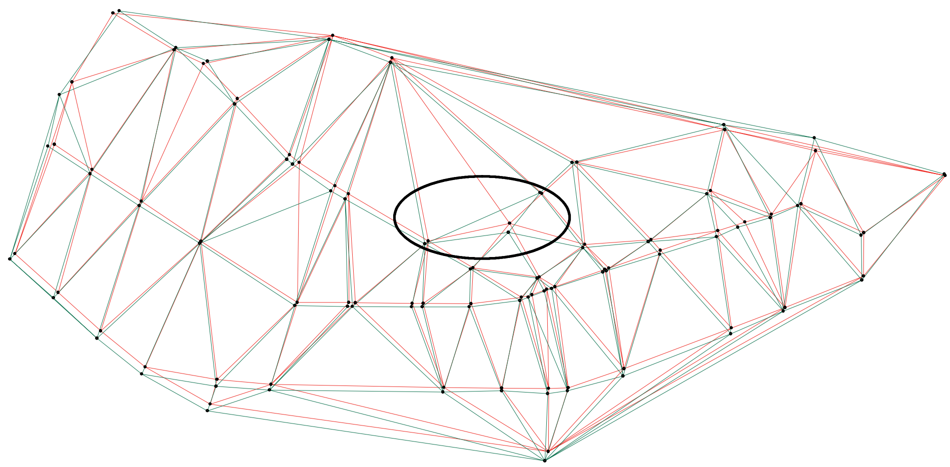


Fig. 5.17: Simultaneous Delaunay triangulation of corresponding points from point datasets with the subject dataset (red) and the reference dataset (green). Section with dissimilar triangles is highlighted in black ellipse

the Delaunay triangulation (in red) of the points from the subject dataset and the corresponding quasi-Delaunay triangulation (in green) of the corresponding points from the reference dataset.

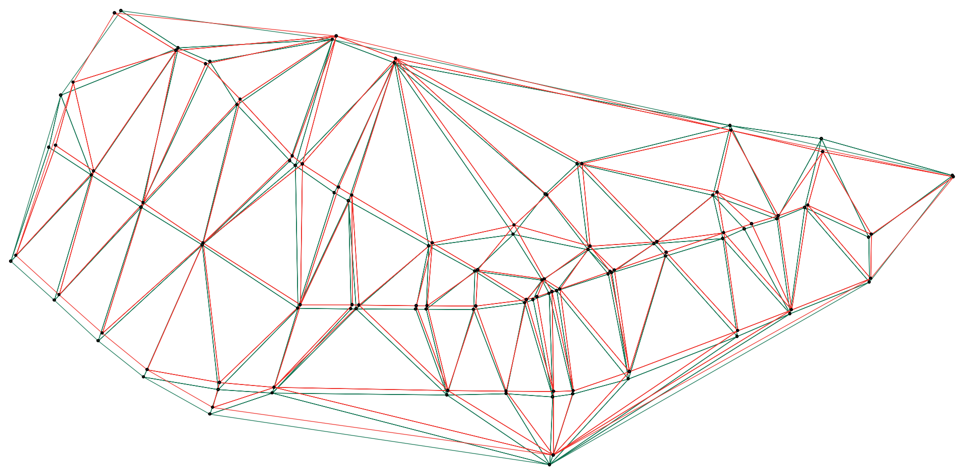


Fig. 5.18: Quasi-Delaunay triangulation of the points from the reference dataset made with respect to the Delaunay triangulation of the points from the subject dataset

5.6.2 Formulation

For each pair of corresponding triangles, a planar affine transformation is determined. The affine transformation equations in 2-D is of the form:

$$
\begin{aligned}
X &= a_1x + a_2y + a_3 \\
Y &= b_1y + b_2x + b_3
\end{aligned}
\tag{5.5}
$$

Where (X, Y) and (x, y) are coordinates of the reference and subject points respectively. a_1, a_2, a_3, b_1, b_2 and b_3 are the parameters of the affine transformation. These include 2 parameters for translation, 2 scale factors and 2 parameters for change in orientation. For a unique solution, exactly three pairs of corresponding points are needed. Therefore, for each pair of corresponding triangles, the parameters of the affine transformation are determined using their corner points. After this, all points within the triangles are transformed into new positions.

5.7 Assessment of quality improvement

The extent to which the geometric quality is improved through the data conflation approach is one of the aims of this thesis. Ordinarily, the extent of improvement can be determined by comparing the geometry of the land parcels in the cadastre before and after the process with some reference measurements. The quality can be assessed in terms of the improvement in position alignment and geometric area. Improvement in position alignment can be determined from corresponding subject and reference paths.

The evaluation of improvement in area can generally be carried out at an individual land parcel level or a group of parcels. The extent of area improvement for any given parcel depends on how close to the "true" value the transformed area gets. Typically though, no reference geometric area measures exist for provisional graphical cadastres, unless field measurements are carried out for the purpose of reference.

Since there are insufficient visible boundary signatures on the orthophoto to allow the digitization of the outline of individual land parcels, the evaluation of the geometric area accuracy improvement can be carried out based on groups of land parcel, which are identifiable and have a high positional certainty. Thus, the outlines of the groups of parcels, which are bounded by roads on all their sides, are first digitized from the orthophoto. The corresponding land parcel groups in the graphical cadastral dataset are obtained by analogously combining adjacent land parcels.

6. Experimentation and Analysis

This chapter presents the implementation of the data conflation approach for geometric quality improvement presented in the previous chapter. The main objective is to assess the extent of geometric quality improvement of cadastral datasets with the data conflation approach. The different steps of the approach presented in the previous chapter are experimented and evaluated using two experimental cases.

6.1 Experimental data

The experimented datasets include a land parcel boundary (or cadastral) dataset and a topographical dataset mainly from the urban fringe of two urban areas in Kenya, Nairobi and Machakos. In the urban fringe, usually urban and rural land uses often overlap and these environments are continuously changing. The two test areas hereafter referred to as Case 1 and Case 2, each covers a spatial extent of about 3.5 Km by 3.5 Km (see Appendix A). The topographical datasets, which are considered as the reference datasets, consists mainly of road network, building footprints and administrative boundaries. The cadastral datasets which mainly consists of land parcel outlines are the subject dataset.

The topographical datasets were captured using stereo-photogrammetric techniques at a representative map scale of 1:2,500. They are generally more up-to-date and accurate in position than their cadastral counterparts. According to the "Accuracy Standards for Large-Scale Maps" (ASPRS, 1990), the limiting RMSEs for these topographic maps at a publication scale of 1:2,500 are respectively, 0.625 m, 1.25 m and 1.875 m for Class 1, 2 and 3.

The cadastral datasets in both cases were however, captured using different photogrammetric techniques and have different positional accuracies. The cadastral dataset from Case 1 and Case 2 were respectively captured using the second and the fifth photogrammetric technique described in Section 2.4.1. Thus, the cadastral dataset from Case 1 are more accurate in position than those from Case 2, which contains significant local positional distortions. Another notable difference is that the cadastral datasets from Case 1 have a more regular spatial arrangement than those from Case 2.

In addition to being of poor geometric quality, the subject analogue maps (the cadastral dataset from Case 1) were not spatially referenced to any coordinate system. An approximate geometric registration was carried out based on similar positions identified manually in the scanned analogue cadastral map and the reference topographical dataset.

Figure 6.1 shows sections of the datasets from the two experimental areas, with dataset from Case 1 and Case 2 on the left and right hand side respectively. The top part of the figure shows the land parcel boundaries superimposed on orthophotos and the bottom shows the corresponding road network and building footprints from the topographical datasets.

Fig. 6.1: At the top are the cadastral dataset (land parcel boundaries) superimposed on orthophotos and at the bottom are the topographical datasets (road network and building footprints) for Case 1 (left) and Case 2 (right) respectively

6.2 Goal and organization of the analysis

The main goal of the analysis is to highlight the possibilities and limitations of each step of the data conflation approach and its effect on the final result. The analysis will focus on the critical aspects of each step in the approach through illustration and analysis of results.

6.3 Derivation of the road network

The extraction of a road network from a cadastral dataset that consists of only privately-owned land parcels is based on the assumption that the empty spaces between the parcel polygons in the dataset represent the road polygons. The process consists of two steps as described in chapter 5. The results from each process are illustrated and evaluated in the following sub-sections.

6.3.1 Extraction of road polygons

The extraction of the road polygons was achieved as a geometry difference between the parcel polygons in the cadastral dataset and their concave hull. Using a concave hull approach presented the possibility to avoid the spurious polygons at the periphery of the dataset extent after carrying out the geometric difference. The challenge in the approach however, was the creation the concave hull.

An algorithm for the creation of the concave hull was created based on the concepts of Thiessen polygons and linear referencing as described in the previous chapter. The only parameter required in the algorithm hereafter referred to as the concave hull algorithm, is the number of iterations. The algorithm was designed for polygon datasets although it can be applied to both line and point datasets.

The concave hull continues to evolve by including more vertices in successive iterations. The number of iterations beyond which no more vertices are added to the concave hulls is 10 and 14 respectively for Case 1 and Case 2. This means that the concave hull does not change anymore. Therefore to ensure that the concave hull attains its final state, the number of iteration should be adjusted accordingly during the execution of the algorithm.

Figure 6.2 show the resulting concave hull after 20 iterations for a set of parcel polygons (shaded in grey) from Case 1. The final concave hull is shown in green colour. The red rectangle shows a section where the concave hull does not coincide with the outline of the parcel polygons despite the addition of more vertices through the segmentation of the polygons. The maximum segmentation distance that was possible for Case 1 to follow the concavity was 20 m, which increased the number of iterations required for the concave hull to reach a stable state. Any distance less than this value would lead to run time exceptions when creating the Thiessen polygons. Therefore, segmentation is only possible to a certain extent and thus solves the problem of concavity to a certain extent. Extreme concavity is still an outstanding issue in the concave hull algorithm. One possibility to solve the problem of extreme concavity is an appropriate adaptation the snake-based approach for contour detection for objects with boundary concavity (Kim et al., 2008).

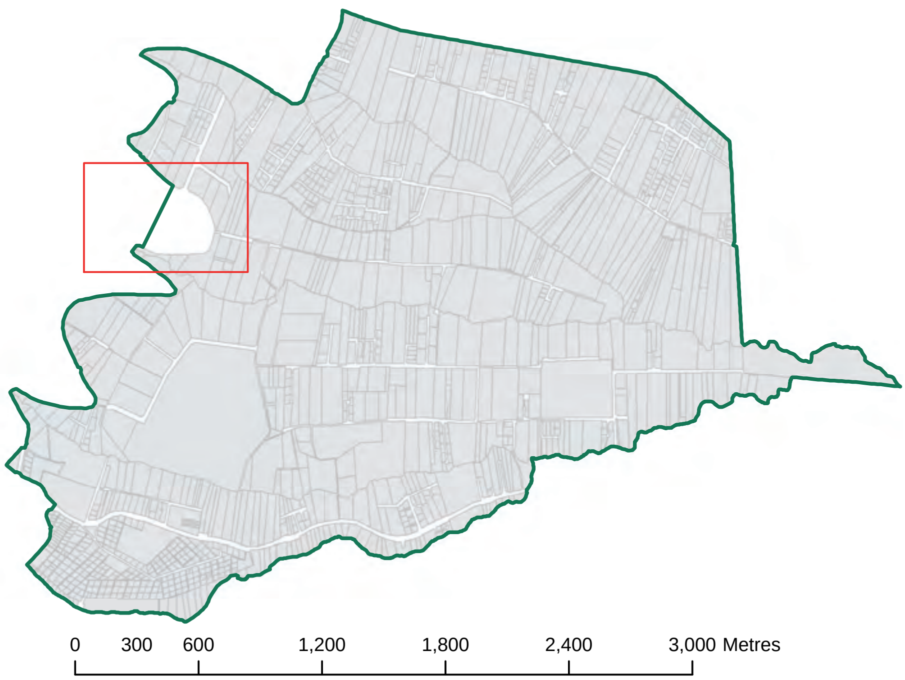


Fig. 6.2: Resulting concave hull for Case 1 after segmentation

Before the road polygons could be obtained as the geometric difference between the concave hull and the input parcel polygons, the few sections of the concave hull that could not follow the concavity were manually edited. Otherwise, the effect of the concavity problem in the subsequent process (i.e., road network derivation) is the creation of spurious road segments at the periphery of the study area. These counterfeit road segments are potential outliers during the feature matching process. Figure 6.3 shows the resulting road polygon shaded in grey together with the outline of the edited concave hull for Case 1.The concave hull and the resulting road polygons for Case 2 are presented in Appendix B.

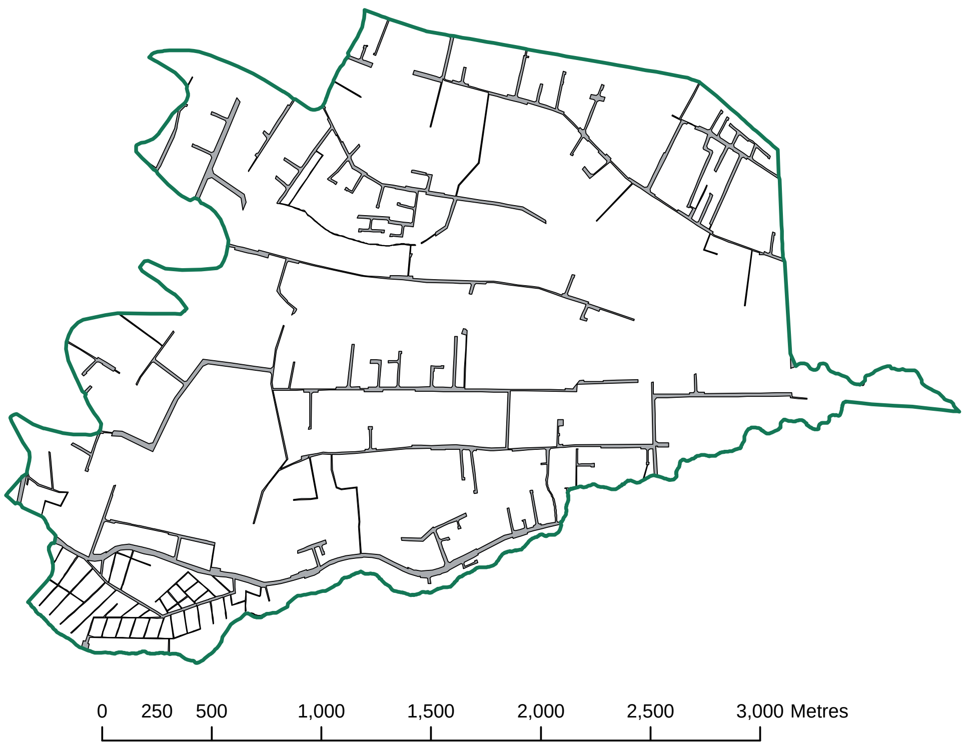


Fig. 6.3: Resulting road polygons (shaded in grey) and the edited outline of concave hull (green)

The main strength of the algorithm lies in the fact that no thresholds specifications are required except for the number of iterations, which can be adjusted as necessary. The main limitation of the algorithm is the presence of extreme concavities in the input geometry. Concavity of some degree can be overcome by the introduction of additional points through the segmentation of the input datasets. This is only possible if the input geometry consists of either line or polygon features. There is however, no control in concavity when the input geometry contains only point features.

6.3.2 Administrative boundary delineation

Apart from the extraction of road polygons from a land parcel dataset, a possible application of the concave hull algorithm is the delineation of the boundary of a city through the amalgamation of the land parcel boundaries. Figure 6.4 depicts a comparison of the concave hull of the land parcel boundaries in Case 1 and the reference boundary (in red outline) contained in the reference topographic dataset. The length of the boundary generated via the concave hull algorithm is about 15 m longer than the reference boundary, although it is 150,000 square metres smaller than the reference. This is because the boundary generated through the concave hull follows only the boundaries of the land parcels ignoring the adjacent road centrelines which are often considered as the actual location of the boundary. Therefore using the concave hull algorithm should consider this fact to obtain the correct boundary.

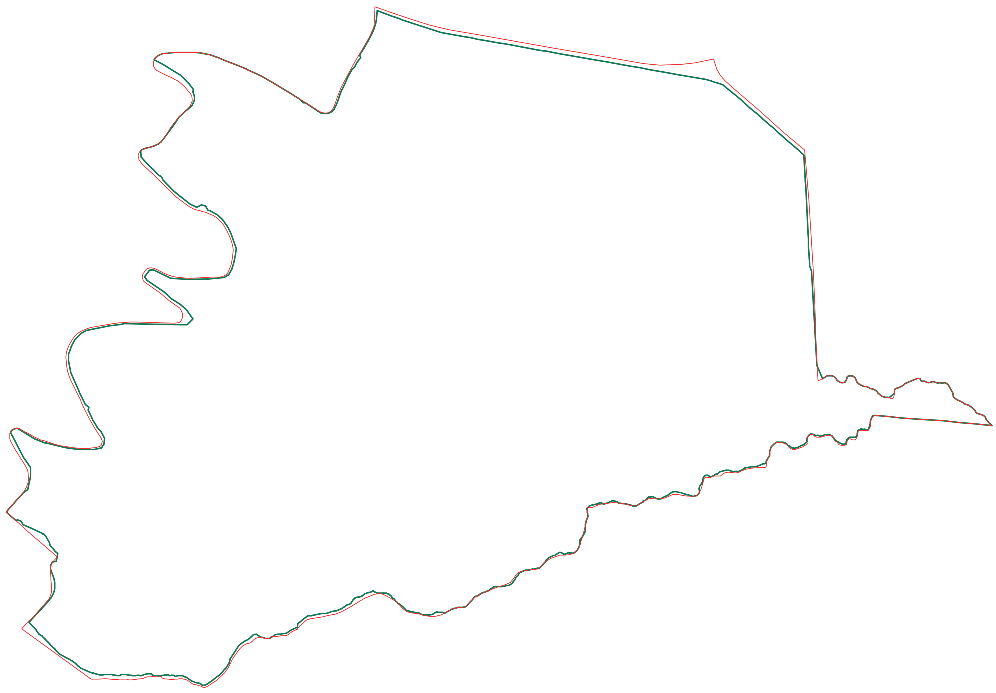


Fig. 6.4: Comparison of boundaries delineated using the concave hull algorithm with the reference boundary

6.3.3 Creation of a network of road centrelines

The road polygons that were obtained from the geometric difference of the input polygons and their concave hull, were collapsed to their corresponding centrelines using straight skeletons based on Delaunay triangulation. The road junctions were also reconstructed according to the procedure described in the previous chapter. Figures 6.5 and 6.6 show the results of the road network centrelines for Case 1 and Case 2 respectively.

Prior to feature matching, the derived road network was subjected to some absolute and relative quality assessment (ISO, 2010). Absolute quality involved a determination of connectivity of the road segments and the presence (or absence) of stand-alone features, i.e., features not connected to the main network. No topological errors were detected because the technique of straight skeletons used to create the road centrelines automatically ensures that the segments are connected for form a network.

In Case 2, a section (highlighted in Figure 6.6) was falsely extracted as segment representing a road centreline instead of a river centreline when compared with the topographical dataset. Despite the assumption that the gaps represent road reserves, it was noted that there was inconsistent representation of rivers on the original cadastral maps. On one hand, the river s are represented as polygons, and as lines on the other. It has to be established in advance if the river features are consistently represented prior to applying the algorithm.

Stand-alone road segments were noted particular at the periphery of the study area involved. These can be identified in Figures 6.5 and 6.6. Stand-alone features are not necessarily mistakes but as terminal segments of another road network that extends beyond the extent of the study area.

Fig. 6.5: Derived road network for Case 1

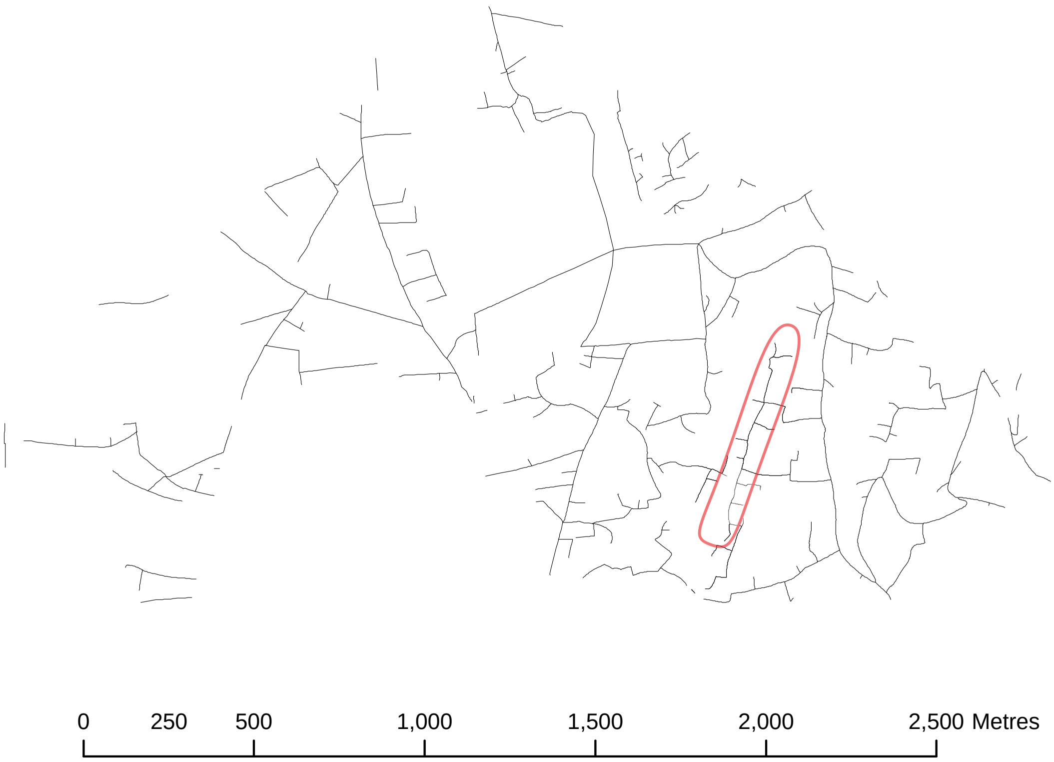


Fig. 6.6: Derived road network for Case 2

The location of the vertices in the input road polygons influenced the straightness of the final centrelines. Figure 6.7 shows an example where centrelines are curved due to the influence of the vertices. The bold grey lines are the road polygons outlines and the black line is the derived centrelines. The ideal straight lines are depicted by the broken black line. This problem can be avoided by simplifying the centrelines using, for example, the Douglas Peucker algorithm (Douglas and Peuker, 1973).

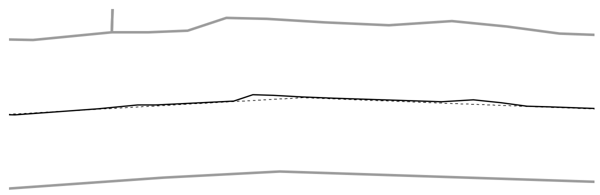


Fig. 6.7: Influenced of the vertices in the input road polygons on the derived road centrelines

Relative quality involves the evaluation of the quality of an object with respect to another considered to be of superior quality. The relative quality of the derived was evaluated using a mapped (reference) road network of better quality. Comparing the derived and the reference road network, it was noted that there were more and shorter segments in the reference road network than in the derived road network. This will therefore have a great influence in the matching process. The short segments in the reference road network dataset, which were however absent in the derived network indicate that there are more existing road features than those implicitly contained in the cadastral dataset.

6.4 Road network matching

The process of matching the road networks was carried out in two stages as described in Chapter 5. The first stage involved point matching based on relaxation labelling, followed by network matching based on path matching. Path matching was considered necessary to obtain additional point correspondences along linear features for improved geometric transformation.

6.4.1 Point feature matching

The actual threshold for the search radius is not critical as long as it is larger than the displacement between the two datasets. For both cases, it was specified as 100 m. Similarly, to calculate a new value P_{ij} at each iteration, only the points within a given distance of the point in question were thus avoiding unnecessary computations. In both cases, a distance of 200 m was specified. To also allow intersections of different degree to correspond, a maximum direction difference between possible corresponding incident arcs of 30° was specified.

Figure 6.8 and Figure 6.9 show the result of point matching for Case 1 and Case 2 respectively. In both figures, there is an even distribution of subject points (grey circles) and the corresponding reference points (black crosses). Because of the scale, the dots representing corresponding points are overlapping.

Because the overall result depends on the matching results from both point and path matching, it is necessary to determine the accuracy of these processes. For both point and path matching processes, precision and recall are used as the quality measures. These measures indicate the fraction of instances for which the correct result is returned.

Fig. 6.8: Case 1 - Distribution of corresponding subject (grey circles) and reference (black crosses) points

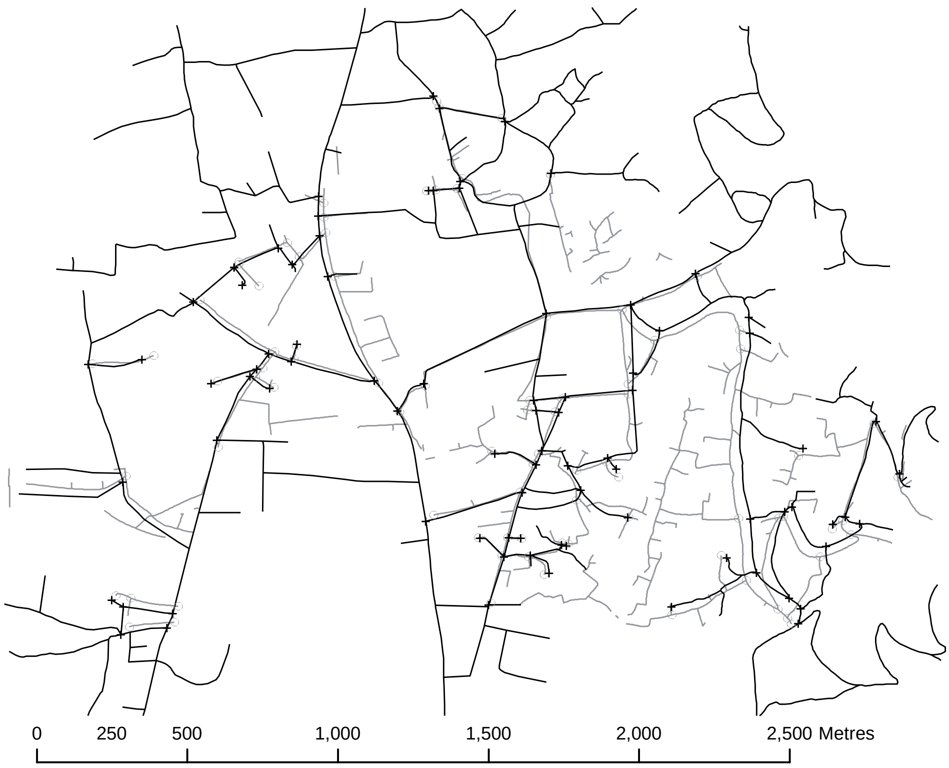


Fig. 6.9: Case 2 - Distribution of corresponding subject (grey circles) and reference (black crosses) points

Given a pair of corresponding features in both the subject and reference datasets, true positive *(tp)* is the number of correctly detected matches. False positive *(fp)* is the number of incorrectly detected matches. A false negative *(fn)* is the number of matches that are not detected by the algorithm. False positives and false negatives are obtained through a visual validation. Precision is then the number of correct matches *(tp)* divided by the total number of matches *(tp + fp)* found by the algorithm (Equation 6.1). Recall is the number of correct matches divided by the total number of actual true matches *(tp + fn)* (Equation 6.2).

$$precision = \frac{tp}{tp + fp} \tag{6.1}$$

$$recall = \frac{tp}{tp + fn} \tag{6.2}$$

Precision can be seen as a measure of exactness or correctness of the matching algorithm, whereas recall is a measure of completeness of matching. Points without matches can be regarded as true negatives *(tn)*.

The evaluation of the data matching results is given in Table 6.1 includes matching results for the two experimental cases. The table presents the total number of points from the subject dataset involved. The points include both the junctions (intersections) and terminations.

From the table, there is a relatively high precision for all the datasets, Case1 has both the highest recall and precision. High recall percentages indicate that the subject and reference datasets are highly similar, while low values indicate that many features in both datasets do not have correspondences. The low recall rate of point matching for Case 2 suggests that there is a significant difference both in the number and positional difference between the features in the subject and reference road networks as evident in Figure 6.9.

Table 6.1: Evaluation of point matching based on relaxation labelling

	Total subject points	True positives *(tp)*	False positives *(fp)*	False negatives *(fn)*	Points with matches	Points without matches *(tn)*	Precision (%)	Recall (%)
Case 1	311	117	33	13	163	148	78	90
Case 2	314	36	9	35	80	234	80	51

These quality measures to some extent indicate the similarity of the datasets in addition to being an indication of the integrity of the matching. The difference in both datasets is also due to some generalizations introduced during the data capture process. During mapping, there is usually subjective interpretation and generalization. For example, Figure 6.10 (left) shows one such case, where there is one road junction 12 from the reference road network (in black), while there are two possible corresponding junctions (93 and 238) in the subject road network (in grey). Figure 6.10 (right) shows a reverse case, where there are two reference junctions (42 and 340) with possible correspondence to one junction (15) from the subject dataset. In such circumstances, the selection of the correct match is purely subjective.

For the subsequent process, all the points with correspondences (true positives, false positives and false negatives) were considered in order to increase the number of correspondences and also to model the local distortions in the subject dataset during geometric alignment more comprehensively.

Relaxation labelling was considered in this research because of its capability to establish correspondences in the presence of local geometric deformation. The main strength of the algorithm based

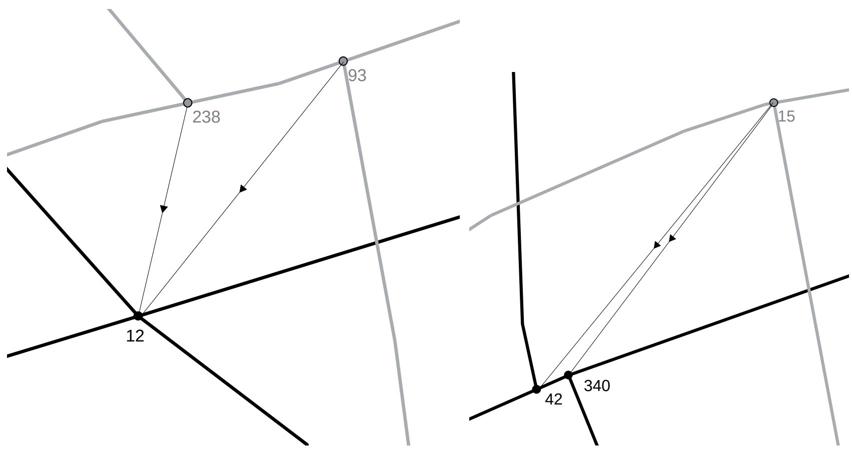


Fig. 6.10: Differences in the datasets due to data capture and generalization

on relaxation labelling is that the knowledge of the approximate relative accuracies of the network datasets is not necessary. The only specification required is an appropriate distance to ensure paring and computation of matching probabilities. The consideration of the angles of incidence in relaxation labelling increased the both the precision and recall by an average of 2%, compared to a situation, where only intersections of the same degree are allowed to correspond.

The downside of the algorithm based on relaxation labelling is that network datasets are considered as the only input datasets. An algorithm based on relaxation labelling can be modified to establish correspondences even where the datasets are not referenced in the same coordinate system.

6.4.2 Linear feature matching

Line feature matching based on path matching was carried out to establish one-to-one correspondences between linear features in the subject and road networks. The process was based on Dijsktra shortest path algorithm and some tolerance for the difference between the lengths of the corresponding paths. For paths to be identified as corresponding the criterion used was based on the relative lengths and a value of 1:2 was empirically set. In addition, a Hausdorff distance of 50 m was specified for the maximum allowable distance difference between any two possible corresponding paths.

Figure 6.11 and Figure 6.12 show the results of path matching for Case 1 and case 2 respectively. The paths in subject and reference datasets with correspondences are shown in grey and black lines respectively. The paths from the subject and the reference datasets without correspondences are depicted in grey and black broken lines respectively.

The number of corresponding path determined during the matching process depends on the time period and methods used to capture the datasets. Significant differences, especially in the time of data capture, would be reflected in the number of similar paths. Therefore the number of such similar paths is an indication of the similarity of the datasets.

The path matching approach used in this study makes it possible to quantify the similarity of the datasets from the path correspondences established during the process. In this study, data similarity was determined based on the total length of the corresponding path against the total length of the entire features in the datasets.

Table 6.2 gives the similarity of the datasets. Similarity is taken as the percentage of the total length of the path in the reference dataset with correspondences in the subject dataset. Similarity

Fig. 6.11: Case 1 - corresponding paths established through the path matching algorithm

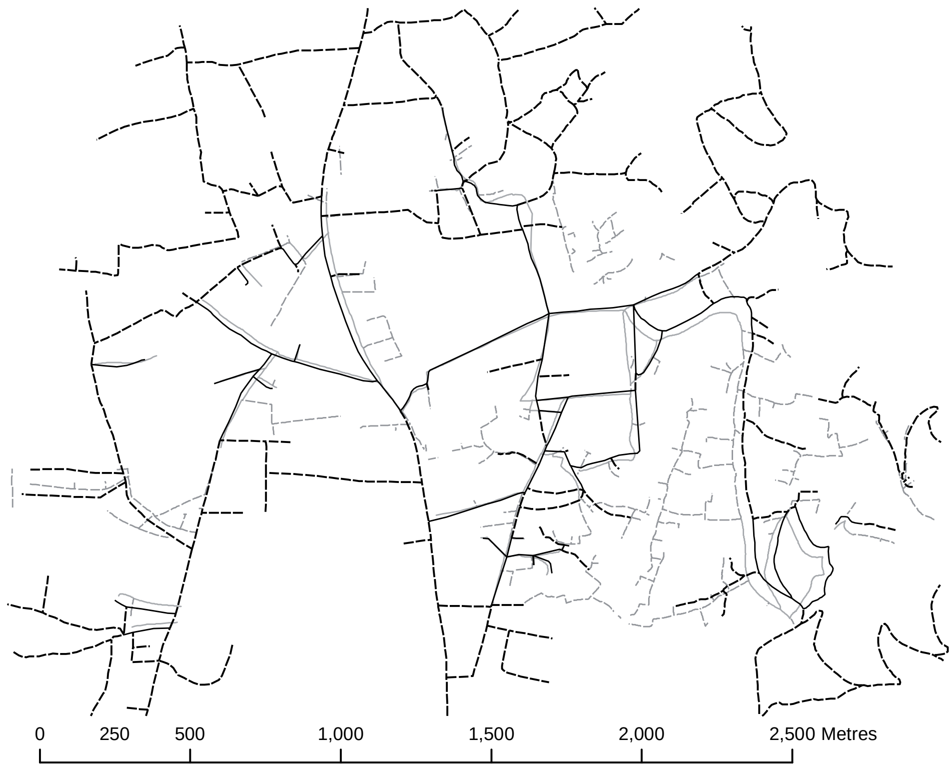


Fig. 6.12: Case 2 - corresponding paths established through the path matching algorithm

as determined here considers only features that are part of the corresponding paths. The similarity values in the table are therefore very conservative.

Table 6.2: Dataset similarity based on corresponding paths

Experiment	Reference (topographical) length (m)		Subject (cadastral) length (m)		Similarity (%)
	All features	*Matched paths*	*All features*	*Matched paths*	
Case 1	72018.5	22670.2 (31.5%)	41777.0	22513.4 (53.8%)	31.5%
Case 2	39263.9	10684.6 (27.2%)	25760.3	10875.3 (42.2%)	27.2%

From the table, similarities are respectively 31.5% and 27.2% for Case 1 and Case 2 experiments. The similarity measures that were obtained to some level indicate the extent to which development is guided by spatial planning. Datasets for Case1 were obtained from an area that could be considered as an urban township; where there is development is guided by spatial planning as evident from the regular arrangement of land parcels (See Figure 6.1) (top part)). On the other hand, most of the subject roads from the cadastral dataset for Case 2 do not actually exist because they have not been developed. This is because the subject dataset for Case 2 is from a rural township, where there is little or no spatial planning at all as evident in the irregular arrangement of land parcels (see Figure 6.1). In this case, the approximate cadastral mapping is not wholly responsible for the low similarity between the datasets.

6.5 Geometric transformation and alignment

The point correspondences obtained during point matching in addition to those obtained during the matching of corresponding paths as vertex correspondences are used to geometrically transform the original cadastral (subject) dataset. The transformation was based on the piece-wise affine transformation considering the local distortions present in the subject dataset. After the geometric transformation of the cadastral dataset, the extent of its geometric quality improvement was evaluated. The improvement of geometric quality of the graphical cadastral dataset was evaluated based on how consistent it is with the mapped road network. This was carried out in three stages. The first stage involved assessment of the linear map accuracy of the cadastral datasets before the improvement. The second stage involves evaluating the extent of improvement in alignment of the cadastral dataset with the reference topographical dataset. The third stage involves the assessment of the extent of improvement in geometric area of the cadastral dataset.

6.5.1 Linear map accuracy

The linear map accuracy of an entire map can be evaluated based on the divergence of the corresponding points obtained during point matching, in which case the accuracy is relative to the reference dataset. Table 6.3 contains summary statistics of the distances between corresponding points. The positional accuracy of the cadastral dataset for Case 1 is better that that of Case 2. The positional accuracy is illustrated by the average of the distance between corresponding points. The positional accuracy of the cadastral dataset for Case 2 is worse as evident in the values of the average and standard deviations. The average accuracy of the Case 2 cadastral dataset was determined as 22.09 m is comparable to the one determined using the knowledge of the processes used (Siriba, 2009), and was about 25 m.

Table 6.3: Summary statistics on identical point (IP) correspondences

	Case 1 IP = 163 (m)	**Case 2** IP = 80 (m)
Min.	0.03	2.11
Max.	33.37	61.07
Range	33.40	63.18
Ave.	4.95	22.09
St. dev	5.28	13.99

Figure 6.13 and Figure 6.14 are visualizations of the planimetric accuracy of cadastral datasets (superimposed) based on the corresponding points. The accuracy maps were generated using an inverse distance weighted (idw) interpolation (Shepard, 1968) of the displacements between the corresponding points.

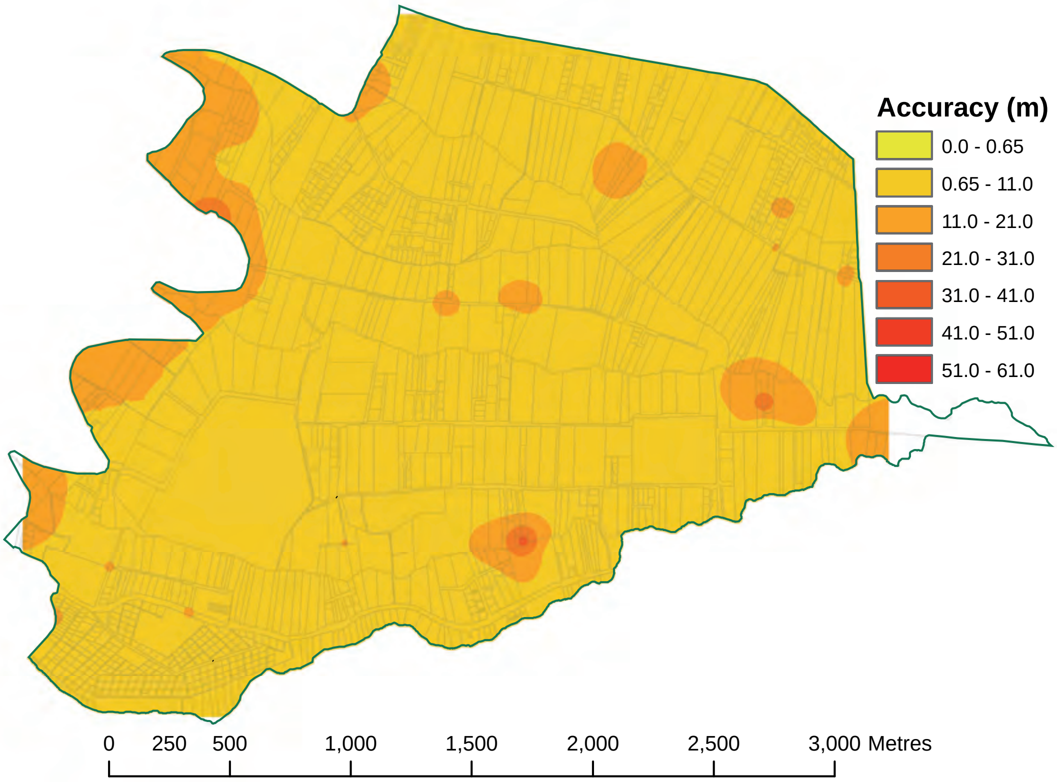


Fig. 6.13: Case 1 – linear map accuracy of the cadastral dataset

6.5.2 Improvement in positional alignment

The mapped road network provides the reference to visually and statistically determine the positional improvement of the land parcels. The corresponding paths in the subject and reference road networks were used to determine the extent of improvement in positional alignment. The road networks are used because they were used as basis for determining the geometric transformation. Figure 6.15 shows both the subject and the reference road network for Case 2. The sections in the boxes numbered 1, 2 and 3 are illustrated at a larger scale in Figure 6.16. In the figure, the original positioning before geometric transformation (top), and, corresponding positioning after geometric accuracy enhancement via transformation is at the bottom of the figure.

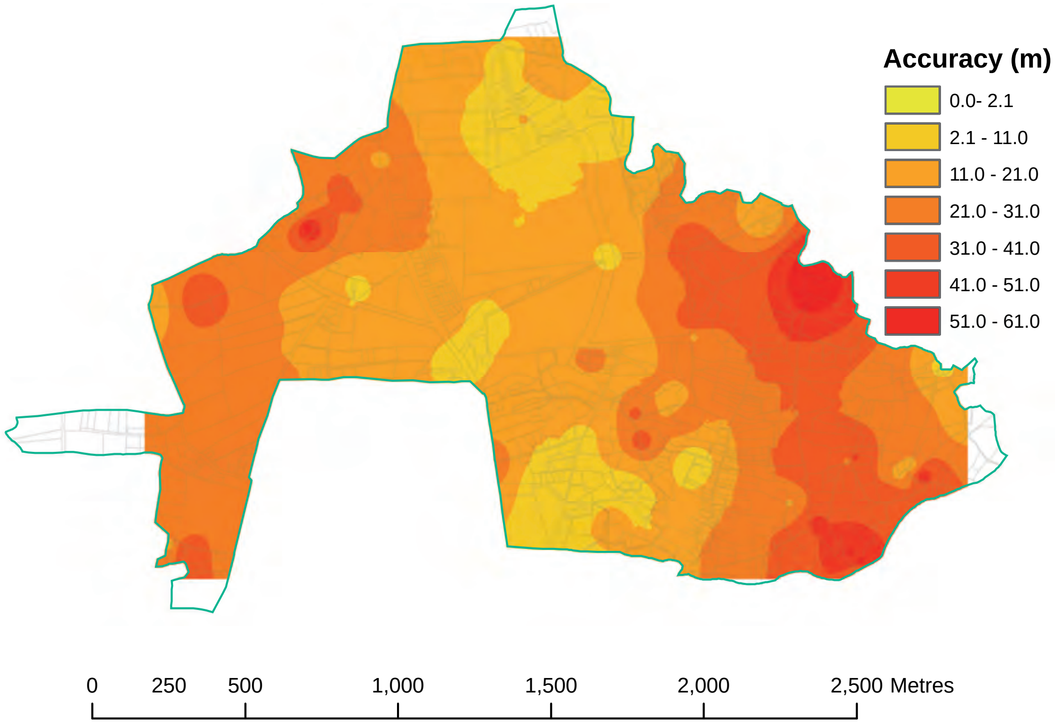


Fig. 6.14: Case 2 – linear map accuracy of the cadastral dataset

Fig. 6.15: Case 2 experimental cadastral after transformation with section in rectangles further highlighted in Figure 6.16

In the bottom row of Figure 6.16, the line segments in the subject road network show an improved alignment with the corresponding line segments in the reference road network. This result satisfies

Fig. 6.16: Superposition of the cadastral (subject) dataset represented by parcels (light grey) and the topographical dataset represented by road network (black) showing extracts from the experimental data. The left, middle and right respectively represent 1, 2 and 3 boxes in Figure 6.15, while the top and bottom parts illustrate the situation before and after transformation respetively

one of the qualities of a good alignment, which is matching the human intuition. The result also shows a significant improvement in the positional accuracy of the subject road network. The bar graph in Figure 6.17 shows the relative alignment in terms of the Hausdorff distance before and after geometric transformation for the corresponding linear features (paths) for Case 2. The grey and the black bars respectively show the situation before and after the transformation. There is however no perfect alignment of the derived (subject) road centrelines that did not have corresponding features in the reference dataset.

Table 6.4 summarizes the results of positional accuracy improvement for the two experimental cases. The accuracy values are given with respect only to the corresponding subject and reference road network features (paths). The improvement of the positional accuracy of the subject dataset was evaluated using the Hausdorff distance of the corresponding linear features in both the subject and reference datasets. In the Case 1, the positional accuracy of linear features in the subject dataset with correspondences in the reference datasets was in the range of 3.60 m to 32.90 m (an average of 8.30 m) was improved to an average of 3.13 m. For Case 2, the average positional error of 28.40 m was improved to an average of 4.00 m. These improvements represent 62% and 86% respectively.

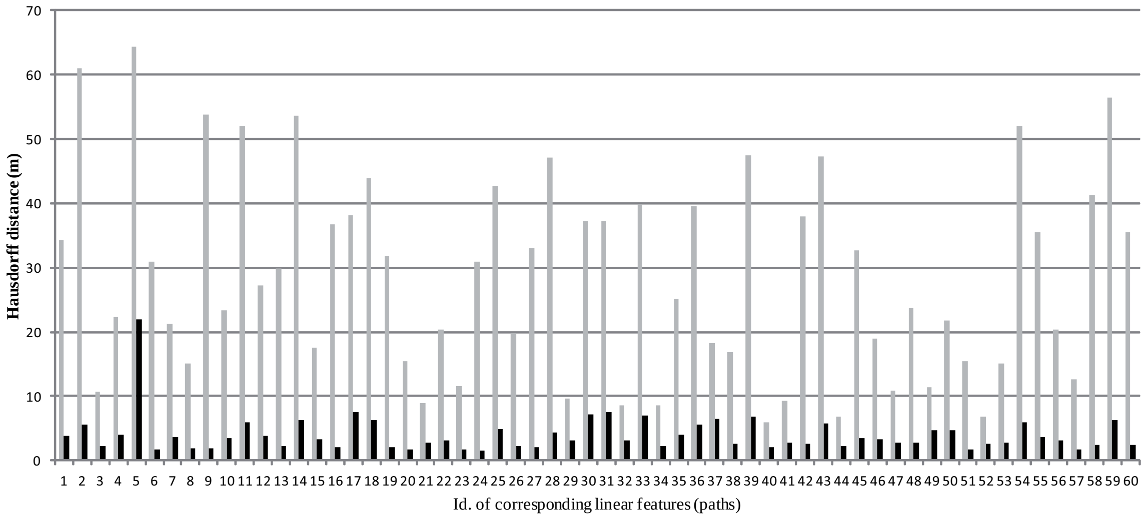


Fig. 6.17: Relative positional accuracy before (in grey) and after (in black) transformation

Following this approach, there is clearly a significant relative improvement in the alignment of features in the subject dataset. The improvement is limited to the number and distribution of the corresponding points used for the transformation. However, in both cases survey-accuracy (< 0.03 m) has not been achieved. This is because the topographic dataset are not themselves survey-accurate.

Table 6.4: Positional accuracy improvement of linear features

	Case 1		**Case 2**	
	Before (m)	After (m)	Before (m)	After (m)
Max.	32.90	16.10	64.30	21.87
Min.	3.60	0.03	5.90	1.54
Mean	8.30	3.13*(62%)*	28.40	4.00*(86%)*
St. dev.	4.20	2.98	15.60	2.91

6.5.3 Improvement in geometric area

Figures 6.18 and 6.19 show the outlines of cadastral (in red outline) datasets before (left) and after (right) the process of geometric quality improvement for sections of datasets from Case 1 and Case 2 respectively. The right hand side of the figures depicts improved alignment of the land parcels.

Although visually the alignment of the land parcels is improved, the determination of the extent of improvement in geometric area for the individual land parcels in the cadastral dataset requires a reference survey-accurate cadastre. However, cadastral datasets of higher accuracy for the experimental datasets were not available. As an alternative, a reference cadastral dataset was digitized from orthophotos, each case consisting of a section of the data extent.

Because of insufficient visible boundary signatures on the orthophoto to allow the digitization of the outlines of individual land parcels, instead the evaluation of the geometric area accuracy improvement was carried out based on groups of land parcel, which were identified clearly and with high positional certainty. Thus, the outlines of the groups of parcels, which are bounded by roads on all their sides, were digitized from the orthophotos. The corresponding land parcel groups

Fig. 6.18: Case 1 - improved geometric alignment of land parcels

Fig. 6.19: Case 2 - improved geometric alignment of land parcels

in the graphical cadastral dataset were obtained analogously by combining adjacent land parcels surrounded by roads. Case 1 and Case 2 respectively had 35 and 27 parcels groups.

The error in each group of parcels in the geometrically improved provisional cadastral dataset was compared to that in the original provisional cadastral dataset to determine whether there was an improvement. Improvement is based on the difference between the initial error in geometric area and the error after local geometric transformation. If the absolute value of the ratio of later to the former error is in the range of 0 - 1 (i.e., $-1.0 \leq x \leq 1.0$), the geometric area is considered positive, otherwise it is negative. Figure 6.20 shows the histogram for the improvement for both cases. The horizontal axis shows the factor by which the remaining error is within the initial error in area. And the vertical axis shows the number of land parcels affected.

In Case 1, a total of 17 out of 35 parcel groups recorded a positive improvement in their geometric area. Collectively, these parcel groups account for 107,972 m^2 out of 219,988 m^2. This implies that there is an overall improvement in area of 50%. In Case 2, a total of 20 out of 27 parcel groups recorded a positive improvement in their geometric area. Collectively, these parcel groups account for 2,221,252 m^2 out of 2,613,731 m^2. This implies that there is an overall improvement in area of

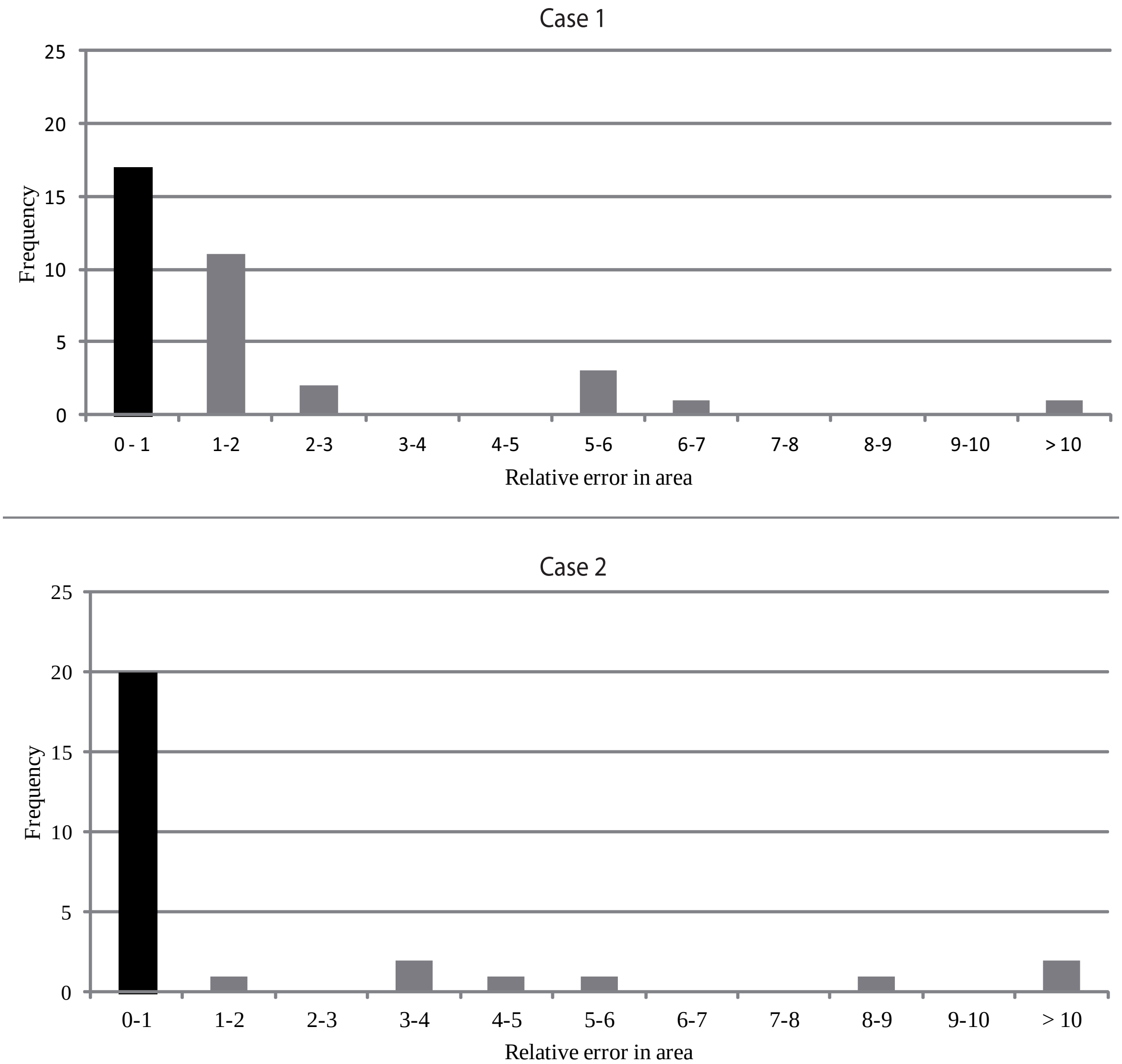


Fig. 6.20: Histograms showing the number of parcel groups exhibiting positive (black) and negative (grey) improvement in geometric area. Case 1 and Case 2 are at the top and bottom respectively

85%. This means there is a more qualitative improvement in Case 2 than in Case 1. This can be attributed to the extreme distortions in the cadastral dataset in Case 2.

6.5.4 Homogenization and adjustment

Local geometric transformation was considered in this study because of the presence of local geometric deformations in the provisional cadastral datasets. The local affine transformation was built on a network of Delaunay triangulation. Affine transformation preserves the co-linearity relations between points as well as the ratios of points along the lines. The area and the angles may not remain the same after transformation. This is the case in particular for objects that overlap more than one triangle.

An example is illustrated in Figure 6.21. The left side of the figure illustrates a rectangular building that overlaps several triangles before transformation. After transformation, the rectangularity of the building is affected as shown on the right hand side of the figure. The main cause of the loss of rectangularity is the disregard of the neighbourhood information between the triangles. That is, during the transformation the topological neighbourhood information is not considered.

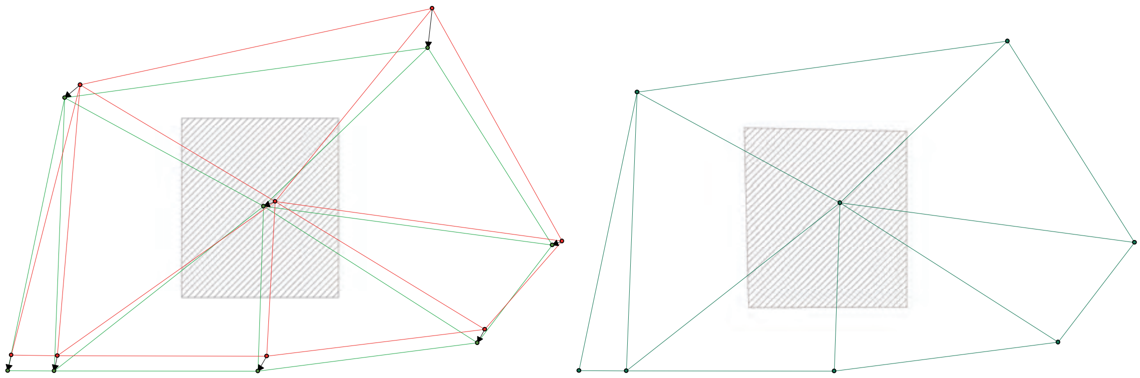


Fig. 6.21: The result of a piece-wise affine transformation showing that rectangularity of features is affected (right)

One possible improvement is to base the local transformation on constrained Delaunay triangulation. This would require the consideration of other objects, like building footprints instead of only land parcel boundaries. The homogenization and adjustment of coordinates approach by Gründig and others (2007) is one possible approach that considers topological neighbourhood information. The approach is also based on Delaunay triangulation. The elasticity of the triangles is achieved by assignment of weights based on the digitization accuracy of the maps. The triangles in the triangulation network carry the neighbourhood topological information. Along the triangle sides, artificial coordinates are subsequently introduced prior to a mathematical adjustment according the least squares method. The method is based on sophisticated mathematical adjustment algorithms and the analysis of coordinates which was not optimal for the case considered in this research. This is mainly because no suitable geometric conditions were present in the cadastral datasets involved. Otherwise, if survey-accurate coordinates are obtained at a later stage via field measurements, then this method should be considered.

7. Conclusions

7.1 Discussion

7.1.1 Approach and objectives

The first main objective was to describe and evaluate the Kenyan cadastre with respect to the requirements of modern land administration and the characteristics of future cadastres. The description of the cadastre helps to highlight the nature, limitations and possible improvements necessary. In particular, it was established that the current cadastral coverage is about 25% and that a greater percentage of that coverage consists of provisional cadastral datasets, which would need to be replaced or improved.

The second main objective of this study was to develop a data conflation technique for geometric quality improvement of provisional cadastral datasets. The motivation for the study was to establish an approach that takes advantage of the availability of high quality topographical datasets to improve the poor geometric quality of existing cadastral datasets. The improvement of poor quality cadastres, for example, to survey-accuracy levels is necessary to have high quality cadastres that can support many land management and other applications.

Different methods that match both linear and polygon features were developed and analyzed in the framework of vector-based data conflation. The overall data conflation approach consists of four main steps. The steps include pre-processing, automatic extraction of road network from a cadastre, establishing feature correspondences through point and linear feature matching algorithms, and finally, transformation of the cadastral dataset via a local geometric transformation technique. The main challenge in the approach is, in most cases, the lack of corresponding ground signatures for features present in cadastral and administrative boundary datasets. The correspondence was established through the road network in the topographical dataset and a similar one derived from the cadastral dataset.

The derivation of the road network from the cadastre dataset took advantage of the topological relationship that exists between the land parcels and the cadastral datasets by gaps. The main assumption is that gaps between the polygons in a cadastral dataset implicitly represent road reserves. Because of historical reasons, roads are not commonly included in cadastres as legal land parcels in some jurisdictions. The geometry difference between the input polygons representing the land parcels and their concave hull contain polygons that represent road parcels. The challenge in this task entailed the creation of the concave hull, which for a given input geometry is often not unique but dependent on the intended application.

It was established however that it is possible to create a unique concave hull especially when the input geometry consists of polygons, without extreme concavities. The concave algorithm developed in this study was based on the concept of Thiessen polygons and the linear referencing. The main advantage of the algorithm is that, only the number of iterations need to be specified, which can be set high enough to ensure that the ultimate concave hull is obtained.

The presence of local geometric distortions in the provisional graphical cadastres required a matching technique that tolerates local distortions. Accordingly, feature matching was carried out using a relaxation-labelling algorithm. High matching precisions of more than 70% were obtained using relaxation labelling. There was an average improvement of precision of 2% when road intersections of different degrees were allowed to match instead of restricting correspondences to intersections of the same degree. A low matching recall of 50% was obtained for the cadastral dataset of poorer positional quality. This can be attributed to the little similarity between the topographical road network and the road network derived from the cadastral dataset. The limited correspondence between cadastral content and the actual ground situation is because the cadastral dataset is not up-to-date and the disparity between planning and the actual development.

Piece-wise affine transformation was used to transform the cadastral dataset considering the presence of local geometric distortions. Despite some distortion of features across adjacent triangles, which cannot however be avoided, there was generally an improvement in both positional alignment and geometric area. It was also observed that improvement in geometric area could not be predicted from the distribution of point correspondences. This is because the adopted transformation model attempts to maintain conformity and linearity at the expense of area. Generally, cadastral dataset of initial poorer geometric quality showed more improvement both in position and area.

7.1.2 Significance of the findings

The findings from this study contribute to a better understanding of the Kenyan cadastre and the current literature on spatial data integration. The outline of the Kenyan cadastre improves the understanding that the quality of a cadastre is largely influenced by the legal traditions of a country. The main lesson to be learnt from the outline of the Kenyan cadastre is that substituting something quickly available (i.e., the provisional maps) for something really intended is costly in the long run because the provisional maps may not be suitable for most applications because of their accuracy limitations.

With regard to data integration, the contribution of this study is threefold. First, given the little work on spatial data integration involving cadastres and boundaries, this study has developed a methodology which can be used and developed further to conflate cadastral datasets with topographical datasets. This approach can be used in jurisdictions where legacy provisional graphical cadastres are still being used and where roads are not explicitly contained in the cadastres.

Secondly, in the derivation of road polygons from a cadastral datasets that does not explicitly contain roads as land parcels, this study presented an algorithm for creating concave hulls. The algorithm produces a predictable and unique hull for polygon features and as long as there are no extreme concavities in the input polygons. This is significant because concave hulls are not commonly used in geo-processing operations as their counterpart convex hull. Therefore, this research contributes to the on-going search for an algorithm for creating unique concave hulls.

Finally, in the presence of local distortions, a feature matching technique was implemented that exploits relaxation labelling technique. The significance of this approach is that the specification of the search radius is not critical, and high matching precisions are possible.

7.2 Outlook

7.2.1 Limitations and possible improvements

Road networks were used as the basis for establishing correspondence between the cadastral and topographical datasets in the adopted approach. Therefore the extent of point correspondences depends on the similarity of the road networks. It was also realized that contrary to the assumption, some of the gaps between the land parcels represented riparian reserves instead of road reserves.

It is therefore important to understanding how roads and rivers are considered in the cadastres. Other features like river networks and fence lines could be considered in future research to establish correspondences between cadastral datasets and topographical datasets. Because there are no conceptual differences, the same approach can be used.

The adopted local geometric transformation was limited to number and distribution of corresponding points. Survey accuracies could also not be obtained because the topographical datasets were also limited in accuracy. It was also not possible to use the individual vertices of the input parcel polygons because there were no corresponding polygons; the extent of improvement is therefore limited. It is recommended to use the data conflation approach alongside other options, like additional field measurements.

In the evaluation of the extent of improvement of the geometric area of the parcels of land, the reference areas were obtained from parcel polygons digitized from the corresponding orthophotos. Although this might be mathematically justifiable, it will be interesting to compare the values of the new geometric areas after transformation and those contained in the land registers, which were not available during the course of this research. This is important because, unlike ordinary topographical objects whose geometric area can change due to geometric transformations, the change in geometric area of land parcels, particularly those that have been legally registered should be treated with a lot of caution. This is necessary to reduce any possible legal complications that may arise. This approach is however limited to provisional cadastral datasets, where the change of geometric areas might not pose much legal challenges. In case, this approach is considered for cadastres that consists of land parcels that have been legally registered, the registered areas should be included in a mathematical adjustment process as possible constraints.

7.2.2 Implications

From the outline of the Kenyan cadastre, it was established that, in its present status, the cadastre is not readily suitable for modern land administration. This is mainly because the cadastre is still analogue and incomplete with a nationwide coverage of about 25%. Moreover, the coverage is not continuous; instead, it consists of maps of different positional accuracies. The cadastral maps that are part of the cadastre are based on different coordinate systems (UTM, Cassini-Soldner and local coordinates) with some of them, particularly the Preliminary Index Diagrams (PIDs), are not geometrically referenced to any coordinate system.

It is generally recommended that cadastral coverage be extended to un-mapped areas using appropriate cadastral mapping techniques that yield intended accuracy results. In addition to using orthorectified satellite imagery by professionals to improve land tenure and cadastral coverage, crowd sourcing is an emerging concept that could be utilized too in land administration. Crowd sourcing as a possible solution to the security of tenure gap is about establishing a partnership between professionals and citizens to encourage citizens to directly capture and maintain information about their land rights (Mclaren, 2011).

Existing analogue cadastral maps should be digitized and harmonized to a common national coordinate system to create a homogeneous cadastral framework. Finally, existing provisional cadastral datasets should be improvement using, for example, the approach developed in this thesis. This would constitute one of the primary tasks in creating a homogeneous cadastral framework as the core of modern land management.

The data conflation approach developed in this study considered cadastral and topographical datasets at the same geometric scales using road network as the basis for quality improvement. Although it is possible to match networks at different scales, the possible reduction in the number of corresponding points that are established during the matching process is a potential problem. This is because datasets at smaller scales have fewer features because most features are eliminated during the generalization process. To ensure that survey-accuracy levels or nearly so are obtained, the topographical datasets should be at the same or larger scale than the cadastral datasets. This implies that the approach developed in this thesis could be limited to the current large-scale topographical coverage or large-scale topographical mapping should be extended as a first step. Another implication of using this approach would be to update both the cadastral and topographical datasets to ensure more and improved correspondence between the datasets.

Bibliography

Adriano, M. and Yasmina, S. M. (2007). Concave hull : A K-nearest Neighbours Approach for the Computation of the Region occupied by a Set of Points. *International Conference on Computer Graphics Theory and Applications*, Barcelona, Spain.

Agouris, P., Stefanidis, A. and Gyftakis, S. (2001). Differential Snakes for Change Detection in Road Segments. *Photogrammetric Engineering and Remote Sensing*, 67(12):1391–1399.

Al-Ruzouq, R. and Dimitrova, P. (2006). Photogrammetric Techniques for Cadastral Map Renewal. *FIG XXIII Congress*, Munich, Germany.

Arvanitis, A. and Koukopoulou, S. (1999). Managing Data during the Update of the Hellenic Cadastre. *FIG Commission 3 Annual Meeting and Seminar*, FIG Commission 3, Budapest, Hungary.

(ASPRS) American Society for Photogrammetry and Remote Sensing (1990). ASPRS Accuracy Standards for Large-Scale Maps. *Photogrammetric Engineering and Remote Sensing*, 56(7):1068–1070.

Barbehenn, M. (1998). A Note on the Complexity of Dijkstra's Algorithm for Graphs with Weighted Vertices. *IEEE Transactions on Computers*, 47(2):263.

Bennett, R., Rajabifard, A., Kalantari, M., Wallace, J. and Williamson, I. (2010). Cadastral Futures: Building a New Vision for the Nature and Role of Cadastres. *XXIV FIG Congress*, Sydney, Australia.

Binns, A., Rajabifard, A., Collier, P. A. and Williamson, I. (2003). Issues in Defining the Concept of a Marine Cadastre for Australia. *FIG and University of New Brunswick Meeting on Marine Cadastre Issues*, University of New Brunswick, Canada.

Blixt, G. (2008). Quality Improvement to Cadastral Information in Sweden. *FIG Working Week*, Stockholm, Sweden.

Boljen, J. (2009). Verbesserung der ALKIS Geometrie durch Digitale Orthophotos ATKIS - DOP20. *Zeitschrift für Vermessungswesen*, 134(1):45–52.

Brenner, C. (2000). *Dreidimensionale Gebäuderekonstruktion aus digitalen Oberflächenmodellen und Grundrissen*. Ph.D. thesis, Universität Stuttgart, Stuttgart, Germany.

Butenuth, M. (2008). *Network Snakes*. Ph.D. thesis, Leibniz Universität Hannover, Hannover, Germany.

Butenuth, M., von Gösseln, G., Tiedge, M., Heipke, C., Lipeck, U. and Sester, M. (2007). Integration of Heterogeneous Geospatial Data in a Federated Database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(2007):328–346.

Casado, M. L. (2006). Some Basic Mathematical Constraints for the Geometric Conflation Problem. M. Caetano and M. Painho (eds.), *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisbon, Portugal.

Chang, D., Dooley, L. and Tuovinen, J. E. (2001). Gestalt Theory in Visual Screen Design: A New Look at an old Subject. *Proceedings of the Seventh World Conference on Computers in Education, Computers in education 2001: Australian topics*, Copenhagen, Denmark, pages 5–12.

Chen, C.-C., Knoblock, C. A. and Shahabi, C. (2008). Automatically and Accurately Conflating Raster Maps with Orthoimagery. *Geoinformatica*, 12:377–410.

Chrisman, N. (2011). In Search of the Data Effect: When Best Available Data Overrides the Real Requirement. *14th AGILE Conference on GIScience*, Utrecht, The Neatherlands.

Chui, H. and Rangarajan, A. (2003). A New Point Matching Algorithm for Non-rigid Registration. *Computer Vision and Image Understanding*, 89:114–141.

Cobb, M. A., Chung, M. J., Foley, H., Petry, F. E., Shaw, K. B. and Miller, H. V. (1998). A Rule-based Approach for the Conflation of Attributed Vector Data. *GeoInformatica*, 2(1):7–35.

Cohen, I., Ayache, N. and Sulger, P. (1992). Tracking Points on Deformable Objects Using Curvature Information. *Proceedings of the 2nd European Conference on Computer Vision*, Santa Margherita Ligure, Italy.

Croitoru, A. and Doytsher, Y. (2003). Accounting for Discontinuities in Cadastral Data Accuracy: Toward a Patch-Based Approach. *FIG Working Week 2003*, Paris, France.

Curtin, K. M., Nicoara, G. and Arifin, R. R. (2007). A Comprehensive Process for Linear Referencing. *Urban and Regional Information Systems Association*, 19(2):23–32.

Dale, P. F. (1979). Photogrammetry and Cadastral Surveys within the Commonwealth. *Photogrammetric Record*, 9(53):621–631.

Dale, P. F. (2006). Reflections on the Cadastre. *Survey Review*, 38(300):491–498.

Deliiska, B. (2007). Thesaurus and Domain Ontology of Geoinformatics. *Transactions in GIS*, 11(4):637–651.

Demir, O. and Çoruhlu, Y. E. (2008). The Graphical Cadastre Problem in Turkey: The Case of Trabzon Province. *Sensors*, 8(9):5560–5575.

Devillers, R., Bédard, Y., Jeansoulin, R. and Moulin, B. (2007). Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for use of Spatial Data. *International Journal of Geographical Information Science*, 21(3):261–282.

Devogele, T. (2002). A New Merging Process for Data Integration Based on the Fréchet Distance. *10th International Symposium on Spatial Data Handling (SDH)*, Ottawa, Canada, pages 167–181.

Devogele, T., Parent, C. and Spaccapietra, S. (1998). On Spatial Database Integration. *International Journal of Geographical Information Science*, 12(4):335–352.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1:269–271.

Donato, G. and Belongie, S. (2002). Approximate Methods for Thin Plate Spline Mappings and Principal Warps. *7th European Conference on Computer Vision Part III Lecture Notes in Computer Science*, Copenhagen, Denmark, volume 2352, pages 21–31.

Döner, F., Thompson, R., Stoter, J., Lemmen, C., Ploegerd, H., van Oosterom, P. and Zlatanova, S. (2010). 4D Cadastres: First Analysis of Legal, Organizational, and Technical Impact with a Case Study on Utility Networks. *Land Use Policy*, 27(4):1068–1081.

Douglas, D. H. and Peucker, T. K. (1973). Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature. *Canadian Cartographer*, 10(2):112–122.

Duckham, M., Kulik, L., Worboys, M. and Galton, A. (2008). Efficient Generation of Simple Polygons for Characterizing the Shape of a Set of Points in the Plane. *Pattern Recognition*, 41:3224–3236.

Edelsbrunner, H., Kirkpatrick, D. and Seidel, R. (1983). On the Shape of a Set of Points in the Plane. *Information Theory, IEEE Transactions*, 29(4):551–559.

Enemark, S., Williamson, I. and Wallace, J. (2005). Building Modern Land Administration Systems in Developed Economies. *Spatial Science*, 50(2):51–68.

Felus, Y. A. (2007). On the Positional Enhancement of Digital Cadastral Maps. *Survey Review*, 39(306):268–281.

Fichtinger, A., Klien, E. and Giger, C. (2009). Challenges in Geospatial Data Harmonisation: Examples and Approaches from the HUMBOLDT project. *12th AGILE Conference on GIScience*, Hannover, Germany.

FIG (International Federation of Surveyors) (1995). Statement on the Cadastre. FIG Publication No. 11, http://www.fig.net/pub/figpub/pub11/figpub11.htm (accessed on 02.08.2011).

FIG (International Federation of Surveyors) (1996). United Nations Interregional Meeting of Experts on the Cadastre - The Bogor Declaration. FIG Publication No. 13A, http://www.fig.net/pub/figpub/pub13a/figpub13a.htm (accessed on 27.07.2010).

FIG (International Federation of Surveyors) (1999). The Bathurst Declaration on Land Administration for Sustainable Development. FIG Publication No. 21, http://www.fig.net/pub/figpub/pub21/figpub21.htm (accessed on 4.08.2010).

FIG (International Federation of Surveyors) (2010). The Social Tenure Domain Model - A Pro-Poor Land Tool. FIG Publication No. 52, http://www.fig.net/pub/figpub/pub52/figpub52.pdf (accessed on 02.08.2011).

Filin, S., Kulakov, Y. and Doytsher, Y. (2005). Application of Airborne Laser Technology to 3D Cadastre. *FIG Working Week 2005 and GSDI-8*, Cairo, Egypt.

Galton, A. and Duckham, M. (2006). What is the Region Occupied by a Set of Points? *GI Science Series, Lecture Notes in Computer Science*, 4197:81–98.

Gielsdorf, F., Gründig, L. and Aschoff, B. (2004). Positional Accuracy Improvement – A Necessary Tool for Updating and Integration of GIS Data. *FIG Working Week*, Athens, Greece.

Gillman, D. W. (1985). Triangulations for Rubber-Sheeting. *AUTOCARTO 7*, Washington, DC.

GoK (Government of Kenya) (1963). *The Registered Land Act Chapter 300 of the Laws of Kenya*. Government Printer, Nairobi, Kenya.

GoK (Government of Kenya) (2010). *The Constitution of Kenya*. Government Printer, Nairobi, Kenya, 194 pages.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199–220.

Hangouët, J. (1995). Computation of the Hausdorff Distance between Plane Vector Polylines. *Auto Carto 12*, pages 1–10.

Haunert, J. H. and Sester, M. (2008). Area Collapse and Road Centerlines Based on Straight Skeletons. *GeoInformatica*, 12(2):169–191.

Henssen, J. L. G. and Williamson, I. P. (1990). Land Registration, Cadastre and Its Interaction - A World Perspective. *FIG XIX Congress*, Helsinki, Finland.

Holl, J. T., Rabley, P., Monacelli, M. and Ewan, D. (2010). The Earthen Vessel: Land Records in the United States. *Property Records Industry Association (PRIA) Annual Winter Conference*, Washington, DC.

Hope, S., Gordini, C. and Kealy, A. (2008). Positional Accuracy Improvement: Lessons Learnt from Regional Victoria, Australia. *Survey Review*, 40(307):29–42.

ISCGM (International Steering Committee for Global Mapping) (2002). What is Global Map? Http://www.iscgm.org/cgi-bin/fswiki/wiki.cgi?page=Application

ISO (International Organization for Standardization Technical Committee 211(ISO/TC 211)) (2010). *ISO/CD 19157 Geographic Information – Data quality*. ISO (Committee Draft), Geneva, Switzerland, http://www.gdmc.nl/oosterom/ (accessed on 05.05.2011).

ISO (International Organization for Standardization Technical Committee 211(ISO/TC 211)) (2011). *ISO 19152 Draft International Standard (DIS): Geographic information - Land Administration Domain Model (LADM)*. ISO (Committee Draft), Geneva, Switzerland, http://www.gdmc.nl/oosterom/ (accessed on 05.05.2011).

Jarroush, J. and Even-Tzur, G. (2006). Re-instating Parcel Boundaries by Real Time Kinematik GPS. *Survey Review*, 38(302):643–653.

Kalantari, M., Rajabifard, A., Wallace, J. and Williamson, I. (2006). A New Vision on Cadastral Data Models. *XXIII FIG Congress*, Munich, Germany.

Karns, D. (1981). Photogrammetric Cadastral Surveys and GLO Corner Restoration. *Photogrammetric Engineering and Remote Sensing*, 47(2):193–198.

Kass, M., Witkin, A. and Terzopoulos, D. (1988). Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331.

Kaufmann, J. and Steudler, D. (1998). Cadastre 2014: A Vision for a Future Cadastral System. *XXI FIG Congress*, Brighton, UK.

Kieler, B., Huang, W., Haunert, J. H. and Jiang, J. (2009). Matching of River Datasets of Different Scales. M. Sester, L. Benard and V. Paelke (eds.), *Advances in GIScience: Proceedings of the 12th AGILE Conference - Lecture Notes in Geoinformation and Cartography*, 2009, Springer-Verlag Berlin Heidelberg, pages 135–154.

Kieler, B., Sester, M., Wang, H. and Jiang, J. (2007). Semantic Data Integration: Data of Similar and Different Scales. *Photogrammetrie Fernerkundung Geoinformation*, 6:447–457.

Kim, S.-H., Alattar, A. M. and Jang, J. W. (2008). Snake-based Contour Detection for Objects with Boundary Concavities. *Optical Engineering*, 47(3):037002–1–037002–7.

Klebanov, M. and Forrai, J. (2010). Implementing a Coordinate-Based Cadastre in Israel: Experience and Perspectives. *XXIV FIG congress*, Sydney, Australia.

Konecny, G. (2008). Economic Considerations for Phototogrammetric Mapping. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII(6A):207–211.

Konecny, G., Lanzon, J. and Mohammed, A. S. (2010). A Geocoded Cadastral Fabric as a Pre-condition for a Sustainable Land Management System. *FIG Commissions 5 and 6 Workshop on Innovative Technologies for an efficient geospatial management of earth resources*, Ust-Kamenogorsk, Kasakhstan.

Lawrence, J. C. D., Humphries, G. J., Simpson, S. R., Gaitta, G. M., Nottidge, C. P. R. and MacArthur, J. D. (1966). Report of the Mission on Land Consolidation and Registration in Kenya. Technical report, Government Printer, Nairobi, Kenya.

Lemmen, C., Augustinus, C. and van Oosterom, P. (2009). The Social Tenure Domain Model - A Pro-Poor Land Rights Recording System. *GIM International*, 23(10), http://www.gim-international.com/issues/articles/id1430-The_Social_Tenure_Domain_Model.html (accessed on 20.10.2011).

Lemmen, C. and van Oosterom, P. (2003). 3D Cadastres. *Computers, Environment and Urban Systems*, 27(4):337–434.

Longley, P. A., Goodchild, M. F., Maguire, D. J. and Rhind, D. W. (2005). *Geographic Information Systems and Science*. Wiley, West Sussex, England, 517 pages.

Marcus, M. and Khan, N. (1959). A Note on the Hadamard Product. *Canadian Mathematical Bulletin*, 2(2):81–83.

Mbaria, C. (2002). Establishment of NDTDB at the Survey of Kenya. *2nd Kenya NSDI Workshop*, pages 2–9.

Mclaren, R. (2011). Crowdsourcing Support of Land Administration- A Partnership Approach. *International Federation of Surveyors - Article of the Month*, pages 1–12.

Mikkonen, K. and Corker, I. (2000). Using Digital Orthophotos to Support Land Registration. *Proceedings of the Twentieth Annual ESRI User Conference*, San Diego, California.

Mohammadi, H., Rajabifard, A. and Williamson, I. P. (2009). Enabling Spatial Data Sharing through Multi-source Spatial Data Integration. *The 11th Global Spatial Data Infrastructure (GSDI) Conference*, Rotterdam, The Netherlands.

MoL (Ministry of Lands) (2010). Achievements of Land Adjudication Surveys. Http://www.ardhi.go.ke/ (accessed on 13.08.2010).

Morgenstern, D., Prell, K. M. and Riemer, H. G. (1989). Digitisation and Geometrical Improvement of Inhomogeneous Cadastral Maps. *Survey Review*, 30(234):149–159.

Mueller, M. (2008). Transformations of Cadastral Descriptions with Incomplete Information into Maps. *Transactions in GIS*, 12(1):83–101.

Mulaku, G. C. and McLaughlin, J. D. (1996). Concepts for Improving Property Mapping in Kenya. *South African Journal of Surveying and Mapping*, 23(4):211–216.

Mustière, S. (2006). Results of Experiments on Automated Matching of Networks at Different Scales. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVI(Part 2/W40):92–100.

Mustière, S. and Devogele, T. (2008). Matching Networks with Different Levels of Detail. *GeoInformatica*, 12(4):435–453.

Mwenda, J. N. (2001). Spatial Data in Land Tenure Reform with Special Reference to Kenya. *International Conference on Spatial Information for Sustainable Development*, Nairobi, Kenya.

Navratil, G. (2011). Cadastral Boundaries: Benefits of Complexity. *Urban and Regional Information Systems Association*, 23(1):21–27.

Ng'ang'a, S., Sutherland, M., Cockburn, S. and Nichols, S. (2004). Toward a 3D Marine Cadastre in Support of Good Ocean Governance: A Review of the Technical Framework Requirements. *Computers, Environment and Urban Systems*, 28(2004):443–470.

Njuki, A. K. (1999). Current Land Survey Laws, Regulations and Requirements and their Impact on the Production and Availability of Large-scale Maps in Kenya. *Regional Workshop on Land Survey and Large-Scale Mapping in Support of Settlements Planning, Land Development and Management*, Nairobi, Kenya.

Njuki, A. K. (2000). The Status of Land Survey and Large Scale Mapping in Kenya. *Land Survey and Large Scale Mapping in Sub-Saharan Africa*, United Nations Centre for Human Settlement (HABITAT), Nairobi, Kenya, pages 48–53.

Njuki, A. K. (2001). Cadastral Systems and their Impact on Land Administration in Kenya. *International Conference on Spatial Information for Sustainable Development*, Nairobi, Kenya.

NRC (National Research Council) (1980). *Need for a Multi-purpose Cadastre.* National Academy Press, Washington DC, 112 pages, http://www.nap.edu/openbook.php?record_id=10989 (accessed on 16.08.2010).

Nyerges, T. L. (1990). Locational Referencing and Highway Segmentation in a Geographic Information System. *Institute of Transportation Engineers*, 60(3):27–31.

Odhiambo, W. and Nyangito, H. (2003). Land Laws and Land Use in Kenya: Implications for Agricultural Development. KIPPRA Discussion Paper No. 15.

OGC (Open Geospatial Consortium) (2002). Topic 12: OpenGIS Service Architecture. G. Percivall (ed.), *The OpenGIS Abstract Specification*, Open Geospatial Consortium, http://portal.opengeospatial.org/files/index.php?artifact_id=1221 (accessed on 08.11.2010).

Olima, W. H. A. and Obala, L. M. (1998). The Effect of Existing Land Tenure Systems on Urban Land Development: A case study of Kenya's secondary towns, with emphasis on Kisumu. *Habitat International*, 23(1):113–124.

Parent, C. and Spaccapietra, S. (2000). Database Integration: The Key to Data Interoperability. M. P. Papazoglou, S. Spaccapietra and Z. Tari (eds.), *Advances in Object-Oriented Data Modeling*, The MIT Press.

Park, J. (2001). Schema Integration Methodology and Toolkit for Heterogeneous and Distributed Geographic Databases. *Journal of the Korea Industrial Information Systems Society*, 6:51–64.

Petrie, G. (1990). Photogrammetric Methods for Data Acquisition for Terrain Modelling. G. Petrie and T. Kennie (eds.), *Terrain Modelling in Surveying and Civil Engineering*, Whittles Publishing Services, pages 26–47.

Rajabifard, A. and Williamson, I. P. (2004). SDI Development and Capacity Building. *7th GSDI Conference*, Bangalore, India.

Ranade, S. and Rosenfeld, A. (1980). Point Pattern Matching by Relaxation. *Pattern Recognition*, 12:269–175.

Ruiz, J. J., Ariza, F. J., Ureña, M. A. and Blázquez, E. B. (2011). Digital Map Conflation: A Review of the Process and a Proposal for Classification. *International Journal of Geographical Information Science*, 29(9):1439–1466.

Saalfeld, A. (1988). Conflation: Automated Map Compilation. *International Journal of Geographical Information Systems*, 2(3):217–228.

Savopol, F. and Armenakis, C. (2002). Merging of Heterogeneous Data for Emergency Mapping: Data Integration or Data Fusion? *Symposium on Geospatial Theory, Processing and Applications*, Ottawa, Canada.

Sester, M., von Gösseln, G. and Kieler, B. (2007). Identification and Adjustment of Corresponding Objects in Data sets of Different Origin. *10th AGILE International Conference on Geographic Information Science*, Aalborg University, Denmark.

Sheeren, D., Mustière, S. and Zucker, J.-D. (2004). How to Integrate Heterogeneous Spatial Databases in a Consistent Way? *Lecture Notes in Computer Science - Advances in Databases and Information Systems*, 3255(2004):364–378.

Shepard, D. (1968). A Two-Dimensional Interpolation for Irregularly-spaced Data. *ACM National Conference*, pages 517–524.

Siriba, D. (2009). Positional Accuracy Assessment of a Cadastral Dataset based on the Knowledge of the Process Steps used. *12th AGILE Conference on GIScience*, Hannover, Germany.

Siriba, D. N., Voss, W. and Mulaku, G. C. (2011). The Kenyan Cadastre and Modern Land Administration. *Zeitschrift für Vermessungswesen*, 136(3):177–186.

Song, W., Haithcoat, T. L. and Keller, J. M. (2006). A Snake-based Approach for TIGER Road Data Conflation. *Cartography and Geographic Information Science*, 33(4):287–298.

Song, W., Keller, J. M., Haithcoat, T. L. and Davis, C. H. (2011). Relaxation-based Point Feature Matching for Vector Map Conflation. *Transactions in GIS*, 15(1):43–60.

Srebro, H. and Shoshani, M. (2007). The Order of Precedence of Boundary Delimitations. *FIG Working Week*, Hong Kong, China.

Steinberg, G. (2000). Implementation of a Legal Digital Cadastre in Israel. *FIG Working Week*, Seoul, Korea.

Steudler, D. (2010). Cadastral Template: A Worldwide Comparison of Cadastral Systems. Http://www.cadastraltemplate.org/ (accessed on 16.07.2010).

Sunday, D. (2006). The Convex Hull of a 2D Point Set or Polygon. Softsurfer, http://softsurfer.com/Archive/algorithm_0109/algorithm_0109.htm (accessed on 03.03.2011).

Swensson, E. (2010). Quality Improvement of the Cadastral Information in Sweden. *XXIV FIG Congress*, Sydney, Australia.

Swynnerton, R. J. M. (1954). A Plan to Intensify the Development of African Agriculture in Kenya. Technical report, Government Printer, Nairobi, Kenya.

Thomas, J. (2004). Modern Land Consolidation - Recent Trends on Land Consolidation in Germany. *Symposium on Modern Land Consolidation*, Clermont-Ferrand, France.

Ting, L. and Williamson., I. P. (1999). Cadastral Trends: A Synthesis. *The Australian Surveyor*, 4(1):46–54.

Tuladhar, A. M. (1996). Spatial Cadastral Boundary Concept and Uncertainty in Parcel based Information System. *International Archives of Photogrammetry and Remote Sensing*, XXXI(Part 3B):890–893.

Uitermark, H. T., van Oosterom, P. J., Mars, N. J. and Molenaar, M. (2005). Ontology-based Integration of Topographic Data Sets. *International Journal of Applied Earth Observation and Geoinformation*, 7((2006)):97–106.

UN (United Nations) (1990). *Guidelines for the improvement of Land-Registration and Land- Information systems in Developing Countries (with special reference to English-speaking countries in Eastern, Central and Southern Africa)*. United Nations Centre for Human Settlement (HABITAT), Nairobi, Kenya.

UN (United Nations) (1992). *Agenda 21*. United Nations, Rio de Janerio, Brazil, http://www.un.org/esa/dsd/agenda21/ (accessed on 27.09.2010).

UNECA (United Nations Economic Commision for Africa) (2003). Land Tenure Systems and Sustainable Development in Southern Africa. Technical Publication, http://www.uneca.org/eca_resources/publications/srdcs/ (accessed on 20.08.2010).

Vaccari, L., Shvaiko, P. and Marchese, M. (2009). A Geo-service Semantic Integration in Spatial Data Infrastructures. *International Journal of Spatial Data Infrastructures Research*, 4:24–51.

van Oosterom, P., Lemmen, C., Ingvarsson, T., van der Molen, P., Ploeger, H., Quak, W., Stoter, J. and Zevenbergen, J. (2006). The Core Cadastral Domain Model. *Computers, Environment and Urban Systems*, 30(5):627–660.

Vassilopoulou, S., Hurni, L., Dietrich, V., Baltsavias, E., Pateraki, M., Lagios, E. and Parcharidis, I. (2002). Orthophoto Generation using IKONOS Imagery and High-Resolution DEM: A Case Study on Volcanic Hazard Monitoring of Nisyros Island, Greece. *International Archives of Photogrammetry and Remote Sensing*, 57:24–38.

Vijayaraj, V., O'hara, C. G. and Younan, N. H. (2004). Pansharpening and Image Quality Interface. *Geoscience and Remote Sensing Symposium - IEEE International*, 4:2558–2560.

Villa, P., Reitz, T. and Gomarasca, M. A. (2007). The HUMBOLDT Project: Implementing a Framework for Geo-spatial Data Harmonization and Moving Towards an ESDI. M. Gomarasca (ed.), *GeoInformation in Europe*, pages 29–36.

Volz, S. (2005). Data-Driven Matching of Geospatial Schemas. A. Cohn and D. Mark (eds.), *International Conference on Spatial Information Theory (COSIT 2005)*, Ellicottville, NY, Lecture Notes in Computer Science 3693, pages 115–132.

Vonderohe, A. P., Chou, C. L., Sun, F. and Adams, T. (1997). A Generic Data Model for Linear Referencing Systems. *National Cooperative Highway Research Results Digest 218*, Transportation Research Board, National Academy Press, Washington, DC.

Vosselman, G. (1992). *Relational Matching*, volume 628. Lecture Notes in Computer Science, Springer Verlag, 190 pages.

Walter, V. and Fritsch, D. (1999). Matching Spatial Datasets: A Statistical Approach. *International Journal of Geographical Information Science*, 13(5):445–473.

Wanyoike, P. K. (2001). The Role of the Registry Index Map (RIM) in Land Management in Kenya. *International Conference on Spatial Information for Sustainable Development*, Nairobi, Kenya.

Wayumba, G. (2004). A Review of Special Land Tenure Issues in Kenya. *Expert Group Meeting on Secure Land Tenure: New Legal Frameworks and Tools*, UNEP, Nairobi, Kenya.

Weissman, R. (1971). Photogrammetry Applied to Cadastral Survey in Switzerland. *Photogrammetric Record*, 7(37):5–15.

Williamson, I., Enemark, S., Wallace, J. and Rajabifard, A. (2010a). *Land Administration for Sustainable Development*. ESRI Press Academic, Redlands, California, 487 pages.

Williamson, I., Rajabifard, A. and Holland, P. (2010b). Spatially Enabled Society. *XXIV FIG Congress*, Sydney, Australia.

Williamson, I. P. (1983). Cadastral Survey Techniques in Developing Countries with Particular Reference to Thailand. *The Australian Surveyor*, 31(8):567–581.

Williamson, I. P. (1985). Cadastres and Land Information Systems in Common Law Jurisdictions. *Survey Review*, 28(217):114–129.

Williamson, I. P. (1996). Appropriate Cadastral Systems. *Australian Surveyor*, 41(1):35–37.

Xu, Z. and Lee, Y. C. (2002). Semantic Heterogeneity of Geodata. *Symposium on Geospatial Theory, Processing and Applications*, Ottawa, Canada.

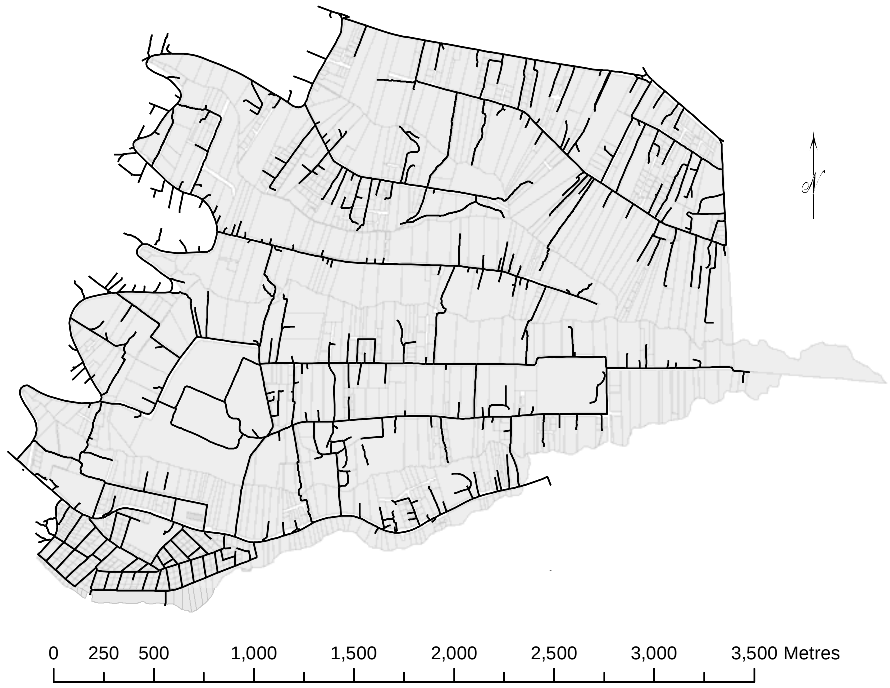
Yuan, S. and Tao, C. (1999). Development of Conflation Components. *Proceedings of Geoinformatics'99 Conference*, Ann Arbor, Michigan.

Zhang, M. and Meng, L. (2007). An Iterative Road-matching Approach for the Integration of Postal Data. *Computers Environment and Urban Systems*, 31(2007):597–615.

Appendix A

The top and bottom figures show the spatial extent covered by the experimental datasets for Case 1 and Case 2 respectively. In each case, the road network from the topographical dataset is depicted by black lines, while the land parcels are shaded in grey.

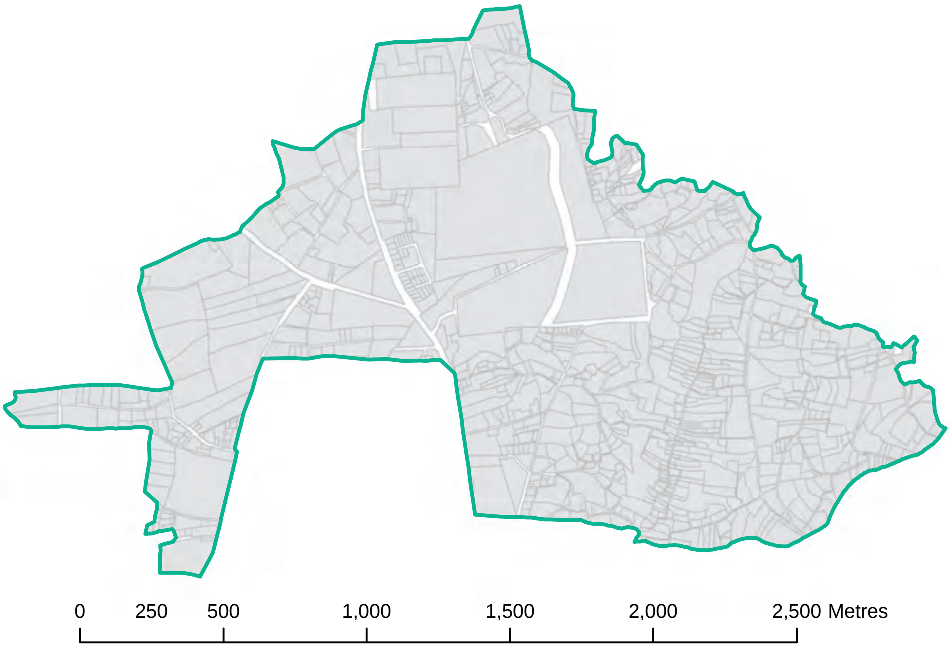
CASE 1

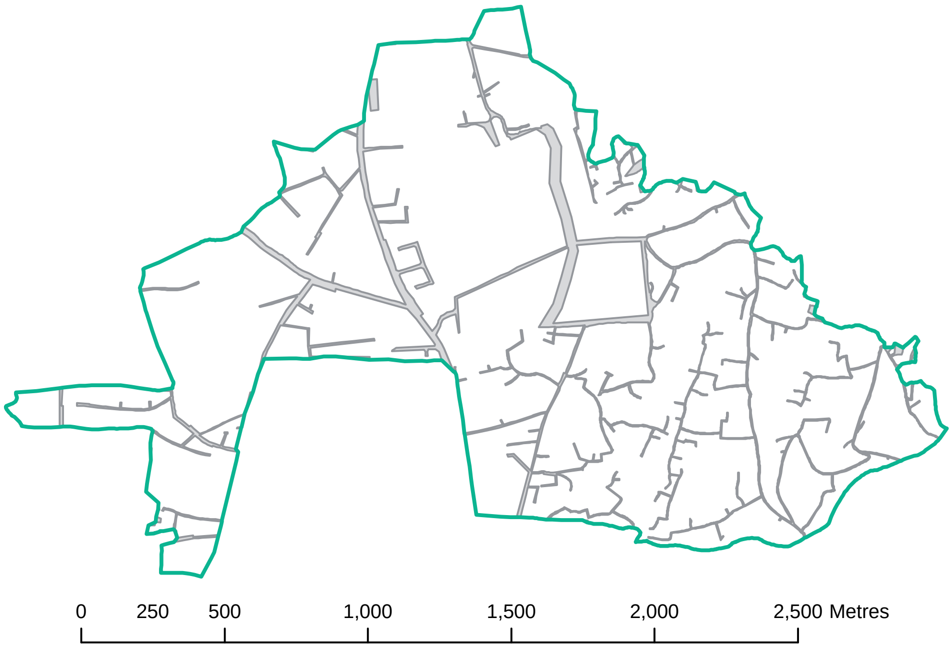


CASE 2

Appendix B

The top figure shows the land parcels shaded in grey with their corresponding concave hull (green outline). The bottom figure shows the concave hull (green outline) together with the resulting road polygons shaded in grey.





Acknowledgements

This dissertation was carried out and written using computing resources at the Institute of Cartography and Geoinformatics, Leibniz University of Hannover and with the financial support from the Germany Academic Exchange Service (DAAD). I appreciate this assistance most sincerely.

My appreciation also goes to Prof. Dr.-Ing. Monika Sester, first for accepting to supervise my research and for providing me with a convenient working environment at the institute. Secondly, for all the guidance, encouraging support, inspiring discussions that gave me the freedom to carry out the research.

I acknowledge Prof. Dr.-Ing. Winrich Voß for acting as my second supervisor and for his motivating and critical remarks on various aspects of land management during our discussions.

I also thank Prof. Yerach Doytsher for acting as a reviewer of this dissertation within a short period of time.

I am also very grateful to Prof. Dr.-Ing. Gottfried Konecny for facilitating my first contacts with my supervisors and for the support in terms of additional reading materials and references.

I thank all my colleagues at the institute for their support in one way or another. I specially thank the data integration group for the fruitful discussions we had during our meetings.

I acknowledge the Government departments in Kenya from where the data used for experiments in the course of the study was obtained.

I thank the University of Nairobi, for granting me a study leave to be away for my studies. In the same vein, I thank my colleagues at the Department of Geospatial and Space Technology for taking up extra workload because of my absence and for the encouraging e-mails that we exchanged.

And finally, let me thank my wife, daughter, parents, brothers, friends and relatives for the encouraging support in one way or another during the course of my study.

Curriculum Vitae

Personal Details

Name	David Nyangau Siriba
Date and Place of Birth	27.05.1975, Kisii, Kenya
Citizenship	Kenyan
Marital Status	Married

Education

01/1982 - 11/1989	Primary School Education at Sensi Primary School
02/1990 - 11/1993	Secondary School Education at Sameta Secondary School
11/1995 - 11/2000	Bachelor of Science in Surveying, University of Nairobi
10/2001 - 05/2004	Master of Science in Surveying, University of Nairobi
06/2008 - 09/2008	German language course at the interDaF e.V. at the Herder Institute, University of Leipzig
10/2008 - 12/2011	Doctoral Studies at the Institute of Cartography and Geoinformatics, Leibniz University of Hannover

Work Experience

09/2000 - 12/2000	Part-time Tutor, Kenya Institute of Highways and Building Technology
01/2001 - 09/2001	Land Surveyor and Mapping Expert, Highland Surveyors ltd., Nairobi
02/2005 - 07/2006	Part-time Lecturer, Department of Surveying, University of Nairobi
08/2006 -	Tutorial Fellow, Department of Geospatial and Space Technology, University of Nairobi