Applying geographic information systems to delineate residential suburbs and summarise data based on individual parcel attributes

Authors:
Stefan A. Sinske[1]
Heinz E. Jacobs[1]

Affiliations:
[1]Department of Civil Engineering, University of Stellenbosch, South Africa

Correspondence to:
Heinz Jacobs

Email:
hejacobs@sun.ac.za

Postal address:
Private Bag X1, Matieland 7602, South Africa

Background: Information aggregation to suburb level is of interest to engineers and urban planners. Readily available suburb boundaries do not always correspond to the suburb names recorded for individual properties in different data bases and unwanted errors are inherent. This mismatch of suburb names at different spatial scales poses a particular problem to analysts. As part of a parallel research project into the development of a robust guideline for suburb-based water demand analyses it was necessary to evaluate a large number of suburbs in terms of various attributes, one of which was the total suburb area.

Objectives: Suburb boundaries were needed to assess the total suburb area. The objective of this research was to develop a novel geographic information system (GIS) application to delineate suburbs with boundaries corresponding to information contained in another data base comprising individual property records. The suburb boundaries derived in this manner may not relate to municipal boundaries, or sociopolitical boundaries, nor do they have to. The fundamentally correct suburb boundary would be the one encompassing what is perceived to be the suburb based on the suburb name in a particular data base that also contains other interesting attributes, such as water use, of individual properties.

Method: The ArcGIS environment was used to delineate suburbs by means of triangulated irregular network (TIN) modelling. Boundaries for suburbs with predominantly residential land use were created that included all residential properties according to the suburb name field as recorded in the treasury system. Other vacant areas were also included so as to obtain the total suburb area. The methodology was developed to assist research in the field of potable water services, but the method presented could be applied to other services that require management of information at suburb level.

Results: This article illustrates how a tedious task of suburb delineation could be automated in the GIS environment. The tool prevents subjective results that would be prone to error. The automated procedure described could effectively delineate a large number of predominantly residential suburbs in a relatively short time span and produce repeatable results. A reasonable outline could only be obtained if a sufficient number of parcels in the area contained the same suburb name. Functionality was added to the tool so that a limit could be set for this purpose. The default was that if more than 20% of the records were erroneous it was considered impractical to delineate a suburb. The derived suburb boundaries correspond to useful information in other data bases and would thus enable more effective management of the information.

Conclusion: A novel procedure to delineate suburb boundaries in the GIS environment was illustrated in this article. Information at two different spatial scales, namely, individual consumers and suburbs, could be married for the purpose of further research into suburban attributes. The tool was applied as part of a parallel research project to delineate 468 suburbs in this manner, results of which were submitted for publication elsewhere.

Introduction

Background

Engineers are regularly faced with the challenge of effective information management in an effort to ensure municipal service delivery to communities, preceded by appropriate planning studies. Water services are generally seen as one of the most crucial municipal services in terms of human survival and public health. This research focuses on potable water supply and, in particular, on methods to deal with water consumption information at the planning stage, where crude estimates of water demand are required. Jacobs and Fair (2012) addressed information management as it pertains to water consumption data and concluded by identifying geographical information systems as a key to the next level of increasing information processing capacity.

The application of geographic information system (GIS) tools in engineering and research into their effective application is not new. Research has been presented of GIS application in various engineering disciplines, such as public transport management (Dondo & Rivett 2004), sewer system analysis (Sinske & Zietsman 2002), river flood plain modelling (Yang, Townsend & Daneshfar 2006) and water master planning (Vorster *et al.* 1995).

Urban development and water demand estimates

Greenfield land is defined as undeveloped land used for agriculture, landscape design or to evolve naturally. These areas of land are usually agricultural or amenity properties being considered for urban development (Wikipedia 2012a). The engineer responsible for planning water services would need to estimate the water requirement of the potential future land users. The eventual land use could for example be residential, commercial or industrial.

Various methods are available for estimating residential water demand in South Africa, with a comprehensive review provided by Jacobs (2008). The most recent publications in this regard were by Van Zyl, Ilemobade & Van Zyl (2008) and Jacobs, Geustyn, Loubser and Van Der Merwe (2004). Further discussion of these water demand estimation methods is beyond the scope of this text; suffice it to say that all the available local methods for estimating water demand are based on the size of individual residential plots. Households living on larger plots use more water per day than those on smaller plots.

Motivation

Greenfield land development studies require planners to make estimates of water use and other service-related variables on a relatively large spatial scale, for example, at suburb level. Details of the expected development at a small spatial scale (individual plots) is often limited at this early stage of planning because of the inherent uncertainties involved in urban development. It would thus make sense to apply a robust method to estimate the water demand for the planned suburb based on the total suburb area only. The delineated area would ultimately include all the roads, parks, public open spaces and private properties, despite much of this not requiring water supply per se.

Research problem: Suburbs, suburb names and suburb boundaries

A suburb is generally defined as a residential area existing as part of a city or within commuting distance of a city. The word is derived from the Latin terms *sub* [under] and *urbs* [city]. Most suburbs have a name and a physical boundary delineating the outer perimeter. There may be exceptions where the line between different suburbs has become blurred over time with no clear distinction between them. In such cases it would be impossible to delineate suburbs by drawing a boundary around it. In an attempt to match individual addresses, Coetzee and Rademeyer (2009) reported that address matching may be complicated by an incomplete or

inaccurate input address that includes an incorrect suburb name. These mismatch problems were noted to be the result of ambiguities originating from uncertainties regarding suburb and/or place name boundaries in that study. Coetzee and Bishop (2009) investigated national address databases and compared two different approaches for harvesting data.

It may not seem clear at first why it would be important to create suburb boundaries as presented in this article. Formerly created suburb boundaries in the required format would certainly be available. The problem is that this may not necessarily be true for each suburb and, even if it were, the information does not necessarily link up between different data sets as desired for a particular research project. For example, predefined suburb boundaries were found to be dissociated in some instances from suburb names for individual plots in the treasury data base.

As part of a parallel research project into the development of a robust guideline for suburb-based water demand analyses (Jacobs, Sinske & Scheepers in press) it was necessary to evaluate a large number of suburbs in terms of various attributes, one of which was the total suburb area. An automated method was needed to delineate suburbs in order to obtain the total suburb area. The derived suburb boundaries needed to correspond to the available water use information for individual consumers stored in the treasury data base. The suburb boundaries derived in this manner may not relate to municipal boundaries or sociopolitical boundaries, nor do they have to. The fundamentally 'correct suburb boundary' would be the one encompassing all properties with the suburb name in a particular data base. Such a boundary may not exist nor may available boundaries be associated with the database to be analysed; it thus needs to be created.

Overview

This article describes a novel procedure that was developed for this purpose. The initial steps of this research involved a review of a geographic information system (GIS) in other fields of engineering in order to assess the potential application in this study. In developing the semi-automatic method to delineate suburb boundaries using GIS it was necessary to extend the available commercial GIS product functions. The conceptual development and subsequent procedures to delineate suburbs for the purpose of obtaining the total suburb area are described in this article with a particular focus on information management. The technical findings regarding water consumption based on suburb areas derived in this manner were reported elsewhere (Jacobs *et al.* in press).

Information management with a geographic information system
Geospatial data handling ability

Wikipedia (2012b) describes GIS as any information system that integrates, stores, edits, analyses, shares and displays geographic information for informing decision-making. GIS is a computer system that handles the location and attributes

of geographically referenced data (Obermeyer & Pinto 1994; Chang 2010). The ability of GIS to process geospatial data distinguishes GIS from other information systems and makes it a valuable tool for engineers in the field of urban services such as water, sewer, gas, electricity and telephone networks. GIS is also valuable in terms of transport planning, as well as in the fields of urban and regional planning (Burrough & McDonnell 1998; Maguire 1992; Obermeyer & Pinto 1994; Shekhar & Chawla 2003; Chang 2010). Some of these GIS applications are discussed below.

Application of a geographic information system in municipal service delivery

The spatial database, graphical display capabilities and internal programming language of a GIS were identified as excellent building blocks for a spatial information system in the public transport services planning field (Dondo & Rivet 2004). The ability of GIS to combine various layers of information can, for example, be deployed in sewer-system analysis. Census enumerator areas (e.g. to derive residential sewage production) and land use area information (for business and industrial sewage production) can be selected graphically from respective layers and be allocated to manholes where the wastewater would enter the system. An analysis run can then directly be performed within the GIS via an embedded programme. Results can be displayed as thematic maps (Sinske & Zietsman 2002). Sinske and Zietsman (2004) reported a GIS-based spatial decision support system for pipe-break susceptibility analysis of municipal water distribution systems. Beuken *et al*. (2010) researched the potential of using GIS for the analysis and management of water distribution networks and pointed out several successful GIS implementations at Dutch water companies.

It is apparent that GIS has found wide application in the field of engineering services, including water and planning. None of the applications addressed the need to match information at different spatial scales and delineate suburbs or any other similar area described by polygons as described in this text.

Complex modelling in a geographic information system

In most of the above applications the standard GIS functionality of an available commercial product was extended with internal programming languages to perform complex modelling. The following software programs were deployed in this study:

- The widely used ArcGIS Desktop 10.0 software package (licence type ArcInfo) from ESRI was used as GIS platform. ArcMap, which is the central application of ArcGIS, was used as the main spatial viewer. File management tasks were performed with ArcCatalog and spatial analyses with ArcToolbox, both part of the ArcGIS Desktop and accessed via ArcMap (ESRI 2010).
- ArcScene is a 3D visualisation application and part of ArcGIS, and was used in conjunction with the 3D Analyst tools of ArcToolbox to perform complex surface modelling for the suburb delineation process. The 3D

Analyst extension of ArcGIS is a system requirement for the above.
- The end results were finalised via ModelBuilder, which is part of ArcGIS.

Suburb delineation using a geographic information system
Definition of a geographic information system parcels and features

In real estate terms, a lot or plot is a tract or parcel of land owned or meant to be owned by an owner or owners. Some countries use the terminology 'parcel of real property' whilst others use 'immovable property', meaning practically the same thing. Each property is described by a polygon in GIS, commonly referred to as a parcel. In between the parcels are other areas of land and interesting geographic features with spatial attributes that may be recorded in the GIS data base as well (some may also be irrelevant parcels). In addition to the parcel polygons the data base may contain point features (e.g. a beacon or centre point of a parcel) and also line features (e.g. a small water canal or hiking route). The most basic type of polygon would be a triangle. Each parcel has a unique GIS-code with associated information for it stored in the corresponding data base.

Chang (2010) defines a feature as any representation of a real-world object on a GIS-map; it could be any shape. In ArcGIS, a feature class stores spatial features of the same geometric type (i.e. point, line, polygon, etc.), same attributes (i.e. common set of attributes) and the same spatial reference (i.e. common mapping co-ordinate system). In the above context, for example, all the parcels addressed by the delineation procedure would be stored in a feature class. Feature classes again are stored in an ArcGIS geodatabase as either standalone or grouped in a feature dataset. These terms are applicable to this study as defined below.

Description of the research problem in terms of a geographic information system polygons

The research problem is firstly explained in terms of GIS terminology before moving on to presentation of the automated procedure for suburb delineation. This description is presented by considering a hypothetical example and uses water consumption as a desired attribute aggregated to suburb level. The suburb name and the number of houses used in this section are completely irrelevant and did not form part of this or further research work with the suburb delineation tool presented in this article. The name Suburb A and the 500 plots (approximately) were simply chosen to clearly illustrate the research problem and the devised method to delineate suburbs. The actual method could be applied to any real suburb or any number of real suburbs in a given area.

The treasury data base would contain information for each of these 500 consumers or residences. Each would have a water meter read monthly, with data stored in the treasury data base. Each consumer's property would be described by

numerous fields in the data base. One of these fields would be the suburb name field with the entry: Suburb A. The town planner would be able to provide an independent GIS data base describing the cadastral layout of the town - this could be seen as a map of the town showing all the properties. This GIS data base typically contains the suburb name and land use for each property in separate data fields. The suburb name in the GIS data base would not always match up with the suburb name in the treasury data base.

In between these 500 parcels comprising residential plots would also be vacant areas that would typically represent roads, parks or public open spaces, but these would not typically be flagged with the suburb name. These vacant areas are often not specifically captured as parcel polygons. It would be obvious to the reader that all the parcels, roads and other vacant areas in between the parcels should actually be part of Suburb A if the total suburb area needs to be considered.

Readily available up-to-date polygons (in GIS format) depicting suburbs in the desired fashion are unfortunately seldom available. Boundary and name changes over time, particularly after local political change in the mid–1990s, resulted in lacunae. This was true for boundaries at the provincial level to suburb and ward level. Another problem was that of duplicate suburb names, for example, a study that would encompass the entire country and where the same suburb name would be found in different cities. The suburb boundary matching desired attributes and encompassing all spaces in between plots could easily and quickly be generated by the method reported in this study, producing repeatable results.

This suburb delineation could be done by hand, in other words by clicking with a mouse around the plots to create a single polygon for the suburb. Such a task would become tedious, subjective and prone to error when repeated for hundreds of suburbs. This article presents an automated procedure that could delineate a suburb and would produce repeatable results. A reasonable outline could, of course, only be obtained if a sufficient number of parcels in the area contained the same suburb name (and same spelling) in the data base. Functionality was added to the tool so that a limit could be set for this purpose. The default was that if more than 20% of the records were erroneous it was considered impractical to delineate a suburb.

Triangulated irregular network modelling

The novel GIS method to delineate suburbs boundaries is based on triangulated irregular network (TIN) terrain modelling. A TIN is a set of adjacent (i.e. connected), continuous and non-overlapping triangles constructed by triangulating irregularly spaced nodes or observation points. These points are vertices with x, y and z co-ordinates. The principles of TIN are described in more detail by Burrough (1986) and Chang (2010). The TIN model, with its network of triangles in the form a sheet, or so-called mesh, is ideal for terrain representation and modelling (Burrough & McDonnell 1998).

Different methods of interpolation are available to form these triangles. The most widely used is called Delaunay triangulation and is implemented in the ArcGIS software suite (ESRI 2010). This triangulation method ensures that all sample points are connected with their two nearest neighbours to form triangles as equiangular or compact as possible. In this manner, it is possible to avoid the formation of too many unwanted sharp, long and skinny triangles (ESRI 2010; Chang 2010; Li & Ai 2010). A finished TIN comprises three types of geometric objects, namely, (1) triangles (facets), (2) points (nodes) and (3) lines (edges). Elevation data is stored at the nodes, whereas slope and aspect data are stored for each facet and remain constant over the facet (Chang 2010). Most GIS software packages implement TIN because one of their data structures and have the ability to export the abovementioned individual components of the TIN as separate polygon, point and line features for further analysis. Apart from TIN, most GIS also implement the grid structure for terrain modelling. One of the biggest advantages of the TIN model over the grid model is the flexibility of TIN to model more detail at certain locations (i.e. terrain specific source data such as roads, rivers, lakes and parcels can be incorporated in the triangulation process). Only high-resolution grids can show these detailed features, but this would not be an optimal solution with regard to data storage because the cell size (which is constant for the grid) will have to be defined as very small over the whole study area (Burrough 1986; Burrough & McDonnell 1998; Chang 2010).

A TIN data model can also be used to represent and model two dimensional (2D) surfaces, as is the case with this research. Li and Ai (2010) discussed the application of Delaunay TIN to detect various spatial and structural characteristics hidden in 2D geometry data (i.e. a type of spatial data mining application). They also pointed out an important aspect of Delaunay TIN, namely, that the triangle element can play two roles: either the component of the polygon feature or the bridge between neighbouring objects. Triangles playing the bridging role are distributed on the principle of 'nearest connection' of Delaunay TIN. Hereby, the neighbourhood relationship is presented by only one triangle no matter how far between these objects, which is a useful characteristic for spatial neighbourhood analysis.

The suburb delineation process presented in this article is also a 2D TIN application based on the abovementioned dual role of the Delaunay TIN triangle. This means that some triangles will be used to cover the entire area of parcels in the suburbs (i.e. TIN triangles located on parcels) and others will span the empty space between parcels (i.e. a TIN triangle located in the empty space bridging the gap between two other TIN triangles located on parcels). The method can delineate a large amount of suburb boundaries all at once and can be executed in ArcGIS via ten geoprocessing steps (Figure 1).

The suburb delineation process requires spatially referenced parcel data as an input, with attribute data fields containing the suburb name and land use. The land use description data is not relevant at this stage. An ArcGIS file geodatabase (.gdb) can now be created in ArcCatalog (Step 1 of Figure 1)
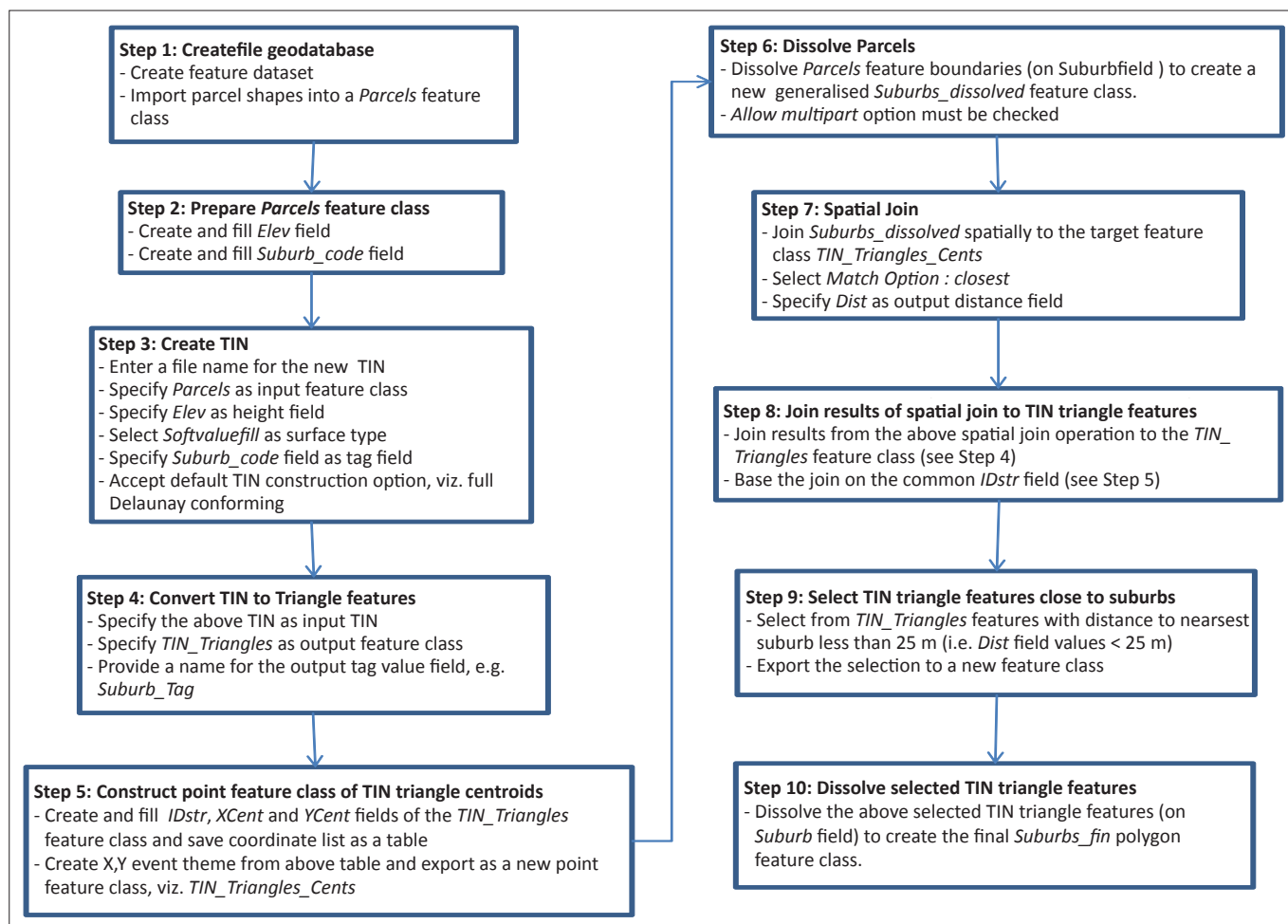
Step 1: Createfile geodatabase
- Create feature dataset
- Import parcel shapes into a *Parcels* feature class

Step 2: Prepare *Parcels* feature class
- Create and fill *Elev* field
- Create and fill *Suburb_code* field

Step 3: Create TIN
- Enter a file name for the new TIN
- Specify *Parcels* as input feature class
- Specify *Elev* as height field
- Select *Softvaluefill* as surface type
- Specify *Suburb_code* field as tag field
- Accept default TIN construction option, viz. full Delaunay conforming

Step 4: Convert TIN to Triangle features
- Specify the above TIN as input TIN
- Specify *TIN_Triangles* as output feature class
- Provide a name for the output tag value field, e.g. *Suburb_Tag*

Step 5: Construct point feature class of TIN triangle centroids
- Create and fill *IDstr*, *XCent* and *YCent* fields of the *TIN_Triangles* feature class and save coordinate list as a table
- Create X,Y event theme from above table and export as a new point feature class, viz. *TIN_Triangles_Cents*

Step 6: Dissolve Parcels
- Dissolve *Parcels* feature boundaries (on Suburbfield) to create a new generalised *Suburbs_dissolved* feature class.
- *Allow multipart* option must be checked

Step 7: Spatial Join
- Join *Suburbs_dissolved* spatially to the target feature class *TIN_Triangles_Cents*
- Select *Match Option : closest*
- Specify *Dist* as output distance field

Step 8: Join results of spatial join to TIN triangle features
- Join results from the above spatial join operation to the *TIN_Triangles* feature class (see Step 4)
- Base the join on the common *IDstr* field (see Step 5)

Step 9: Select TIN triangle features close to suburbs
- Select from *TIN_Triangles* features with distance to nearrest suburb less than 25 m (i.e. *Dist* field values < 25 m)
- Export the selection to a new feature class

Step 10: Dissolve selected TIN triangle features
- Dissolve the above selected TIN triangle features (on *Suburb* field) to create the final *Suburbs_fin* polygon feature class.

FIGURE 1: Suburb delineation procedure.

and the shapefile, containing the parcels, could be imported as feature class within a new feature dataset. The feature dataset provides a logical structure (almost like a file folder) wherein feature classes can be grouped together. The ArcGIS file geodatabase with unlimited storage space was chosen instead of the ArcGIS personal geodatabase, which has a storage limit of 2GB.

The elevation field (in the *Parcels* feature class) can be filled with any constant elevation value (such as e.g. 1 m) for all parcels (Step 2 of Figure 1) because the TIN will be used for 2D analysis only. The suburb code in the *Parcels* feature class (Step 2 of Figure 1) must be a unique integer code for each suburb name. The TIN model can best be created in the ArcScene environment with the *Create TIN* tool (Step 3 of Figure 1), accessible from the integrated ArcToolbox (note, the 3D Analyst extension of ArcGIS is required). The 2D TIN model will be built based on the parcel vertices, which all have the abovementioned 1 m spot height allocated. Important is to set the surface type to *Softvaluefill*. This will ensure that the parcel boundaries will be enforced in the triangulation as breaklines (i.e. TIN triangle edges will not cross the parcel boundaries [Figure 2a]). Furthermore, the TIN triangles inside these parcel polygons will hereby be attributed (i.e. filled) with the corresponding suburb code tag value (for cross-reference checking). These TIN triangles

can now be extracted from the TIN model and saved as a new polygon feature class (Step 4 of Figure 1) via the ArcToolbox conversion tool *TIN Triangle*.

Geoprocessing

A series of geoprocessing steps (Step 5 to Step 8 of Figure 1) are now required to determine the name of the closest suburb and corresponding distance for each TIN triangle. The latter is measured from the geometric centroid of the triangle (Step 5 of Figure 1) to the closest parcel edge in the suburb. This can be accomplished in ArcMap via the *Calculate Geometry* function in combination with the *Spatial Join* and *Dissolve* tools (accessible from the integrated ArcToolbox). The dissolve operation based on the suburb name (Step 6 of Figure 1) merges all individual parcels in a suburb into one multipart suburb polygon, in order to improve the spatial join operation time (Step 7 of Figure 1). This temporary suburb polygon is only used in the spatial join operation. The spatial join results include the name of the closest suburb and the corresponding distance and can now be joined to the TIN triangle features (Step 8 of Figure 1) for further queries and analyses.

The distance to the nearest suburb can be queried (Step 9 of Figure 1) to obtain a selection of TIN triangles close to suburbs. A TIN triangle can be regarded as close to a

suburb when it is either completely within a suburb parcel (distance will then be zero) or within approximately 25 m from a suburb parcel. The latter scenario is when the TIN triangle is located in a street or in a nearby unidentified (vacant) land use area either somewhere in the interior of the suburb or in the outer border regions between suburbs. The abovementioned distance selection process will assign in these border regions approximately half of the TIN triangles to the one suburb and half to the other (i.e. they slot together almost like a jigsaw puzzle [Figure 2b]).
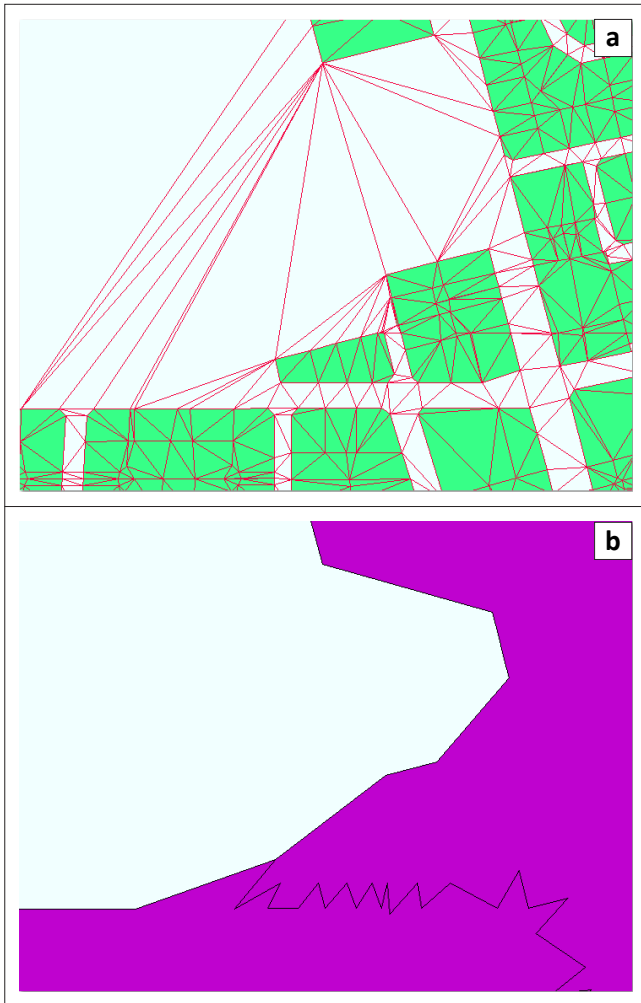
FIGURE 2: Transforming (a) individual parcels to (b) suburb areas using a triangulated irregular network (TIN).

Final selection for suburb polygons

The invalid long and skinny TIN triangles located in the undefined land use areas on the edges of a suburb (as the result of the TIN interpolation process) will also mostly all be filtered out by the distance selection process. Some smaller ones may be missed by the process and can afterwards be wiped manually for 'aesthetic' reasons. For the suburb area calculation, however, they are insignificant and could remain in the system.

The final selection of TIN triangles can now be dissolved (Step 10 of Figure 1) based on the suburb names from the above spatial join results in order to obtain the final delineated suburb polygons. It can be recalled that this boundary does not need to match up to any other boundary - it needs to delineate the outer edge of a number of plots that were flagged in a given data base as being part of this suburb, plus all spaces in between.

Data summary output

Prior to the analysis, the land uses in the suburbs need to be identified according to the types, namely, residential, open space, business, industrial, et cetera. The land use information must be summarised per suburb in order for the model to extract the predominantly residential suburbs. The summarisation process is illustrated in Figure 3.

Composite land use information needs to be summarised per suburb (Step 1 inFigure 3) by generating a summary table on the *Parcels* feature class. This can be accomplished in ArcMap with the integrated ArcToolbox function *Summary Statistics* and specify the *Suburb* and *Land_use* fields as the two *Case* fields on which the summary should be based. The resultant summary would contain the various information required as output in multiple suburb records and could be linked with the *Suburbs* feature class (containing the final delineated suburb boundaries) via a one-to-many relationship. This type of relationship would, however, make further processing by the model unnecessarily complex.

A simpler link could be accomplished in ArcMap by selecting and exporting from the composite table the abovementioned land uses into separate tables and consecutively joining them (i.e. one-to-one relationship) with the *Suburbs* feature
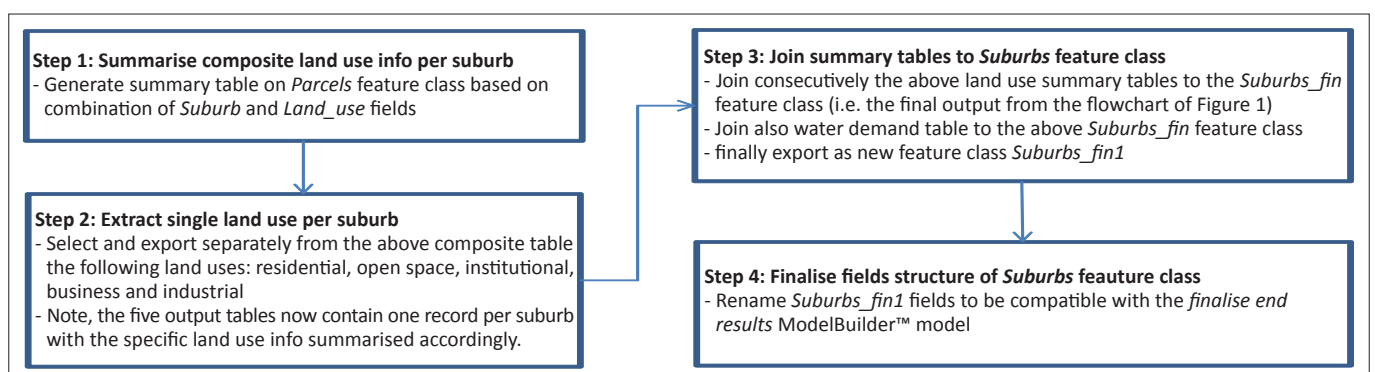

FIGURE 3: Land use summarisation.

class (Step 2 and Step 3 of Figure 3). The *Suburbs* feature class now contains one record per suburb with the land use information contained in separate fields, which is ideal for further processing. The procedure checks that the number of parcels per suburb deviate less than 20% from the number of parcels as recorded for the corresponding suburb in the treasury database. Suburbs with more than 20% deviation in number of parcels are excluded from the selection because in these cases there are obvious fundamental differences in the suburb boundaries between the two data sets and the delineated suburb would not be considered valid for the purpose of deriving its total area.

Conclusion and future research needs

This article illustrates how a tedious task of suburb delineation could be automated in the GIS environment. The article shows how information at two different spatial scales, namely, (1) individual consumers and (2) suburbs, could be married for the purpose of further research into suburban attributes. The suburb boundaries obtained from the system also encompass the vacant areas and roads (in between the parcels). The automated procedure employed built-in logic to enable the selection of predominantly residential suburbs and to derive the total suburb area. The tool was employed as part of a parallel research project into suburban water demand to delineate 468 suburbs in this manner, results of which were submitted for publication elsewhere (Jacobs *et al*. in press).

The GIS based information system presented in this article could further be improved by implementing the following possible enhancements:

- The semi-automatic suburb delineation process and the land use summarisation process could be implemented directly as models in the ModelBuilder environment, subsequently reducing the analysis time.
- The invalid (long and skinny) TIN triangles located on the edges of the suburb after the TIN interpolation process add inaccuracies to the suburb delineation procedure in some cases. An improved selection of TIN triangles could be obtained with almost no invalid triangles on the suburb edges by reducing the threshold settings. Only a few valid triangles located on wide roads and traffic circles would be wrongly missed by this finer threshold. The distance threshold cannot be reduced significantly for study areas containing suburbs with many unidentified (vacant) land areas (i.e. those not captured as parcels) because these vacant areas would then wrongly be excluded from the suburb area calculation.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationship(s) which may have inappropriately influenced them in writing this paper.

Authors' contributions

H.E.J. (University of Stellenbosch) instigated this research, selected the journal and handled editorial matters. The GIS applications were developed by S.A.S. (University of Stellenbosch). Both authors contributed equally to the text in this manuscript.

References

Beuken, R.H.S., Van Daal, K.H.A., Pieterse-Quirijns, E.J. & Zoutendijk, F.J.M., 2010, 'The use of GIS for analysis of water distribution networks', in J. Boxall & C. Maksimovic (eds.), *Proceedings of the 10th International Conference on Computing and Control for the Water Industry, CCWI 2009 – 'Integrating water systems'*, Sheffield, UK, September 01–03, 2009, pp. 93–98.

Burrough, P.A., 1986, *Principles of geographical information systems for land resources assessment*, Oxford University Press, Oxford.

Burrough, P.A. & McDonnell, R.A., 1998, *Principles of geographical information systems*, Oxford University Press, Oxford.

Chang, K.T., 2010, *Introduction to geographic information systems*, 5th edn., McGraw-Hill, New York.

Coetzee, S. & Bishop, J., 2009, 'Address databases for national SDI: Comparing the novel data grid approach to data harvesting and federated databases', *International Journal of Geographic Information Science* 23(9), 1179–1209. http://dx.doi.org/10.1080/13658810802084806

Coetzee, S. & Rademeyer, I.M., 2009, 'Testing the spatial adjacency match of the Intiendo address matching tool for geocoding of addresses with misleading suburb or place names', *In-proceedings International Cartography Conference*, Santiago, Chile, November 15–21, 2009, n.p.

Dondo, C. & Rivett, U., 2004, 'Spatial information systems in managing public transport information', *South African Journal of Information Management* 6(2), 8 pages.

ESRI, 2010, *ArcGIS Desktop Help*, Environmental Systems Research Institute, Redlands, CA.

Jacobs, H.E., Sinske, S.A & Scheepers, H.M., in press, article submitted for peer review and possible publication to *Urban Water Journal*, submitted September 2012.

Jacobs, H.E., 2008, 'Chronologiese oorsig van Suid-Afrikaanse riglyne vir residensiële gemiddelde jaarlikse waterverbruik met erfgrootte as onafhanklike veranderlike: Navorsings- en oorsigartikel', *Suid-Afrikaanse Tydskrif vir Natuurwetenskap en Tegnologie* 27(4), 240–265.

Jacobs, H.E., 2008, 'Residential water information management', *South African Journal of Information Management* 10(3), 12 pages.

Jacobs, H.E. & Fair, K.A., 2012, 'A tool to increase information-processing capacity for consumer water meter data', *South African Journal of Information Management* 14(1), 7 pages.

Jacobs, H.E., Geustyn, L.C., Loubser, B.F. & Van Der Merwe, B., 2004, 'Estimating residential water demand in southern Africa', *Journal of South African Institution of Civil Engineering* 46(4), 2–13.

Li, J. & Ai, T., 2010, 'A triangulated spatial model for detection of spatial characteristics of GIS data', in Y. Wang (ed.), *Proceedings of the 2010 IEEE International Conference on Progress in Informatics and Computing, PIC 2010*, Shanghai, China, 10–12 December 2010, IEEE, Washington, DC, pp. 155–159.

Maguire, D.J., 1992, An overview and definition of GIS, in D.J. Maguire, M.F. Goodchild, & D.W. Rhind (eds.), *Geographical information systems: Principles and applications*, Vol.1, pp. 9–20, Longman, New York.

Obermeyer, N.J. & Pinto, J.K., 1994, *Managing geographic information systems*, The Guilford Press, New York.

Shekhar, S. & Chawla, S., 2003, *Spatial databases: A tour*, Pearson Education, Inc., Upper Saddle River, NJ.

Sinske, S.A. & Zietsman, H.L., 2002, 'Sewer-system analysis with the aid of a geographical information system', *Water SA* 28(3), 243–248. http://dx.doi.org/10.4314/wsa.v28i3.4891

Sinske, S.A. & Zietsman, H.L., 2004, 'A spatial decision support system for pipe-break susceptibility analysis of municipal water distribution systems', *Water SA* 30(1), 71–79. http://dx.doi.org/10.4314/wsa.v30i1.5029

Van Zyl, H.J., Ilemobade, A.A., & Van Zyl, J.E., 2008, 'An improved area-based guideline for domestic water demand estimation in South Africa', *Water SA* 34(3), 381–392.

Vorster, J., Geustyn, L.C., Loubser, B.F., Tanner, A. & Wall, K., 1995, 'A strategy and master plan for water supply, storage and distribution in the East Rand region', *Journal of South African Institution of Civil Engineering* 37(2), 1–5.

Wikipedia, 2012a, 'Greenfield land', viewed 22 August 2012, from http://en.wikipedia.org/wiki/Greenfield_land

Wikipedia, 2012b, 'GIS', viewed 22 August 2012, from http://en.wikipedia.org/wiki/GIS

Yang, J., Townsend R.D. & Daneshfar, B., 2006, 'Applying the HEC-RAS model and GIS techniques in river network floodplain delineation', *Canadian Journal of Civil Engineering* 33(1), 19–28. http://dx.doi.org/10.1139/l05-102