

Functional classification of the Gauteng provincial road network using the South African Road Classification and Access Management Manual (TRH26)

Pryaska Veramoothea¹, André Breytenbach¹, Ethel Baloyi¹

¹Built Environment, Council for Scientific and Industrial Research, Gauteng, South Africa, pveramoothea@csir.co.za

DOI: <http://dx.doi.org/10.4314/sajg.v4i3.7>

Abstract

The built environment consists of various land uses and activities connected by a road network. The efficiency and effectiveness of the road network directly impacts economic growth and societal movement. This study involved the functional classification of the Gauteng provincial road network using the South African Road Classification and Access Management Manual (TRH26) as a guideline. In combination with high resolution up-to-date satellite reference imagery, dwelling and building inventories, road centrelines and other base layers, it was found that the TRH26 provided a well-structured approach to classifying the provincial road network according to the functional properties of each road link. It resulted in an accurate spatial representation of the current Gauteng provincial road network when using an appropriate geographic information system - providing a wealth of formal and technical information regarding each road and street via the comprehensive attribute database. At link level it was possible to synchronise the upgraded network with the financial asset registry in order to gauge the current state of the provincial roads within Gauteng, and to identify those that are in need of attention and/or rehabilitation to optimise efficiency and safety. This approach can be standardised throughout the country to make inferences to road safety, speed regulation, driver behaviour, traffic studies and accessibility to services.

1. Introduction

The transport network linking the various land uses contained by the urban landscape is an essential element of the built environment. With the difficult economic position that South Africa currently finds itself in, it is imperative to have a cost effective and efficient road network. At the same time, the high volume of traffic-related accidents calls for a road network that is safe for all road-users (Fitzpatrick *et al.*, 2003). It is known that the physical road network affects traffic and driver behaviour, and this in turn directly influences energy consumption and the environmental impacts associated with the emissions thereof (Wilkie *et al.*, 2012). This further affects the sustainability of the land use in practice as it affects the flora and fauna in the area (Jaarsma, 1997). The Committee of Transport Officials (COTO) defines an optimal road network as one that is congruent to its land use, as the land use determines the factors involved for what type and class of road should run through (COTO, 2012). However, even though the importance between land use and development is well known, an equally beneficial relationship between the two is often very difficult to maintain due to economic constraints and zoning schemes (DEAAT, 2002).

Maintenance of the road network is imperative as it is the primary means of mobility and access in both economic and societal sectors. Economic growth in South Africa is therefore largely dependent on an efficient and fully functional road network, as it affects the movement of goods and services rendered nationally and internationally. Societally, a well maintained road network will ensure road safety for all users, provided the rules of the road are adhered to. Economically, it is important that the classification of a road represents the road's functionality effectually, as it can enhance the delivery of goods to the land use by making it efficient and effective (COTO, 2012). Arterial roads of a higher class significantly influence economic development by providing the main route of transportation of goods and services, as well as significantly influence societal movement as it provides a safe, effective and efficient route to travel on (DEAAT, 2012). Therefore the need exists to accurately geo-locate or capture and describe each roadway link using existing and appropriate data and software.

Road extraction and classification can however be a very time-consuming exercise when not automated to some degree (the Gauteng provincial road network used in this paper eventually contained 2530 road links and 1975 nodes). Geographic information systems (GIS) and remote sensing, with the aid of digital imagery and modern global positioning system (GPS) technology and navigational tools, allow for this task to become less tedious, as it allows for the semi-automation of the extraction and classification routines. This is done by using algorithms to look at the characteristics and properties of the road and its surroundings, and thereby classifying them in accordance to the classification specifications (Ye *et al.*, 2006), which in this case is the functional classification of a provincial road network.

The objectives of this study were to acquire accurate and up-to-date spatial data, to develop a GIS workflow to prepare this data to include a variety of important attributes indicative of the current situation of the Gauteng provincial Road Network Inventory (RNI), and to functionally classify each road within the road network according to the TRH26 manual. The TRH26 guidelines ensured that all classifications were made according to Road Infrastructure Strategic Framework for South Africa (RIFSA) standards put forth by the Department of Transport to be used nationally. As

RIFSA standards require data that is not older than two to five years (depending on the asset), the Road Asset Management System (RAMS) has to be updated every two years.

2. Background

Studies on road networks began abroad in the 1960's and focused mainly on the topology of the landscape. Where currently GIS commonly plays a huge role in such network analyses, along with the vast amount of data available today, past studies had to be conducted through graph-theoretic network analysis with many data restrictions, reduced computational capacity and lack of modelling capabilities (Xie and Levinson, 2007). Wilkie *et al.* (2012) describes GIS data useful for road network analyses as normally consisting of polyline data and containing various attribute fields. In this study, line and point data were used, which is common in network analyses. The ESKOM SPOT Building Count (SBC) data used in this study enabled one to observe a change in building density over time. Building density data was acquired nationally from 2006 to 2012, and included land use attributes. This allowed for one to assess the appropriate type of road infrastructure required, according to the building density, land use, and expected growth based on the observed pattern of growth over time.

In 2012, KwaZulu Natal's eThekweni municipality also adopted the TRH classification system to optimise the functionality, accessibility and mobility of the (then) existing 15000km road network. In addition to the road network, the KZN Department of Transport also classified non-motorised transport locations, public transport services, pedestrian crossings and traffic-calming devices, and all this extra information was recorded in order to be taken into account when upgrading their infrastructure (Rodin, 2013). An important limitation to the KZN road classification was the data integrity and ensuring that the data was up to date – similar to this study. In order to ensure their data integrity, roads were verified via visual inspection and by using traffic count data (measures that were not included in this study, as per the TRH26 guidelines). Once the challenge of upgrading the data was complete, the classified dataset formed the basis of many future municipality projects.

According to the Federal Highway Administration of 1997, the process of functional classification involves the grouping of every road within a road network according to its structure and the service it provides (Fitzpatrick *et al.*, 2003). Functional road classification therefore primarily addresses aspects of access and mobility. Local streets mostly serve the purpose of land access, where arterial roads represent mobility of through-traffic. Collector roads offer some access and some mobility functions. Roads of a higher class should have little to no access roads, as this would affect the efficiency of the road. Conversely, local streets have to be limited to access roads, as the speed and amount of traffic associated with mobility roads can be dangerous to not only users of the road, but also pedestrians (DEAAT, 2012). Intersections and access roads reduce the mobility function due to lower speeds for the turning of vehicles and the presence of pedestrians (Sampson, 2010). Mobility routes should be continuous, with limited access, and should contain a longer length between nodes (Holderness and Stander, 2008). Figure 1(a) shows the proportion of service per road type through access or mobility areas. Figure 1(b) shows the road types that should be considered as mobility roads and those that should be considered as access roads.

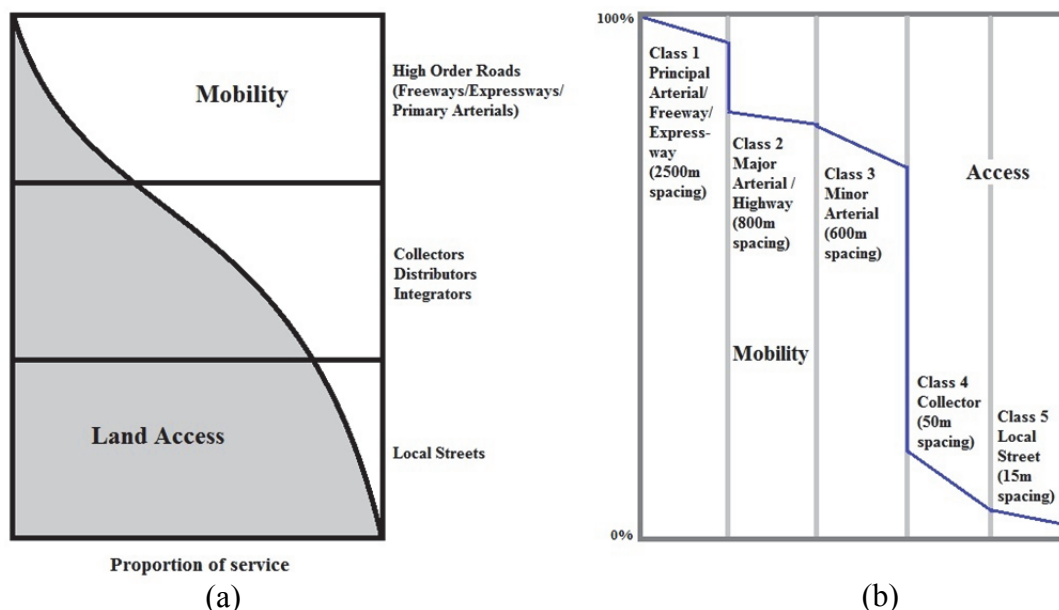


Figure 1. (a) Mobility vs access roads (DEAAT, 2012, pg 9); (b) Road types that should be considered as mobility or access roads (Sampson, 2010)

The product of this study was ultimately used to update the Financial Asset Registry, in order to qualify for funds from the Provincial Roads Maintenance Grant for road maintenance, emergency maintenance, road upgrades, repairs to

roads and bridges, and to improve the efficiency of the coal haulage network. According to the Division of Revenue Bill of 2013, grants provided from the 2015/2016 financial year will be based solely on performance of the RAMS in place.

3. Material and Methods

3.1 Study Area

Gauteng occupies less than 2% of the South African landscape, but contributes significantly to not only the South African economy, but also the African GDP. Despite its small area, it is home to over 12 million people. For this reason, access to services and mobility to areas of employment and social activity is crucial.

This study involved the enhancement and classification of the function of the Gauteng provincial RNI for 2013/2014 for the Gauteng Provincial Department of Roads and Transport (GPDRT), as a follow up to the 2010 network and old classification system. All newly classified roads – both paved and gravel – were therefore located within the current Gauteng provincial boundary, with the GPDRT confirming ownership of each provincial road. It amounted to just over 5781 kilometres of carriageways that excluded metropolitan roads and streets, as well as the growing number of SANRAL toll roads.

3.2 Data

The source data supplied by the GPDRT was in survey point format, which had to be converted to line features. Each line had a number of attributes assigned to it as inherited from the 2010 classification, nodes, descriptions and lengths. The nodes and their unique identifiers in the 2010 dataset had not yet been captured as a point dataset. Dual carriageways also appeared in the 2010 dataset, but were to be correctly captured as separate entities digitally.

The data was viewed, prepared, and classified on ESRI ArcMap 10.3 using supporting imagery that included ESRI base layers, the 2012 South African national SPOT 5 mosaic, and National Geo-spatial Information (NGI) aerial photography. Base layers for the study included the 2014 release of the AfriGIS road centreline dataset, cadastral information and recent dwelling inventories. Last-mentioned inventories –processed to determine the rural/urban boundaries and highlight residential and non-residential growth nodes – included the SBC data over a number of years and GeoTerraImage (GTI) Building Base Land Use datasets for 2001 and 2010 (Breytenbach, 2014).

4. Methodology

The road classification criteria used was in accordance with the TRH26, South African Road Classification and Access Management Manual, Version 1.0. This manual was created by the Committee of Transport Officials (COTO), to create a simple and standard approach to road access systems and mobility that -is congruent to the needs of the economy, logistics sector and infrastructure development.

An overview of the editing and classification workflow can be seen in Figure 2. The Gauteng provincial roads dataset for 2010 and the Gauteng provincial survey dataset for 2013 were used for the updated 2013/2014 classification. The new road network had to be represented by true road centrelines, therefore caution was taken in order to accurately represent the roads. This was achieved by using the up-to-date high resolution satellite and aerial imagery in conjunction with the AfriGIS Road Centrelines 2014 dataset. The 2010 dataset did take into account dual carriageways, however this was only captured in the attribute table. Thus the 2010 dataset represented dual carriageways as a single line, which had to be updated in the 2013/2014 dataset as two distinct separating and converging centrelines per dual carriageway.

A well-established road network consists of nodes and road links, which together display topological and geometric changes. The nodes denote locations in a function of space, whilst the links denote the active connections of the road network. The links also possess different properties according to the physical structure of the road, and therefore are not identical to one another (Xie and Levinson, 2007). Road intersections show the highest number of collisions in urban areas, therefore taking cognisance of these intersections is integral for road safety (COTO, 2012).

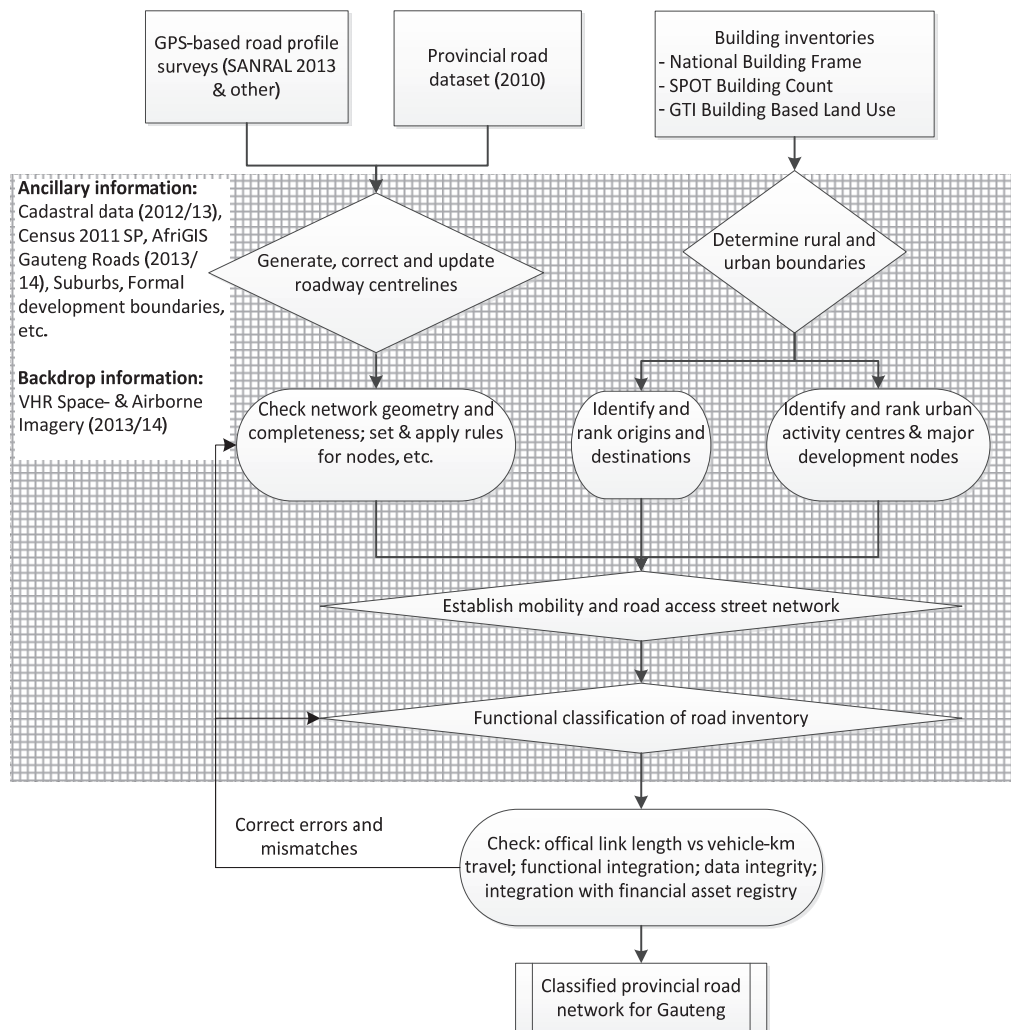


Figure 2. Technical workflow used for the functional classification of the GP provincial roads network, as per the TRH26 guidelines

Each road was considered individually and nodes were added to each provincial road intersection, where administrative boundaries were crossed and where surface types changed (e.g. paved to gravel). The line feature between the captured nodes then became the road “links” which were classified according to its location and properties. Each node and link was given its own unique identifier and the lengths of each link were calculated and recorded. The primary key consisted of the road name, the survey direction and sequential numbering from the first road link to the last, e.g. D1017_P0001, D1017_P0002, and so on. Link distances aided in applying the distance limitations set out by the TRH26 in terms of total road length (Table 1).

With the addition and upgrade of the road network, a number of new nodes needed to be added. The new added nodes were denominated by its proximity to existing nodes and these are yet to be given unique identifier codes. Certain intersections had insufficient existing nodes to name the all the new nodes in the proximity, thus a systematic alphabetical system was devised to capture the new nodes per intersection, as can be seen in Figure 5(a). If a new node needed to be added out of an intersection, the new node allocation convention that was used in this study allowed for an “N” to be placed before the name of the closest node. As the 2010 dataset only represented dual carriageways with a single centreline, all dual carriageway intersections contained only a single node. The updated double centreline representation of dual carriageways dictated that intersecting dual carriageways needed to be upgraded from a single node to up to four nodes. The node allocations and spatial location thereof aided in distinguishing the road segments as primarily access or mobility routes.

Each link had a positive (“P”) description if it followed the official survey direction of the road (all links inclusive in a single entity e.g. D103). A negative symbol (“N”) was assigned for all opposing legs in dual carriageways. In addition to this, the general directions of each dual carriageway were recorded (N, S, E, W). For the remaining omnidirectional roads, which had no physical barrier between the oncoming and ongoing lanes, an “O” was assigned. The surface type of each road was recorded during this process as well.

The 2010 functional classification was carried through to the attributes of the 2013 dataset, however a more extensive classification was done according to the TRH26 (Table 1) for the 2013/2014 revision. Xie and Levinson (2007) stated that a network with a vast set of attributes, such as those used in this study, is a network of “a higher level of complexity.” The complete set of attributes can be found in Addendum A.

The first distinction that was made was that of whether the road belongs to a rural or urban area. This was decided by looking at the inclusion and/or proximity of the road to the development boundary, as well as considering the underlying residential and non-residential kernel densities derived from the building inventories. In Figure 3 it is apparent that in highly developed areas, and within the development boundary, the roads were classified as “urban”, whereas the remaining roads were classified as “rural”.

Table 1. Summarised criteria for the classification of roads (TRH26)

Class	1	2	3
Rural	>50km	>25km	10-100km
	Continuous	Continuous	Start and end node
	Multiple Roads	Multiple Roads	Non-continuous
Urban	Connect to R1	Connect to R2	Connect to R3
	Start and stop at other U1	Start and stop at U/U2	Start and stop at U2/U3
	High density 5km	High density 1.5km	High density 1.2-1.5km
	Medium density 8km	Medium density 3km	Medium density 1.2-2km
	Low density 10km	Low density 5km	Low density as required
	>20km	>10km	>2km

Class	4/4a	4b	5/5a	5b
Rural	<10km		<5km	
	Link to local streets		Within residential area	
	Origin and destination roads			
	No Through traffic			
Urban	Connect local roads		Access individual properties	
	Streets and arterial roads		Only connect to U4	
	Origin/destination along street		Commercial activity	
	<2-3km if no through traffic	<2km	Shopping	
	1way/2way	Public transport	Residential	Residential
	CBD streets	Residential streets		

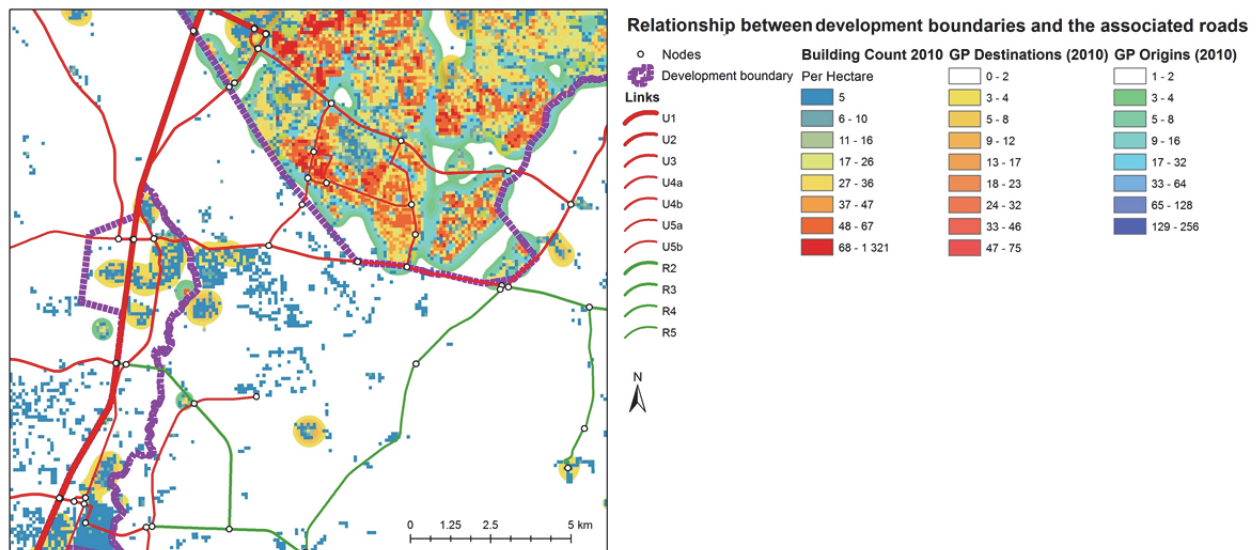


Figure 3. Functional road classification as affected by the development boundary and calculated building densities per hectare

Thereafter six classifications were made: (1) principal arterial, (2) major arterial, (3) minor arterial, (4) collector road, (5) local road and (6) walkway (path or track). The summarised criteria for these classes can be found in Table 1. The walkway (path/track) classification was omitted from this study as only provincial roads were taken into consideration.

5. Results and Discussion

An accurate and fully classified RNI may aid in the planning, design and construction of new roads becoming more efficient, as the function of a road can determine whether it should be upgraded, needs supporting roads, or needs no

interference. This, in turn, allows for the most economic decision to be made regarding the upgrade of the road(s) in question.

An effective road network encompasses roads that provide adequate connectivity and mobility between land uses. This function has to be appropriate at the micro, meso and macro scales to be effective. The nodes and links created a visual representation of the spatial connectivity and mobility of the entire road network. Figure 4 shows the final Gauteng RNI, split into its respective nodes and links. More nodes exist in busier areas with numerous intersections. A common misconception exists that if nodes are far apart and the associated links are fairly equal in length, then the road in question is homogeneous. However, it can be seen from the many attributes that most roads are heterogeneous in nature (Xie and Levinson, 2007). Figure 4 also shows an overview of the fully classified network of provincial roads.

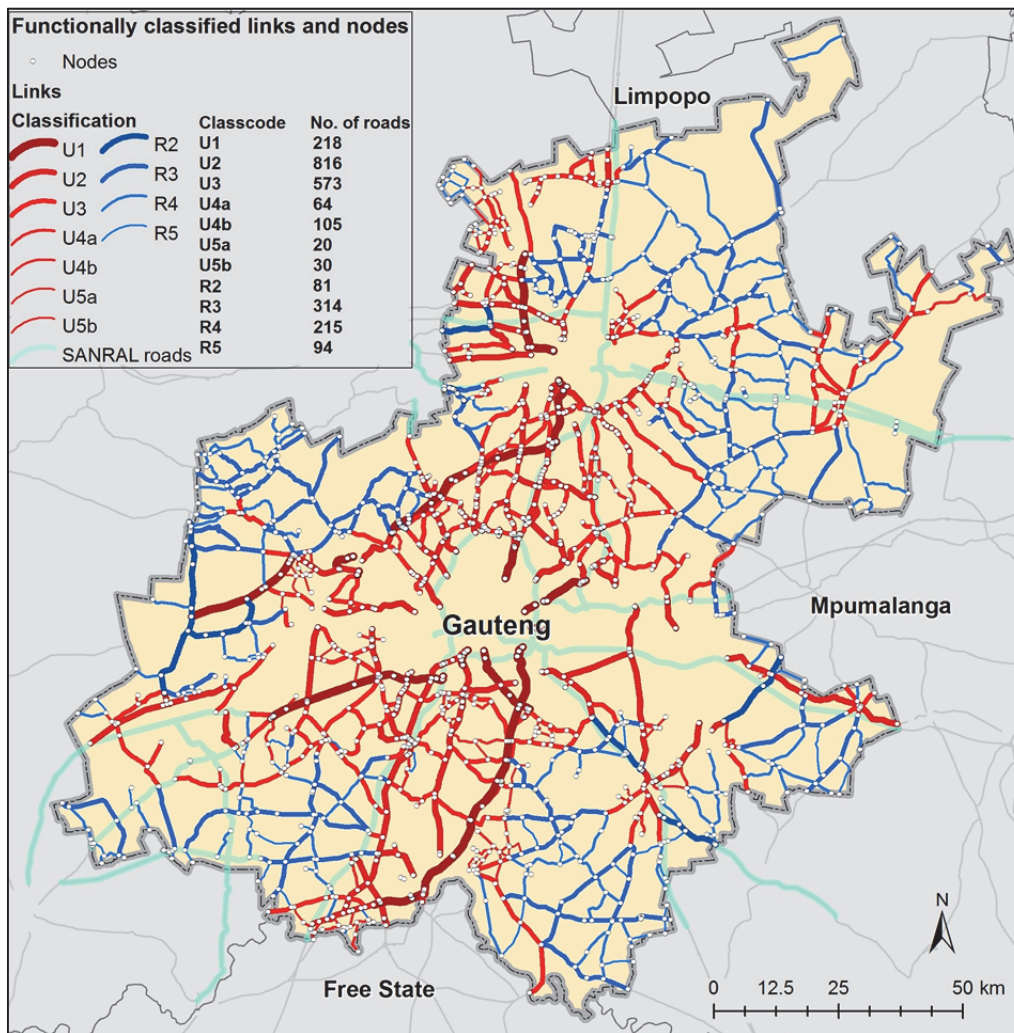


Figure 4. The Gauteng provincial road network, split into nodes and links

With the upgrade of roads, a number of new nodes needed to be added at intersections. In Figure 5(a), Node 1163 was available. It was then split into 1163A-D to denote the nodes of the intersection. New nodes needed to be placed in close proximity to node 1163, and were given the NODE_ID of N1163A-H as there were no closer nodes to associate with. These instances proved to be very challenging and required much care when correctly naming the new nodes as well as describing the links affected. Figure 5(a) also displays dual carriageways represented by its true centrelines, an important upgrade in the 2013/2014 classification. The representation of dual carriageways as two roads of the same entity, allowed for each section (positive/negative) to be treated individually. In the event of road works starting on one side before proceeding to the opposing carriageway for example, the two roads would end up having different road surfaces and quality grades as the process of upgrade occurs. Each road would thus, in effect, have a different maintenance regime, yet the asset register can still be correctly updated accordingly once matched.

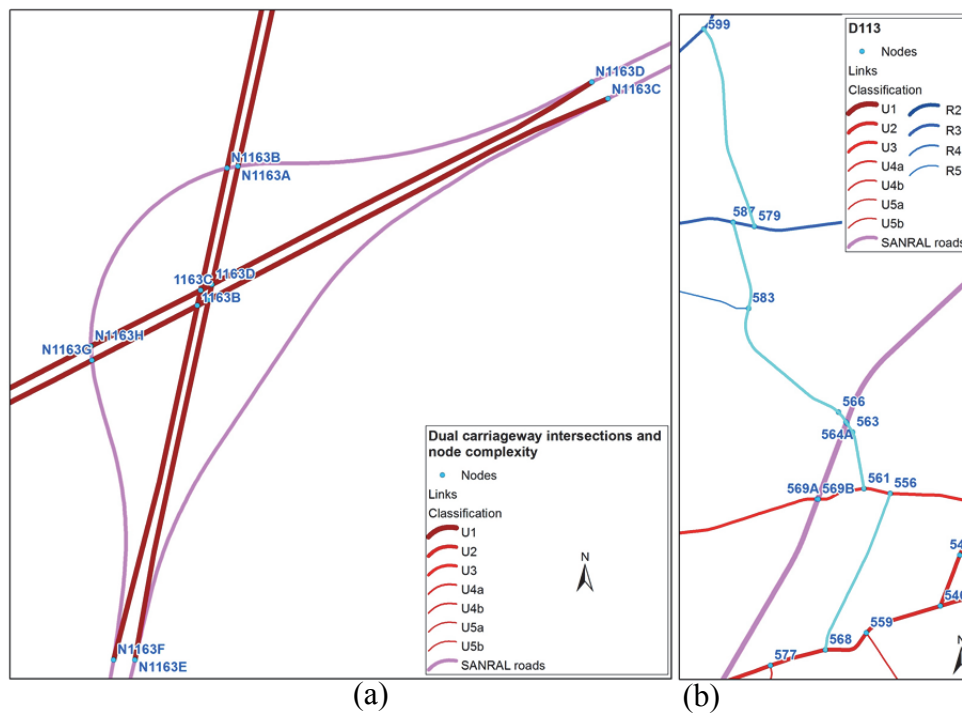


Figure 5. (a) Diagram showing the complexity of a dual carriageway intersection where many new nodes were added (N1163 A-H); (b) Diagram showing the non-continuity of the D1113. Highlighted areas depict the D1113

Link kilometre start and end readings needed to be taken with extreme care as some roads discontinue at intersections with other roads, and continue in different locations. The link kilometres of the re-continuing road would then have to start from the discontinued link kilometre. An example of a non-continuous road can be seen in Figure 5(b). Roads that contained the addition of new links and/or nodes required that all road links be recalculated and the link descriptions updated accordingly. The 2010 dataset contained significant errors in link calculations and representations. The 2013/2014 dataset therefore provided a more accurate representation of the current links and lengths thereof.

The total length of provincial road links in 2010 was 5074.20 km. The almost fourteen percent increase of effective road length in the new 2013/2014 dataset highlights the dynamic nature of road infrastructure development and shows the importance of upgrading the RAMS on a regular basis. It was also found that the shorter links ($\leq 400\text{m}$), often between two dual carriageways, in many cases represented bridges. This could be inferred from the 2010 data attribute table by the presence of an “OORBRUG”-tag (translation: bridge) in the road name field, as well as through visual inspection of the satellite imagery. The link location and kilometre reading therefore also assisted in spatially locating bridges as part of the Bridge Management System, a subcomponent of the RAMS.

The process of accurately representing the road centrelines spatially was largely a manual process. The 2010 dataset was the last available functional classification to make reference to, and so after a four year lapse, there were many changes within the road infrastructure. Not only does the manual interpretation contribute to an element of human error, but the process proved to be tedious and time consuming. The GIS workflow developed, accompanied by the TRH26 guidelines for the functional classification of roads, allowed for a systematic approach to upgrading the provincial road network whilst provided clear-cut decision making rules for the classification process. As the road links were considered in detail, the road in entirety had to be considered as well, to ensure that the road length rules applied to the classification of the road. In certain instances the road classification differed from link to link if the function changed from rural to urban, or vice versa, and this was accounted for in the guidelines.

An accurate road network also allows for measures of road access to be placed where needed. Access control, such as toll gates, enables a higher speed on arterial roads with less congestion. Uncontrolled access can impact on factors such as air quality, damage, road accident damage and traffic, and could, if not controlled, lead to new roads being constructed. Road access systems allow for this to be monitored and maintained, so as to avoid the cost implications of a new road. The increased level of mobility and efficiency can optimise public transport (COTO, 2012).

The results yielded from this study showed a great deal of information about the roads that our citizens use on a daily basis, as well as how they have developed according to traffic pressures since 2010. The importance of having this system in place for every province is important as it will allow for inferences on a number of factors such as road safety and road accident volumes, as well as highlight parts of the road network that need attention. The TRH26 offers a standardised measure of classification, and as a result this can be adopted in every province and the results compared and contrasted. In 2003, the South African National Roads Agency Limited stated that all access to national roads will

be designed in accordance to the National Guidelines for Road Access Management in South Africa manual, as used here (SANRAL, 2003). In this respect, standardisation of functional road classification is optimised.

6. Conclusion

The Western Cape Road Access Guidelines stated that stakeholders sought after a means for managing the transportation and planning parameters for arterial roads (DEAAT, 2012). This study proved that the functional classification method guideline by the TRH26 can provide this as a solution to such sought after measures. The attributes are easily attainable, despite the current manual calculations that need to be done to attain them. The attributes provide a comprehensive state of the network, which allows for other studies to use this data in a variety of other applications. As many of the attributes are quantifiable, with routine maintenance of the RNI, one can track the progression of the network, both spatially and temporally (Xie and Levinson, 2007).

The TRH26 guidelines have proven to be an efficient method of road classification. The process can be optimised by frequent upgrading of the RNI, so as to avoid verifying road works that have occurred a significant period prior to the classification. Since a Node Rule did not exist prior to this study, a protocol had to be agreed upon and was followed, however further attention needs to be given to this aspect, as new roads will continuously be added to the road network.

An up-to-date classified road network impacts on areas of finance, maintenance and construction, which allows for optimal and efficient urban and rural development. This project also served as the CSIR's contribution to the TRH guidelines as a working document to be updated with every future classification as situations change and data can be better represented.

The updated road network will allow for more accurate traffic studies and analyses. The traffic counts used by the CSIR are that of permanent and mobile traffic count stations. These counts have to be shared amongst the number of roads intersecting at the count station. True road centrelines will allow for road traffic counts to be calculated correctly.

7. References

- Breytenbach, A, 2014, 'PG Report: Spatial Urban Dynamics – Spatial Change Tracking', Technical Paper, Technical Report No: CSIR/BE/SPS/IR/2014/0039/B, SIMS No. 5229, CSIR, Pretoria.
- Committee of Transport Officials, 2012, *TRH26 South African Road Classification and Access Management Manual*, Version 1.0., viewed 02 May 2015, <<http://www.roadsandtransport.gpg.gov.za/legislation/Documents/SA%20Road%20Classification%20and%20Access%20Manual.pdf>>.
- DEAAT, 2012, *Road Access Guidelines*, Provincial Administration of Western Cape, Department of Economic Affairs, Agriculture & Tourism: Transport Branch, Cape Town, viewed 02 May 2015, <https://www.westerncape.gov.za/other/2012/8/road_access_guidelines_latest.pdf>.
- Fitzpatrick, K, Carlson, P, Brewer, M & Wooldridge, MD, 2003, 'Design speed, operating speed, and posted speed limit practices', *82nd Annual Meeting of the Transportation Research Board, Washington, DC*. 2003.
- Holderness, C.K, & Stander, H.J, 2008, 'Road Access Management in Practice: The Port Elizabeth Case', *Proceedings of the 27th Southern African Transport Conference (SATC 2008)*, Pretoria, South Africa, July 2008, pp 579-592.
- Jaarsma, CF, 1997, 'Approaches for the planning of rural road networks according to sustainable land use planning', *Landscape and urban planning*, vol 39, no. 1, pp 47-54.
- Rodin, J, 2013, *Roads classification in accordance with national policy*, media release, 01 November 2013, Engineering News, viewed 24 April 2015, <<http://www.engineeringnews.co.za/article/roads-classification-in-accordance-with-national-policy-2013-11-01>>.
- Sampson, J, 2010, Roads classification and access management: What authorities need to know, Presentation to IRF/SARF 2010 Conference, Preserving Africa's Road Network, viewed 02 May 2015, <http://www.teachamerica.com/accessmanagement.info/pdf/2_1145-1215sampson.pdf>.
- SANRAL, 2003, *Policy in respect of road planning and design*, views 16 May 2015, <<http://www.nra.co.za/content/policyinrespectofroadplanninganddesign2003.pdf>>.
- Wilkie, D, Sewall, J & Lin, M C, 2012, 'Transforming GIS data into functional road models for large-scale traffic simulation', *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 6, pp. 890-901.
- Xie, F & Levinson, D, 2000, 'Measuring the structure of road networks', *Geographical analysis*, vol. 39, no. 3, pp. 336-356.

Ye, FM, Lin, SU & Tang, JL, 2006, 'Automatic road extraction using particle filters from high resolution images'. *Journal of China University of Mining and Technology*, vol 16, no. 4, pp. 490-493.

Acknowledgements

We wish to acknowledge the GPDRT for allowing the use of their data for this publication, as well as Pierre du Plessis for his help.

Addendum A

Attributes used in 2013/2014 classification of the Gauteng provincial road network:

Attribute	Description
GIS_ID	Unique sequential identifier for each entity in the dataset
ROAD	Road name
DIRECTION	The direction in which the road continues
CLASSCODE	Classification code
CLASSDESC	Classification description
SURFACE	Paved or gravel
LINK_NO	Unique identifier for links within a road
LINK_KM	Length of link
FROM_DESC	Description of where the link is departing from
TO_DESC	Description of link destination
FROM_KM	Sequential distance following previous link
TO_KM	End of the link's kilometres
FROM_NODE	Node at the beginning of the link
TO_NODE	Node at the end of the link
FNODETYPE	Type of node at the beginning of the link
TNODETYPE	Type of node at the end of the link
CATEGORY	Urban or rural
ROADTYPE	Grade of road according to the previous classification system
TYPE_DESC	Description of the type of road
FNC_CODE	Functional code, according to the previous classification system
FNC_CLASS	Functional classification, according to the previous classification system
ROUTE_NR	Route number, often found on road signs
ROUTE_DIR	Route direction; positive, negative or omnidirectional
OWNER	Owner of the road
YEAR_SURV	Year the road was last surveyed in
DBKEY	Unique road ID
KM_DIFF	Error factor from equation $[\text{Abs}(([\text{TO_KM}] - [\text{FROM_KM}])) / ([\text{TKM_EXCEL}] - [\text{FKM_EXCEL}]) * 100) - 100]$
FRM_EXCEL	Distance at the beginning of a link, according to the 2010 dataset spreadsheet
TKM_EXCEL	Distance at the end of a link, according to the 2010 dataset spreadsheet