Part A – Literature Review

Paper 1

Ringrose, S., Vanderpost, C. and Matheson, W. (1996) 'The use of integrated remotely sensed and GIS data to determine causes of vegetation cover change in southern Botswana', Applied Geography, 16(3), pp. 225–242.

Summary

The aim of this paper is to provide spatial information about different vegetation types in the dry savanna area in Malwelwe, Botswana. Then to combine their distribution with village and borehole locations to consider the impacts of human pressures upon vegetation cover. Remotely sensed images were used to identify the distribution of four main vegetation types using spectral reflectance characteristics. These locations were mapped and converted to polygons for GIS-based analysis. The locations of villages and boreholes were then digitised and buffers created around them. The percentage of each vegetation type within each buffer zone was calculated to assess the intensity and nature of human pressures upon natural vegetation cover.

Results showed a positive correlation between the location of villages and boreholes and the location and extent of depleted land and sparse vegetation cover.

Algorithm and Software

Buffer Analysis. ARC/INFO software (first version of ArcGIS)

Data and Data Source

Buffers of 2, 3, 4 and 6-km radii were created around point locations of villages and boreholes. Their locations were digitised from available maps that ranged in scale.

Results

After the buffers were created around the villages and boreholes they were overlaid on the vegetation map (Figure 1). The buffers created were circular, as they were created around points, rather than lines or polygons. The percentage of the buffer zone for each vegetation type was then calculated. About 60% of the total sparse degraded area was found within the 2km buffers of either a village or borehole and about 95% within 3km. These results therefore established a strong relationship between the degradation of land and the presence of human activity.

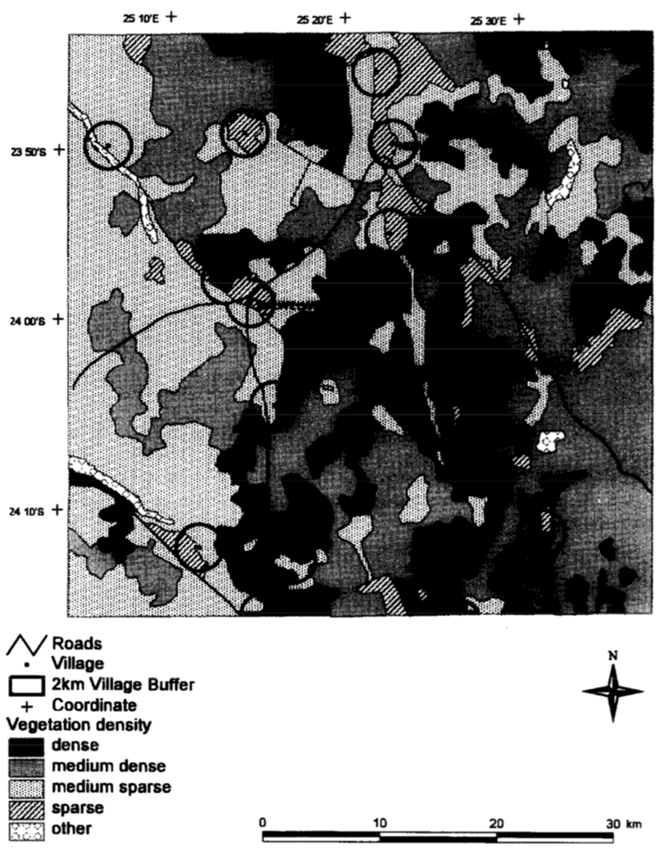

Figure 1: *The map included within the paper showing the extent of the 2km buffers around the villages in relation to the vegetation distribution.*

<u>Data and Algorithm Quality Issues</u>

The first issue not mentioned by the authors was that the buffers were created around points, which represented the villages and boreholes. For boreholes, a point would be an accurate representation. However, for a village, defined in this paper as having a minimum of 50 residents, the placement of the points may have created inaccuracies. This is because the size of a village and the ambiguity involved with defining its extent, makes it difficult to be accurately represented as a point. Furthermore, it is not made clear whether a polygon was first digitised and the centroid generated, if the points were digitised from existing points on the available maps, or if the centre of the village was judged manually when digitising. Consequently, the buffers created may not be an entirely accurate and precise representation of the distances from the villages in reality.

Another issue not addressed was the way in which buffers measure distance in an 'as-the-crow-flies' manner. This leads to the area the buffer represents not considering distance travelled via roads or paths. In this case, there may be a route between two villages that people walk or drive along causing vegetation to be damaged at a larger distance than the buffer distance calculates. Consequently, conclusions made about the distance that human impact affects vegetation will not be entirely correct.

Paper 2

Li, X., Zhang, L. and Liang, C. (2010) 'A GIS-based buffer gradient analysis on spatiotemporal dynamics of urban expansion in Shanghai and its major satellite cities', *Procedia Environmental Sciences*, International Conference on Ecological Informatics and Ecosystem Conservation (ISEIS 2010), 2(Supplement C), pp. 1139–1156.

Summary

The study explores the spatial-temporal characteristics of urban expansion in the Shanghai region, China, and the urbanisation of its eight major satellite cities. Satellite images for the years 1987, 1990, 1995 and 2000 were firstly used to extract the extent of the urban areas. Circular buffer zones were then created around Shanghai and the other cities. Changes in density of the urban areas were quantified for each buffer zone over the time period to determine the urbanisation rate and compare it between satellite cities.

The results show that the changes and rate of urbanisation was largely influenced by distance from the urban centre, with distinct spatial variations primarily resulting from interactions between the urbanisation of Shanghai with that of its satellite cities.

Algorithm and Software

Buffer Analysis. The specific software was not mentioned. The authors only write “GIS-based buffer analysis was adopted in our research…”, however it is likely the software was ArcGIS.

Data and Data Source

The buffer zones were created surrounding the centre of Shanghai and the satellite cities. The urban area of each was determined from Landsat data. For Shanghai, buffer zones were created around a polygon representing the urban area. For the satellite cities, buffer zones were created around the geometric centre (a point) of each city, as it was difficult to determine the urban centre.

Results

Two different buffer systems were established, both involving the creation of buffers zones increasing in size. The first system created 2km width buffers covering the entire metropolitan area surrounding the urban area of Shanghai (Figure 2, left). The second system created 10, 1km width buffers around Shanghai’s urban centre and 10, 0.5km width buffers around the centre of each satellite city (Figure 2, right).

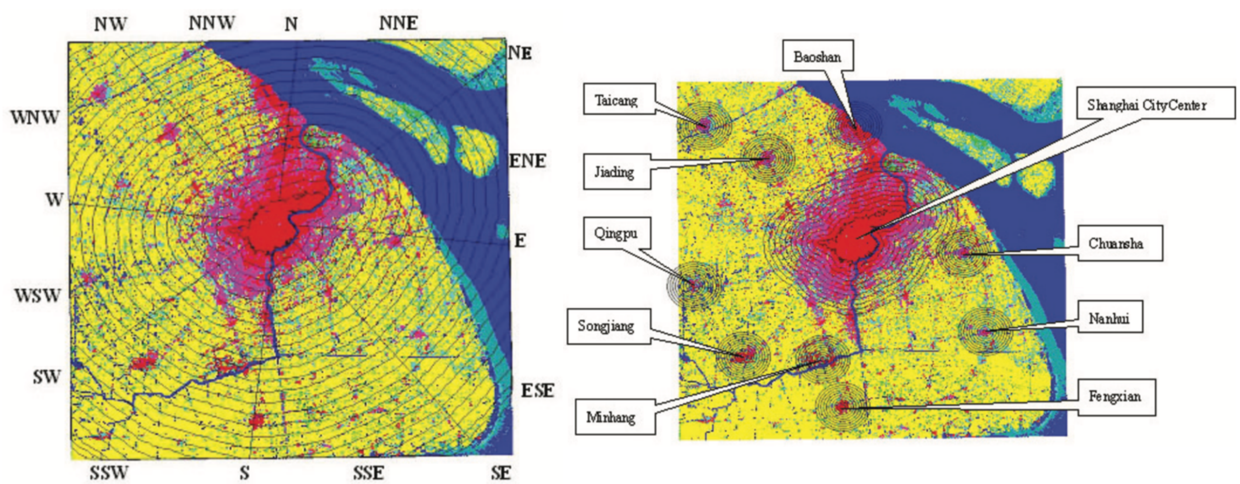

Figure 2: *The first buffer zone system with 2km width buffers created around only Shanghai’s urban centre (left). The second buffer zone system with 1km and 0.5km buffers created around both Shanghai and its eight satellite cities, respectively (right).*

Data and Algorithm Quality Issues

The authors of this paper took into consideration the uncertainty of creating buffer zones around a point when representing an ambiguous area. It was stated that the urban area was “used to **represent** the urban centre as a **baseline**” (emphasis added), thus addressing the ambiguity. Furthermore, for the satellite cities, the urban centre was “generally not obvious”, so the geometric centre was generated for buffer creation instead.

These authors also have not considered that buffers are created using an ‘as-the-crow-flies’ distance. This means the buffers are the same distance from the centre all the way around in a circular manner. However, urban growth does not occur at the same rate all around the city. Therefore, when considering the distance that the city has urbanised over time, the figure will be generalised to whichever area has reached the furthest buffer zone.

It has not been stated in this paper whether the buffers were created using a Euclidean or geodesic method. As a World projected coordinate system was used and the study area was within one UTM zone, Euclidean buffers would have worked well for this research. Geodesic buffers, however, take into account the actual shape of the earth, so are much more accurate for analysis on a global scale.

Part B – GIS Comparison

1 Introduction

When carrying out GIS work, it is vital for the user to understand the tools and software they are using. Without this knowledge efficiency may be reduced, or incorrect results generated.

For example, in May 2003, The Economist published an article (corrected article: 2003b) showing concentric circles for ranges that could be reached by missiles launched from North Korea. However, due to a fundamental misunderstanding of the Mercator projection, the ranges identified were largely underestimated (Chrisman, 2017). The creator obviously did not know how to generate accurate distances from the buffer tool and therefore did not successfully implement it. Consequently, it is important for GIS users to fully understand the tool they wish to use and how it can be used incorrectly. Through testing exception cases for an algorithm within this assignment, the motivation in terms of reducing errors from lack of knowledge is made clear.

In terms of efficiency, choice of software may be influential as each package is marked by limitations. Consequently, one package may better fit for a certain purpose whilst another package better for a different purpose. Additionally, there may be variation between the way different packages carry out a single algorithm. Hence, comparing the implementation of an algorithm in ArcGIS and QGIS, provides understanding into which package may be better equipped for purpose.

2 Background

2.1 Buffer

The buffer tool creates polygons of a specific distance around a discrete object. Buffers are used regularly for visualising the distance from an object, for example in site selection to define that a car park must be located 100m from a shopping centre.

The buffer zone is created firstly by establishing the boundary around every point, node and line segment of the given object, which are then combined to create the buffer zone for that object (Figure 1).

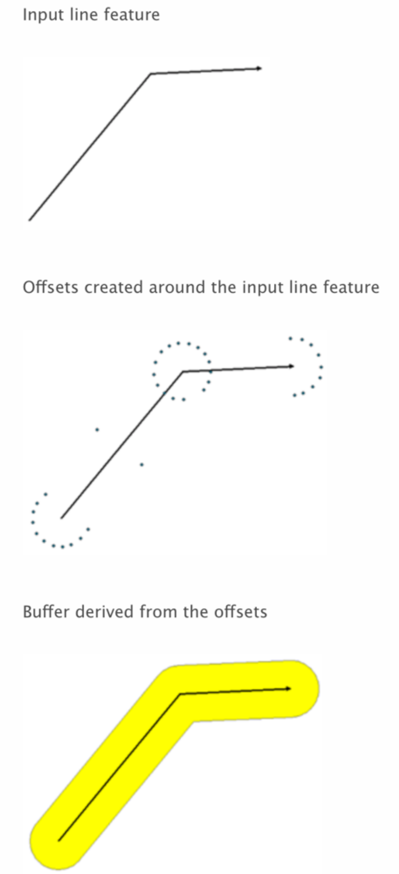

Figure 1: *The process in which a buffer is created around a line object (ESRI, n.d.)*

Potential failures may occur when creating buffers on a global scale, due to the coordinate system causing projection distortions. This can be eliminated using geodesic buffers, which will always produce geographically accurate results as they are not affected by distortions introduced by projected coordinate systems (Flater, 2011). Euclidean buffering, on the other hand, can produce misleading and technically incorrect buffers. Secondly, the buffers created may not accurately represent distance when considering a complicated building shape.

2.2 Centroids

Centroids are generally defined as the nominal centre of gravity for an object (De Smith et al., 2015). Therefore, creating a centroid for a polygon is done through using its centre of gravity to calculate the centre point.

Centroids are used to provide a default position for labelling, or act as a representation for analytical purposes. For example, when calculating the distance between two polygons, the centroid may be used to give an appropriate representation for each polygon.

Although not technically a failure, the centroid may be calculated to be outside of the feature's boundary due to the centre of gravity falling on the outside. Despite this being an accurate result, it may not be fit for purpose. For example, if the centroids are being calculated for labelling.

3 Data

The data used is from two sources: Open Street Map via Geofabrik and Natural Earth. The former provided polygon shape files for buildings in Kent, which were clipped to the town of Gillingham. From Natural Earth, a point layer was downloaded for the locations of the world's populated places, in which 6 cities were selected. Two polygons were also manually digitised using QGIS for centroid testing.

4 Method

Firstly, to carry out buffer and centroid generation, the appropriate tools had to be selected in ArcGIS and QGIS (Table 1) and different cases identified (Table 2).

Table 1: *The names of the tools used in ArcGIS and QGIS for generating buffers and centroids.*

Software	**Tool Name**	
	Buffer	**Centroid**
ArcGIS	Buffer	Feature to point
QGIS	Fixed distance buffer	Polygon centroids

Table 2: *The three different cases identified in order to test each tool.*

Buffer	***Centroid***
Individual 'simple' and 'complicated' building polygons used to analyse the intricate detail of buffer creation	One individual building polygon used to analyse the intricate detail of centroid creation
All 9435 building polygons within the town of Gillingham used to assess running time	All 9435 building polygons within the town of Gillingham used to assess running time
Six city points used to assess the impact of coordinate systems and consequent buffer distortion	The polygon shapes manually digitised to assess exception cases

5 Results

The following section shows the results for the outlined cases in each package using two different coordinate systems. The time taken to carry out the analysis for comparison was difficult to assess because QGIS does not give execution time information. Consequently, to provide a comparison, the time taken for the execution of all 9435 buildings in Gillingham in each tool was measured using a stopwatch. Obviously, this is not the actual time taken by the computer, however, it gives a relative time for comparison. Furthermore, the time taken using a stopwatch compared to times given by ArcGIS were accurate within seconds so this method is fairly reliable.

5.1 Buffer

Firstly, buffers around the individual building selected were generated (Figures 2 and 3) in the WGS coordinate system. In QGIS, there was initial confusion regarding units used for distance. After consulting the GIS stack exchange page, it was noticed that many other QGIS users have been confused by this, as the buffer size is applied in the layer CRS units (Stack Exchange User, 2012). In this case in degrees, which was set to 0.0001. In ArcGIS, however, the units could be specified so was set to 10m. Buffers were then created around a 'simple' and 'complicated' building shape (Figures 4 and 5).

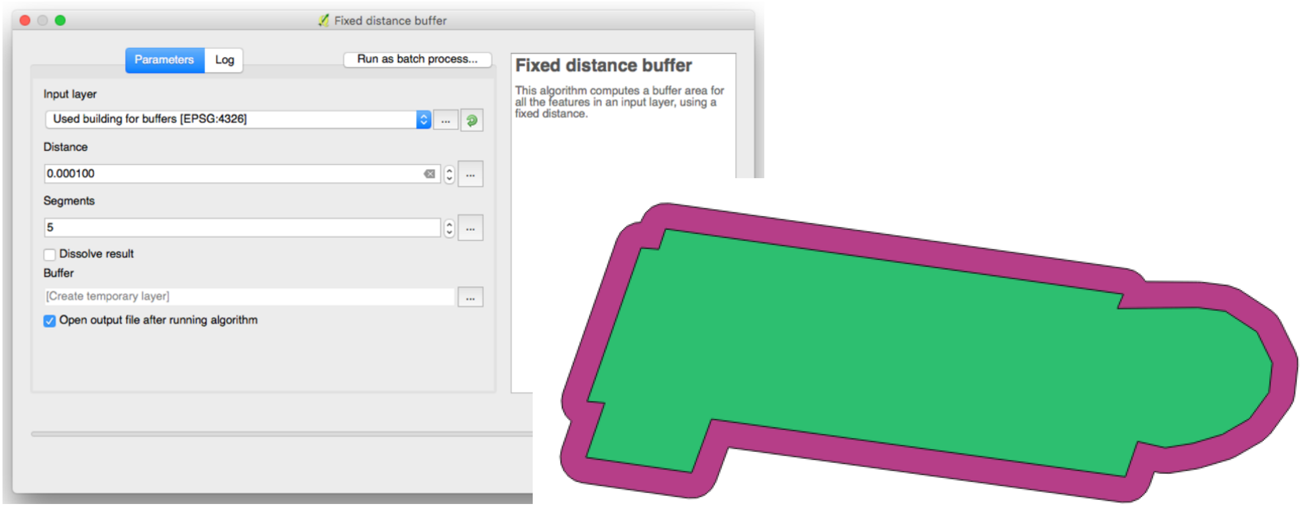

Figure 2: *The buffer created around the building in QGIS.*

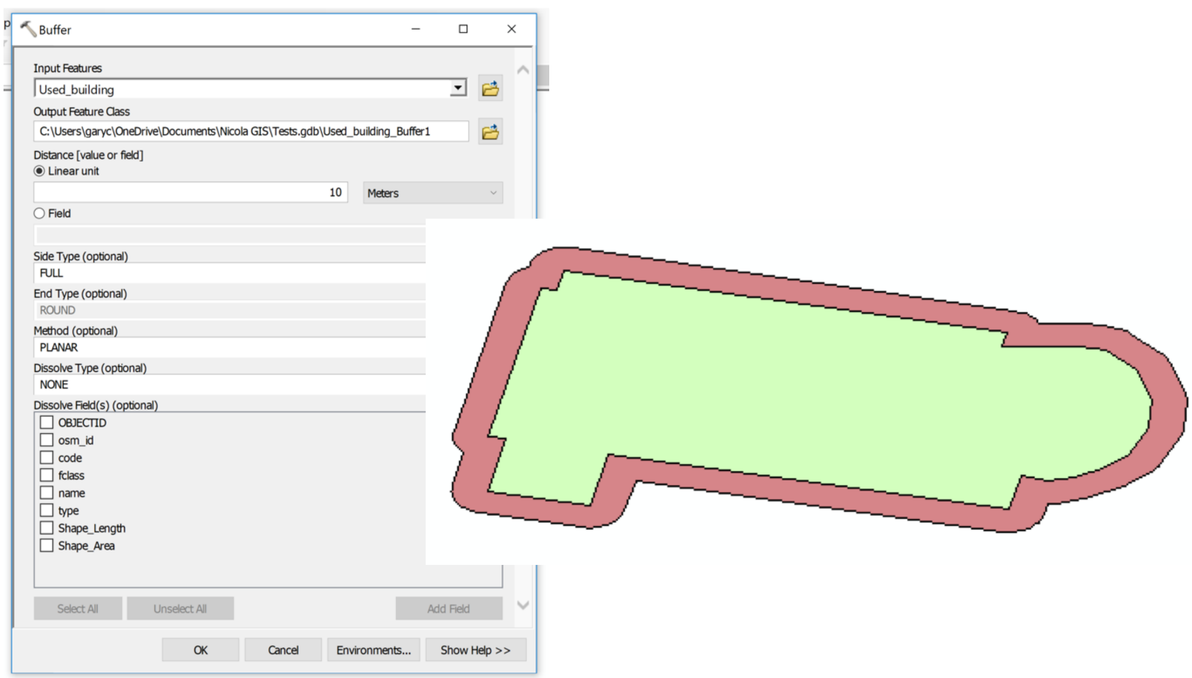

Figure 3: *The buffer created around the building in ArcGIS.*

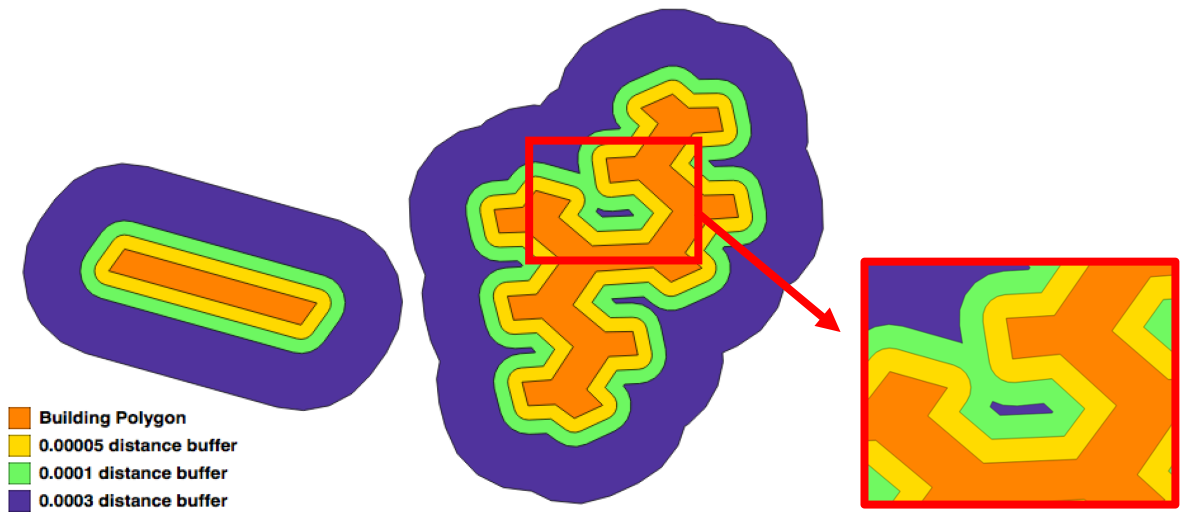

Figure 4: *Three varying distance buffers created around 'simple' (left) and 'complicated' (right) building shapes in the WGS coordinate system in QGIS. The zoomed in inset shows how the green buffer distance creates a buffer with a hole within it, due to the complicated shape of the building. If showing the Euclidean distance that can be walked from any part of the building then this would not cause a problem. However, if creating buffers for representing pollution extent, for example, this hole would cause an error.*

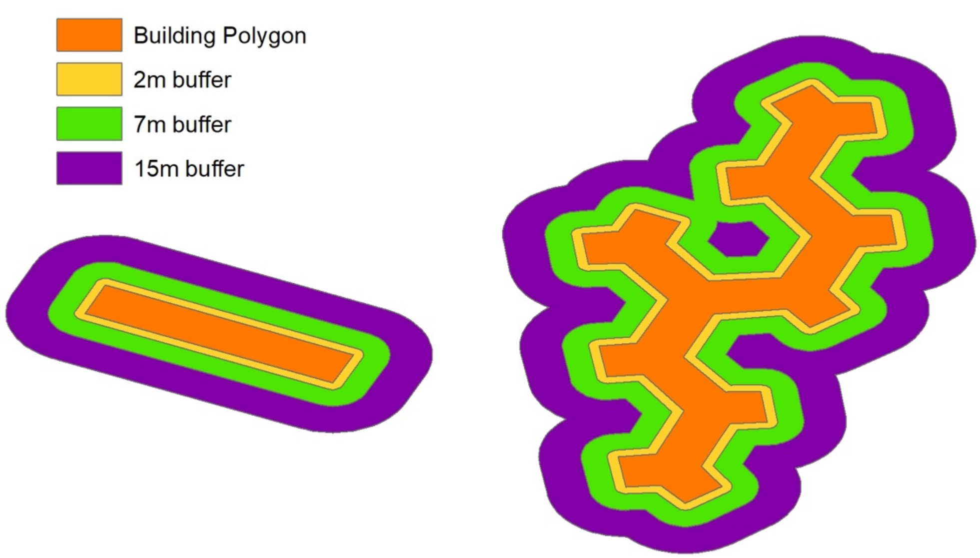

Figure 5: *Three varying distance buffers created around 'simple' (left) and 'complicated' (right) building shapes in the WGS coordinate system in ArcGIS. The same issue with the buffer being created with a hole can be seen within the 7m buffer created.*

Secondly, buffers were created around all of the 9435 in Gillingham (Figures 6 and 7). This took 3.68 seconds in QGIS and 38.19 seconds in ArcGIS.

Figure 6: *The buffers created around all of the buildings in Gillingham in QGIS.*

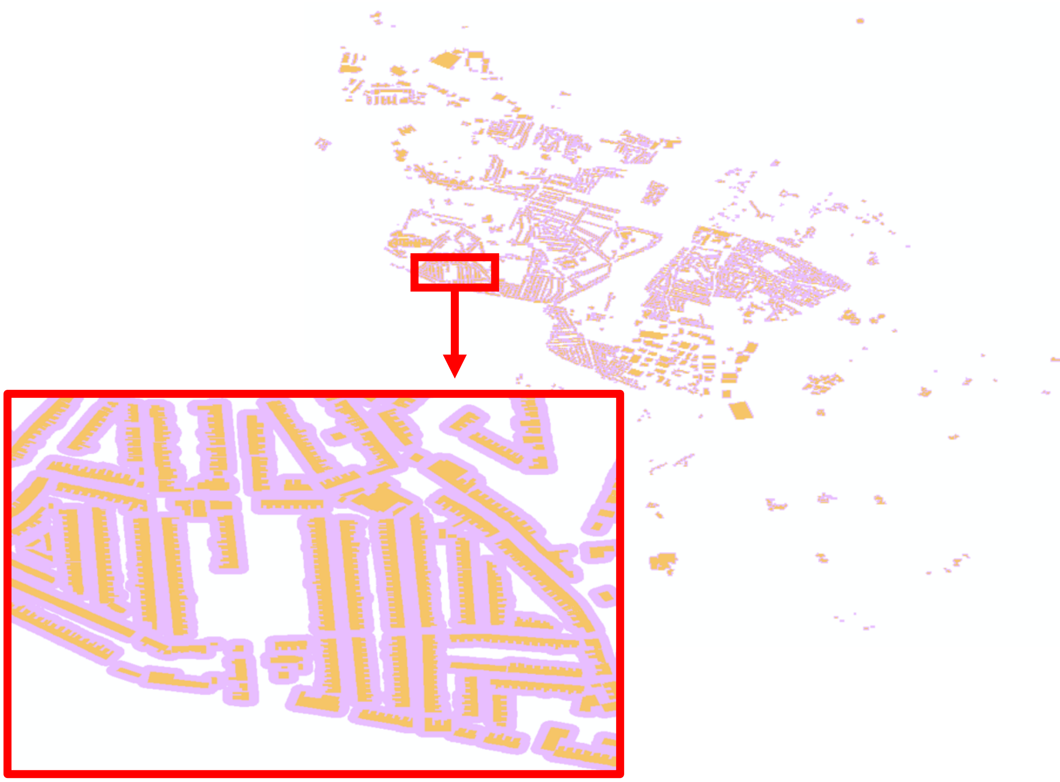

Figure 7: *The buffers created around all of the buildings in Gillingham in ArcGIS.*

Next buffers were generated around points across the world to test how the software created the buffers; using a Euclidean or geodesic method. Esri state that if the input features are dispersed, then geodesic buffers should always be used (n.d.) because they takes into account the shape of the globe.

ArcGIS gives two options – Planar and Geodesic – for buffer method. The former is the default and automatically determines between Euclidean or Geodesic buffers based on the coordinate system. Each of these options were tested to create 1,000km buffers, using the WGS coordinate system and successfully generated geodesic buffers (Figure 8). When creating buffers in the planar method using the British National Grid system, Euclidean buffers were made as it is a projected system.

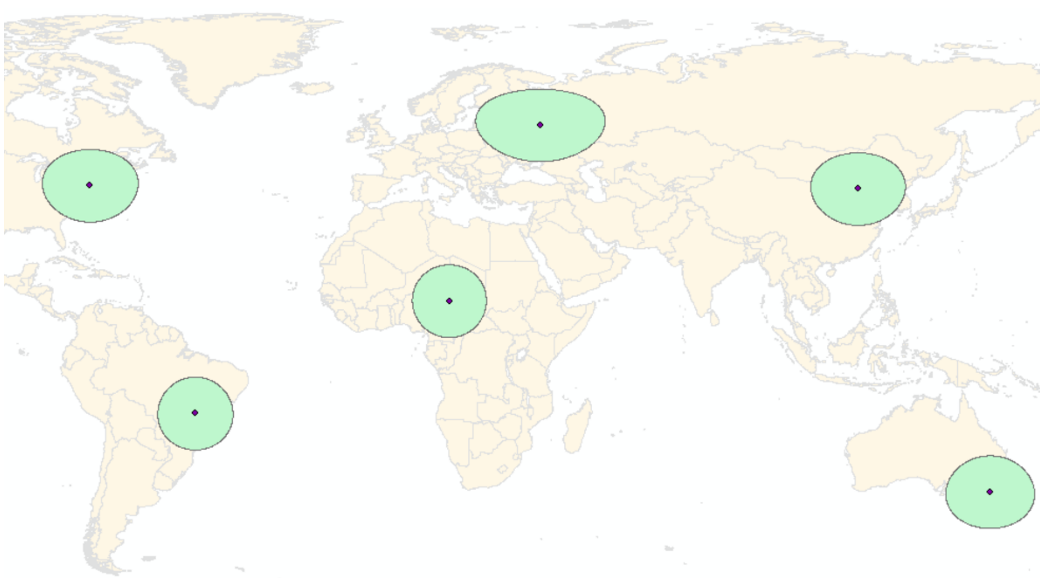

Figure 8: *The geodesic buffers created around the six cities in ArcGIS. The buffers of points further from the equator (e.g. Moscow in Russia) show considerably more distance distortion than those closer to the equator (e.g. N'Djamena in Chad).*

In order to show the tool used incorrectly (as was initially in The Economist, 2003b), the planar method along with an angular units (degrees) was selected. This generated Euclidean buffers, which are inaccurate in high distortion areas away from the equator (Figure 9).

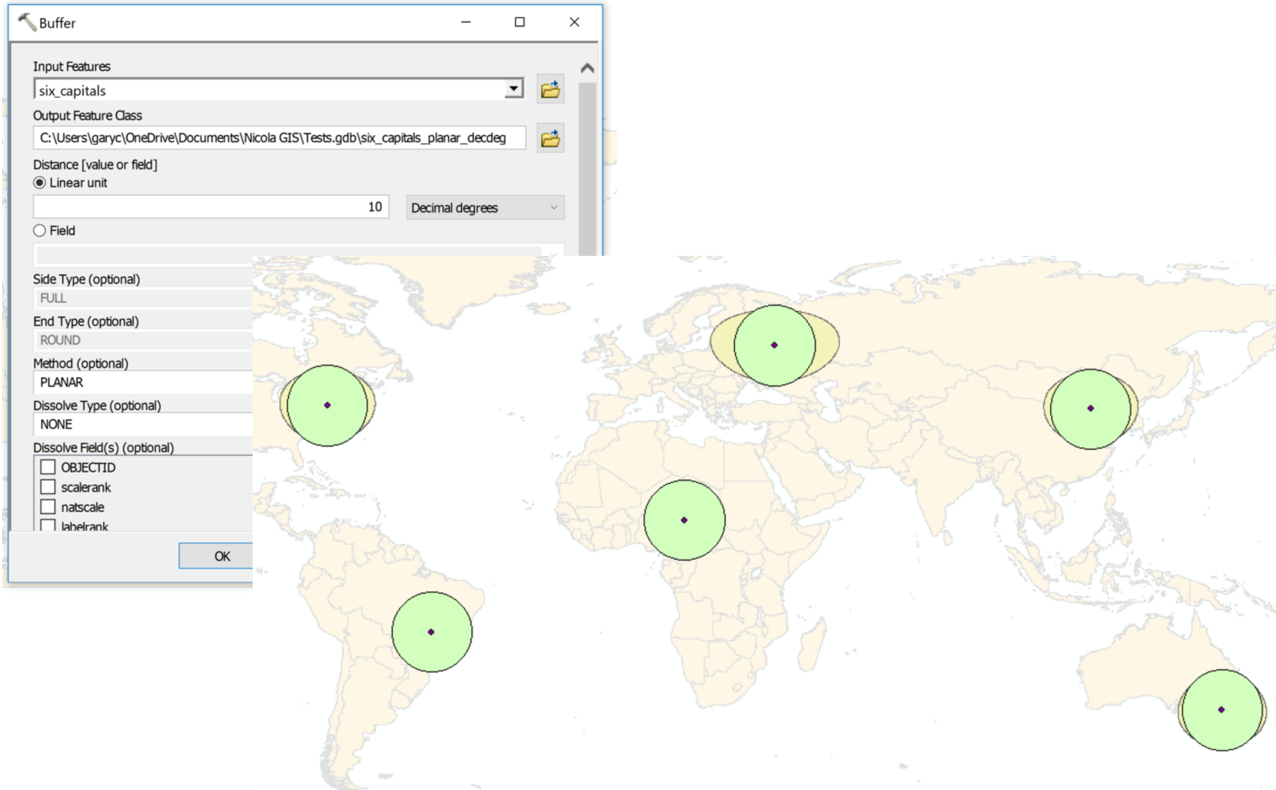

Figure 9: *The Euclidean buffers created around the six cities in ArcGIS shown in green. The previously made geodesic buffers can be seen in yellow to showcase the inaccuracies created by the buffers at high distortion areas further away from the equator.*

QGIS does not have Euclidean or geodesic buffer options. The buffers created were generated as circles thus suggesting a Euclidean distance was used, rather than geodesic which would be much more accurate (Figure 10).

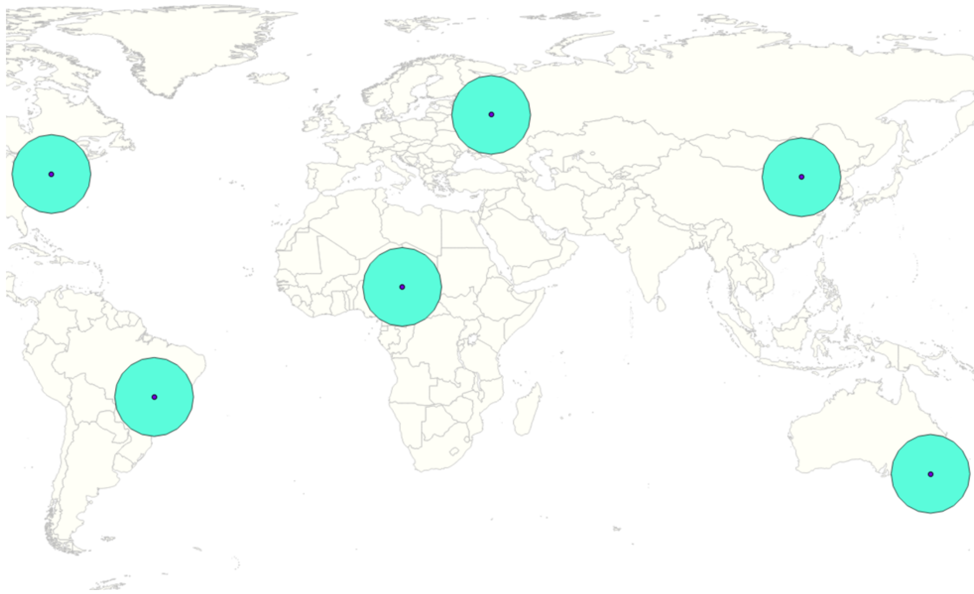

Figure 10: *The buffers created around the six cities in QGIS. It can be seen that the size and shape of each is the same, thus distortion created by the WGS coordinate system has not been accounted for.*

Another issue with the method in which QGIS creates buffers is an input must be provided for the number of 'segments' relating to the number of polygon edges per quarter and will affect the accuracy of the buffer created (Figure 11). The difference in area was calculated using the measure tool to show the difference in accuracy (Figure 12). This creates a further issue when buffering around a point (Figure 13). In ArcGIS, there is no segment tool, therefore all buffers created will be consistent.

Furthermore, a prominent time difference was noted between the time taken for buffers of varying segment size. Consequently, if aiming to generate high accuracy buffers, then the execution time will be higher.

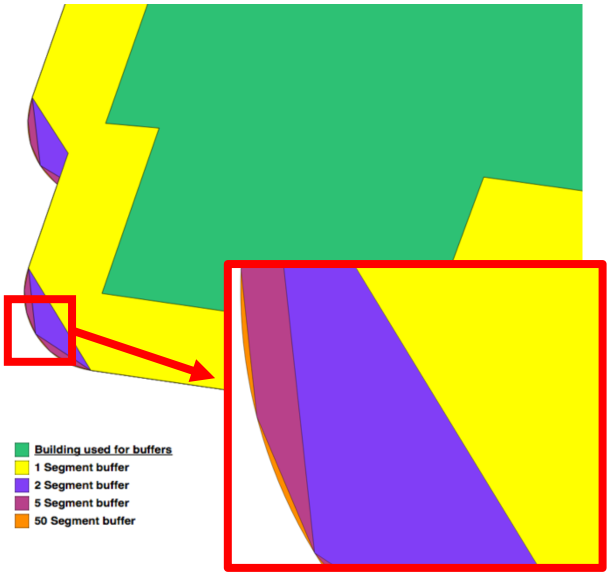

Figure 11: *The four buffers that were created around the building. The default value is 5, so this value was used along with values of 2, 1 and 50. It can be seen that after zooming in on the corner of the building (inset), all four of the buffers have discrepancies in area.*

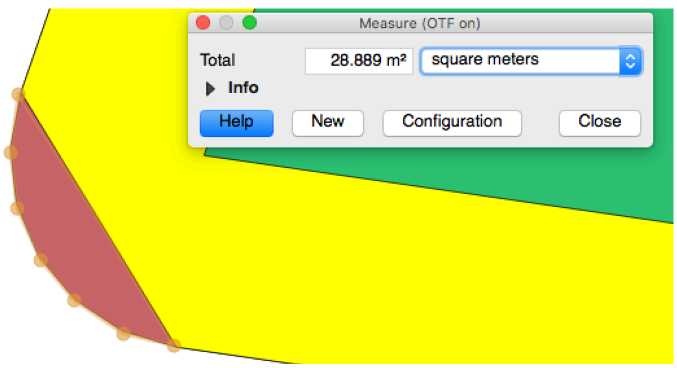

Figure 12: *The difference in area between the 1 segment and 5 segment buffers for the shown corner of the building. It can be seen that there is an approximate 29m^2 difference between the two buffers.*

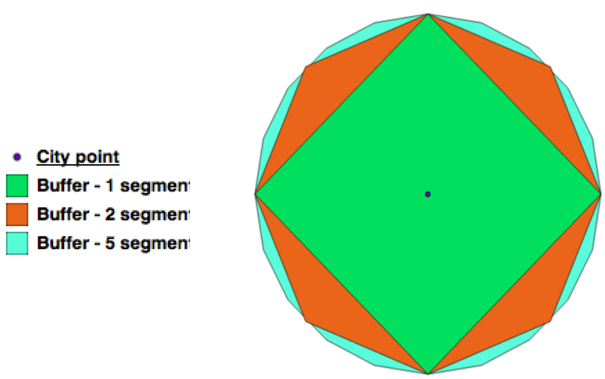

Figure 13: *The issue of choosing a segment value when buffering around a point. A huge variation in results would be generated from each of the three buffers, which may not be apparent to a user with little knowledge of how the buffer tool works in QGIS.*

5.2 Centroids

Firstly, centroids were generated for the individual building selected (Figures 14 and 15) in the WGS system. The coordinates for the centroid point in both packages was the same (0.534668, 51.401034), acting as validation.

Secondly, centroids were calculated for each of the buildings in Gillingham (Figures 16 and 17). This took 1.53 seconds in QGIS and 4.23 seconds in ArcGIS.

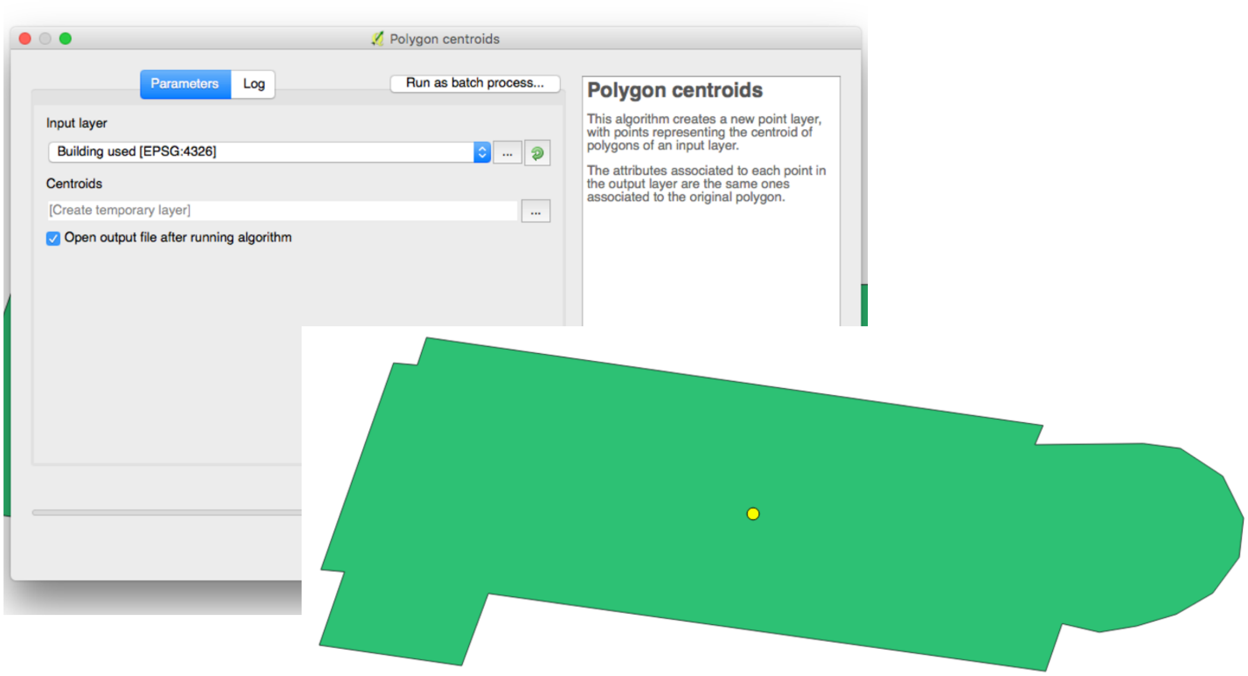

Figure 14: *The polygon centroid calculated for the building in QGIS.*

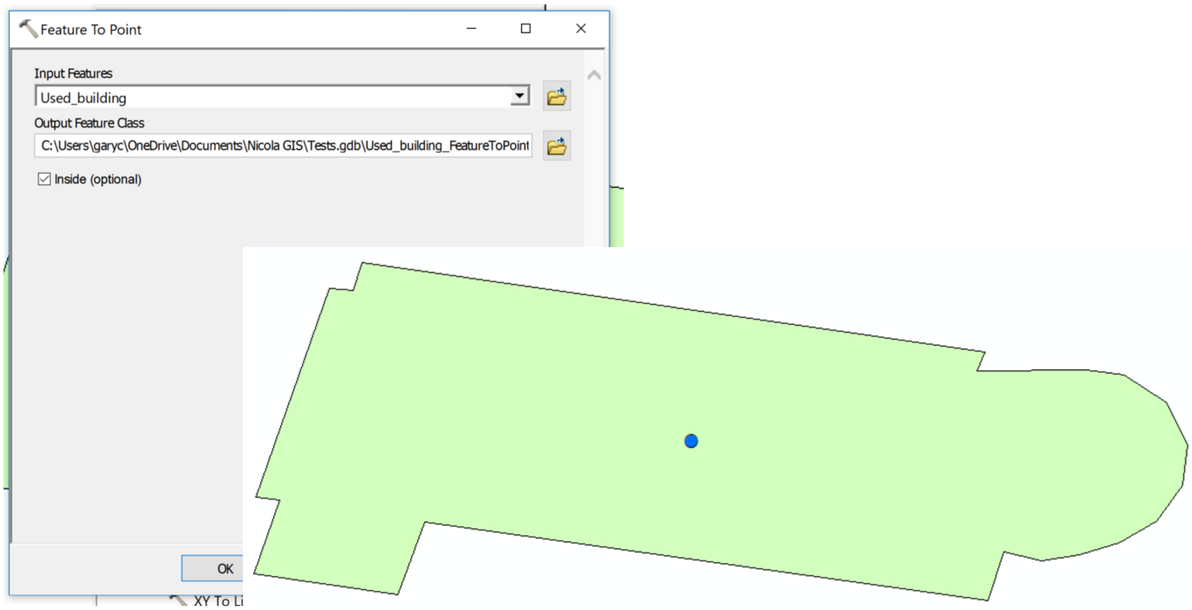

Figure 15: *The polygon centroid calculated for the building in ArcGIS.*

Figure 16: *The centroids for every building polygon in QGIS.*

Figure 17: *The centroids for every building polygon in ArcGIS.*

Next, two polygons were manually digitised in QGIS to test exception cases. The main exception case considered was for polygons where the centre of gravity falls outside of the polygon boundary, thus most likely generating a centroid outside of the polygon. The centroids generated were, as thought, outside of the boundary (Figure 18). The coordinates for the centroids points in QGIS and ArcGIS, in the WGS system, were once again the same. Due to both centroid points falling 'inside' of the polygons, it becomes difficult to distinguish between them presenting an issue if the points were used for labelling the polygons.

In ArcGIS, there is an option called 'Inside' for ensuring the generated centroid point is within the polygon (Figure 18). QGIS's centroid tool is not equipped with this option, therefore ensuring the centroids each fell inside of the polygon could become tedious. This may involve manually editing each point, downloading a plugin or using the Random Point tool, although these points would not be centroids.

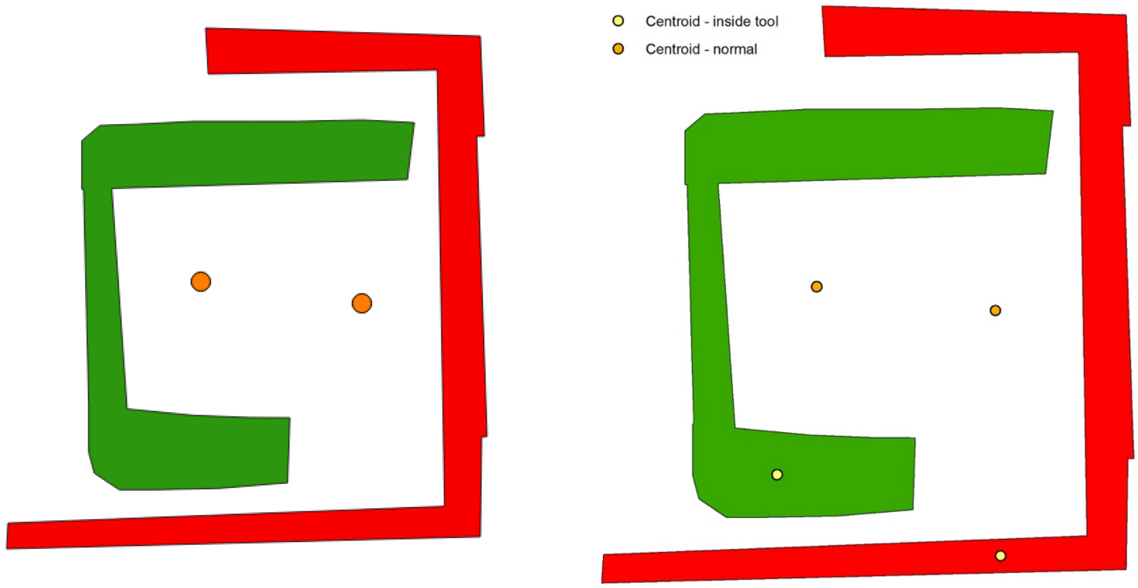

Figure 18: *The centroid points generated for the two manually digitised polygons in QGIS (left) and ArcGIS (right). Two sets of centroid points were created in ArcGIS using the tool normally and then selecting the 'inside' setting to ensure the points fall within the polygon boundary.*

6 Discussion

The results, in general, were as expected for both tools. However, key differences were identified between the ways in which buffers and centroid points were created due to the functionality of each tool in each package (Table 3).

For the buffer tool, both packages gave the same result for the buildings within Gillingham and the simple and complicated buildings. Although, the ability to enter the segment size in QGIS can create vastly different buffer shapes, thus leading to variation. However, if the same segment size was used consistently throughout a project the results would be reliable. It would be important, therefore, for the user to ensure they were aware of this.

The differences between the packages when creating buffers on a global scale in a geographic coordinate system were distinct, thus leading to different results. This is due to the ability to create Euclidean and Geodesic buffers in ArcGIS but not QGIS. In practice, if using a projected coordinate system, such as British National Grid, there would be no issue with using

QGIS. However, when using a geographic coordinate system, such as WGS, on a global scale project it would be important to use geodesic buffers.

Table 3: *The key differences between the ways in which QGIS and ArcGIS generate buffers and centroids identified from the results.*

QGIS	**ArcGIS**
BUFFER	
The distance units are the same as the CRS of the layer, which can create confusion.	The distance units can be selected allowing for easy use when creating buffers.
There is no option to select either Euclidean of geodesic buffers. The results suggest Euclidean buffers were made despite the CRS being a geographic system (WGS).	A choice can be made between generating Euclidean or geodesic buffers, allowing for reduced distortion depending on the CRS.
The number of segments must be inputted, which can be advantageous in reducing execution speed, but can also lead to the buffers being over simplified and not representative.	The shape of the buffers is automatically created with no choice being given towards the segment size.
CENTROID	
There is no option to ensure the centroids are within the polygon boundary.	The 'inside' selection can be used to ensure the centroids are within the polygon boundary. This is advantageous for many purposes, such as using centroids to label.

Likewise with buffering, the results for the centroid tool gave the same results in both packages. Additionally, the results remained the same (despite expected alignment) within both projections. The only difference observed relates to the functionality of the tool in QGIS, as there is no option to ensure the centroid point is located within the polygon boundary. ArcGIS has this option and therefore handles exception cases better.

The time taken for the tool to execute in each package revealed a somewhat surprising disparity, with ArcGIS performing significantly slower. Consequently, if the result will not be affected by the limitations of QGIS outlined in this report, then using this package may be more efficient in practice. Additionally, QGIS has the ability to install plug-ins that may allow for the errors identified to be eliminated. For example, there is a plug-in named 'realcentroid' that says it always generates an internal centroid.

Finally, there was no observed impact from the data quality in creating buffers or centroids in this assessment. However, in practice there may be accuracy errors generated, such as the centroid falling in the wrong position if the data source was not completely accurate. Additionally, for other algorithms there may be an impact caused by data quality, such as interpolation algorithms.

7 Conclusion and Further work

To conclude, it is clear that there are variations between the way in which QGIS and ArcGIS generate both buffers and centroids. In a normal case, such as creating a buffer around an individual building or generating its centroid, there are no differences between the results. However, there are exception cases where results differ, such as creating buffers around objects around the world or choosing to use the 'inside' selection in ArcGIS when generating centroids. Consequently, it is vital that the user has a strong understanding of not only how the algorithm works, but also the way it works within their chosen package, along with any differing functionality the tool may have.

Although a number of exception cases and different scenarios were used to test both tools, there a likely many other potential failure points that have not been considered during this analysis. Additionally, different variations of each algorithm were not included such as line buffering, variable width buffering, or creating a centroid between multiple objects, amongst others. Consequently, further work may consider assessing the results and errors created when using these variations, along with testing the tools using different data sources and coordinate systems.

References

Chrisman, N. R. (2017) ‘Calculating on a round planet’, *International Journal of Geographical Information Science*, 31(4), pp. 637–657.

De Smith, M., Goodchild, M. and Longley, P. (2015) *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*, 5th ed.

ESRI (n.d.) ‘How Buffer (Analysis) works’, [online] Available from: http://pro.arcgis.com/en/pro-app/tool-reference/analysis/how-buffer-analysis-works.htm (06.01.18).

Flater, D. (2011) ‘Understanding Geodesic Buffering’, *ArcUser*.

Stack Exchange User (2012) ‘Understanding QGIS buffer tool units? - Geographic Information Systems Stack Exchange’, [online] Available from: https://gis.stackexchange.com/questions/23856/understanding-qgis-buffer-tool-units (06.01.18).

The Economist (2003a) ‘Correction: North Korea’s missiles’, *The Economist*, [online] Available from: http://www.economist.com/node/1788311 (05.01.18).

The Economist (2003b) ‘When bluff turns deadly’, *The Economist*, [online] Available from: http://www.economist.com/node/1748566 (05.01.18).