Representations, Structures and Algorithms

RQJG1

16140263 | MSc GIS | December 16, 2016

Table of Contents

1 – Literature Review

1.1 – REVIEW OF 'A GIS-BASED METHODOLOGY FOR SAFE SITE SELECTION OF A BUILDING IN A HILLY REGION'

1.1.1 – Summary

'A GIS based methodology for safe site selection of a building in a hilly region' is a peer-reviewed, academic article, published in the journal 'Frontiers of Architectural Research' in March 2016. The article critically reviews a GIS site selection process that uses survey questionnaires and current building regulation information to assist in the identification of locations suitable for safe construction projects. Various ArcGIS algorithms such as weighted overlay and buffering were employed to interpret the raw data and produce a raster image depicting the appropriate suitability of sites for construction with special consideration to safety during construction.

1.1.2 – Algorithm and Software

ESRI ArcMap was identified as the software that was used to perform the analysis of the study area. Various algorithms were used to conduct the research. As part of the site selection analysis, areas were identified as unsuitable due to building regulations such as those within:

- 1.2m of an electric power line,
- 18m of a septic tank,
- 6m from existing building, and
- 2m from natural run-off.

Thus, in order to eliminate such areas from the possible site selection output raster, the ESRI ArcMap buffer algorithm was executed. This allowed the researchers to isolate areas which would conflict with current building regulations and remove them from the study.

1.1.3 – Data and Data Source

Elevation raster data; a road feature class; an electric line feature class; sub-soil dispersion system feature class; existing buildings feature class and a land use raster dataset were imported as the raw data sources for the case study. These feature classes and elevation rasters were produced using the information from the architectural drawings of the current buildings in the study area.

1.1.4 – Results

The results of the study were to classify each site within the study area with a scale value (SV) between 1 and 10, signifying to what extent each pixel of the raster is a safe site for construction using a GIS methodology. Results were calculated using a weighted overlay raster tool which calculates a single value output based on the relevant importance of various factors such as slope angle and distance from existing infrastructure. In order to ensure that the results complied with the appropriate building regulation requirements, the ESRI buffer algorithm was implemented to identify and remove areas which are not suitable for construction.

1.1.5 – Data and Algorithm Quality Issues

The study was formed of two general sections: the section evaluating data such as the questionnaire surveys and building regulations that were used to derive the methodology, and the section which assessed to what extent the methodology was successful with respect to a case study. The case study used data from a document illustrating an architect's drawings of the topography and current infrastructure in the study site. Thus, it is reasonable to suggest that in order to increase accuracy of the results, data from a greater variety of sources would be beneficial.

With regards to the algorithm quality, the ESRI buffer algorithm produces reliable results (section 3.6), however, it is noteworthy that the researchers do not mention which projections or coordinate systems are used. Unless this is considered, it is possible that projection or coordinate system inconsistencies may decrease the accuracy of the results of the buffer algorithm.

1.2 – REVIEW OF 'A SPATIAL MULTI-CRITERIA EVALUATION FOR SITE SELECTION OF OFFSHORE MARINE FISH FARM IN THE LIGURIAN SEA, ITALY'

1.2.1 – Summary

'A spatial multi-criteria evaluation for site selection of offshore marine fish farm in the Ligurian Sea, Italy' is an academic article published in the journal 'Ocean & Costal Management'. The research process involved:

- the identification of factors that affect the suitability of fish farming in the study site and relevant data collected;
- the modification of the data to ensure that it is fit for purpose and compatible for comparison;
- outputs are overlaid and united into a single raster output of sit suitability, and
- results are validated by comparing the identified possible sites with existing farms.

The article describes an effective method of using GIS tools and algorithms to identify potential sites for fish farms in the study site using a variety of data sources and GIS algorithms.

1.2.2 – Algorithms and Software

During the research, two GIS software packages were used: Geographic Resource Analysis Support System 7.1.0 (GRASS) and Quantum GIS 2.6.0 (QGIS) and were selected as they are both free and open-source software systems which will allow the results of the study to be more easily implemented and applied in the field. The article states that the GRASS software was used to perform the analysis and therefore the *r.buffer* algorithm was implemented on the raster and the *v.buffer* tool on the vector based datasets. QGIS was employed to combine all of the various results into a single, suitability raster map based on the weighted overlay output.

1.2.3 – Data and Data Source

The sources of data used in the reviewed research article vary from those which are clearly stated and the relative merits assessed, to those where the source is merely mentioned. For example, wave data was provided by the WAVEWATCH III model for the study site. The resolution of the data and timestamp are presented as data from 1979-2010 for the Ligurian Sea and the data is validated by the Italian Rete Ondametrica Nazionale (RON) buoy data for the same area. However, the bathymetric

data, nautical map data and Ligurian marine sediment map data sources and data format are merely stated (Dapueto et al., 2015).

1.2.4   – Results

In a similar fashion to the article reviewed in section 1.1, the relative weightings of factors used in the weighted linear combination (WLC) tool that produced the final suitability map were derived from surveys from several parties with expertise in the field. The results detected 65 suitable locations for the installation of an offshore wind farm in the study area. However, it is noted that 'local analyses should be conducted to deeply examine specific factors' before such farms should be installed (Dapueto et al., 2015).

1.2.5   – Data and Algorithm Quality Issues

The quality of data sources used, the justification of methodology and validation indicate that the accuracy and reliability of the results is excellent. However, in order to improve the results, the researchers should have stated the relative merits and justify all data sources, not just select few. Conversely, the researchers state the limitations of their work and offer suggestions on how to expand their research to implement it on other study sites, a notable difference between this article and the article reviewed in section 1.1.

2   – TERMINOLOGY DEFINITIONS

2.1 – Intersects

2.1.1   – ArcGIS

"A geometric integration of spatial datasets that preserves features or portions of features that fall within areas common to all input datasets." (ESRI, 2016a)

2.1.2   – QGIS

"Overlay layers such that output contains areas where both layers intersect." (QGIS, 2016a)

2.2 – Touches

2.2.1   – ArcGIS

"To be selected, the geometry of the source feature must fall inside the geometry of the target feature including its boundaries. For example, a polygon representing the United States contains the state of Texas and is selected even though they share common boundaries along their southern borders." (ESRI, 2016b)

2.2.2   – QGIS

"Test for if geometry touch another (uses GEOS)" (QGIS, 2016b)

3 – GIS COMPARISON

3.1 – Introduction (200 words)

In the following section, algorithms that identify points contained within polygons and produce a buffer around a feature will be analysed in both the ESRI ArcMap 10.4 software package and the Quantum GIS package 2.18.1. Although the two algorithms perform different functions, in reality, they can be used in conjunction with each other for the GIS operation of calculating if an object is within a predefined distance of another object (Longley et al., 2015). That is, for example, if a utility company wishes to detect which households are within a threshold distance of the current utility infrastructure network; they could calculate a buffer around the infrastructure network, and identify households (points) which are located within the buffer (polygon). Such processes are used regularly in site selection GIS studies (Cheng & Thompson, 2016, Dapueto et al., 2015, Kumar & Bansal, 2016, Noorollahi et al., 2016).

Longley et al. (2015) state that the production of a buffer is '*one of the most important GI system operations available to the user*'. Thus, it is imperative for a GI scientist to be equally aware of the relative merits of the algorithms that are employed, as their limitations. The following sections will review to what extent the algorithms are successful in both British National Grid OSGB36 and WGS84 and their limitations reviewed.

3.2 – Background (200 words)

Huang and Shih (1996) critically review eight possible point-in-polygon algorithms: grid method, ray intersection method, sum of angles method, swath method, sign of offset method, sums of area method, orientation method and the wedge method. They conclude that the grid method is the most effective and computationally efficient for identifying points within a raster polygon data file. Additionally, the swath method is appropriate for point-in-polygon queries that involve more than one point as it is less computationally expensive.

The drop line algorithm (or semi-line algorithm) as a method of detecting if a point in contained within a polygon is discussed by Worboys and Duckham (2004) and De Smith et al. (2006). Potential failures arise when the generated line passes directly over a line or a node of the polygon. Such an issue can be eliminated by allocating a weight value to the point, if an exception is reported (De Smith et al. 2006).

The buffer algorithm takes two parameters: the distance that the buffer must create around the objects, and the geometry of the objects themselves. However, buffers are not limited to the exterior to the objects, but can be used to isolate an area of an object where all identified locations remain more than a pre-specified distance away from the exterior of the original shape (Albrecht, 2007).

3.3 – Data (200 words)

The data evaluated in this study was produced, for the purposes of this study, using the 'add feature' tool of Quantum GIS 2.18.1 (QGIS) in the WGS84 coordinate reference system. The datasets produced were designed to test the reliability of the point-in-polygon (PIP) and buffering algorithms of the two GIS software packages. Thus, unusual shapes and complex polygons were formed and points located in areas designed to investigate how robust the algorithms are (Figures 1–8).

3.4 – Method (400 words)

After conducting research and a review into literature regarding the relative merits and limitations of the PIP and buffering algorithms, a shape file was produced in QGIS using the WGS84 coordinate reference system, designed to test the durability and reliability of the two algorithms. Following this step, the shape file was exported from QGIS and imported into the ESRI ArcMap 10.4 software where more complex polygons and points were created, also in the WGS84. This step assesses whether the software package used to design the polygon data affects the reliability of the algorithms implemented later.

Subsequently, the shape file was resaved and the ESRI '*Spatial join*' algorithm was executed to identify the points that the algorithm attribute to the polygons based on their location. After the results had been collected, the coordinate reference system of the layers were changed to the coordinate system was also changed to British National Grid, OSGB36 (BNG) using the '*Define projection*' tool located in Data Management tools, Projections and Transformations toolbox. Then, the same '*Spatial join*' algorithm was run and the results also collected.

In order to test to what extent the QGIS software compared with its ESRI equivalent, the coordinate reference system of the QGIS file was also changed to British National Grid OSGB36 and the analysis tool '*Points in Polygon*' algorithm was re-run and the results collected and visualised in section 3.5.

Regarding the buffer algorithm, the process was repeated with the same polygon and point shape files and the results also collected and illustrated in section 3.5.

3.5 – Results (250 words)

The results of the study have been visualised in figures 1,2,3 & 4, below. Figure 1 and 3 illustrate the result of the QGIS '*Points in Polygon*' query, executed in the WGS84 and British National Grid (OSGB36) coordinate reference system, respectively. Each polygon has been labelled with a number, corresponding to the id column of the Attributes table also included in Figure 1.
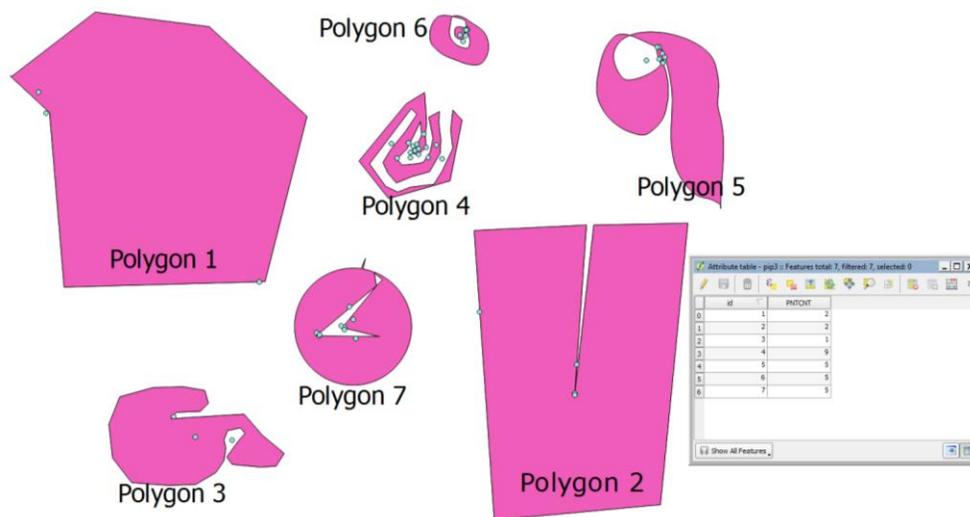

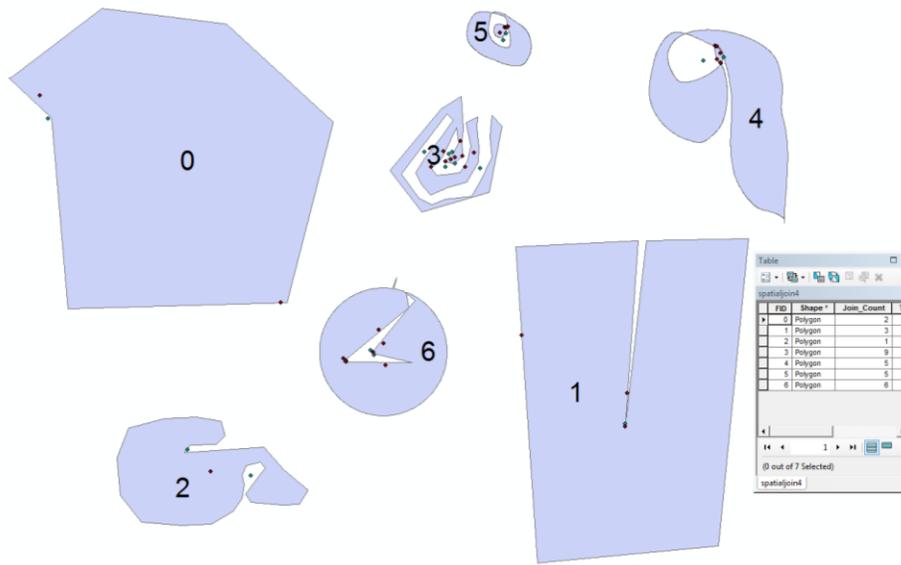*Figure 1 - QGIS 'Point in polygon' results in WGS84*

Figure 2 - ArcMap 'Spatial join' results in WGS84

In contrast, figures 2 & 4, demonstrate the results of the ESRI ArcMap '*Spatial join*' algorithm in WGS84 and British National Grid, respectively. A summary of the results can be found in the corresponding Attributes tables which have been included in the figures. Each polygon is assigned a number which has been superimposed onto the polygon and matches the value in the FID column of the attribute table. A summary of the attribute tables can be found in table 1.
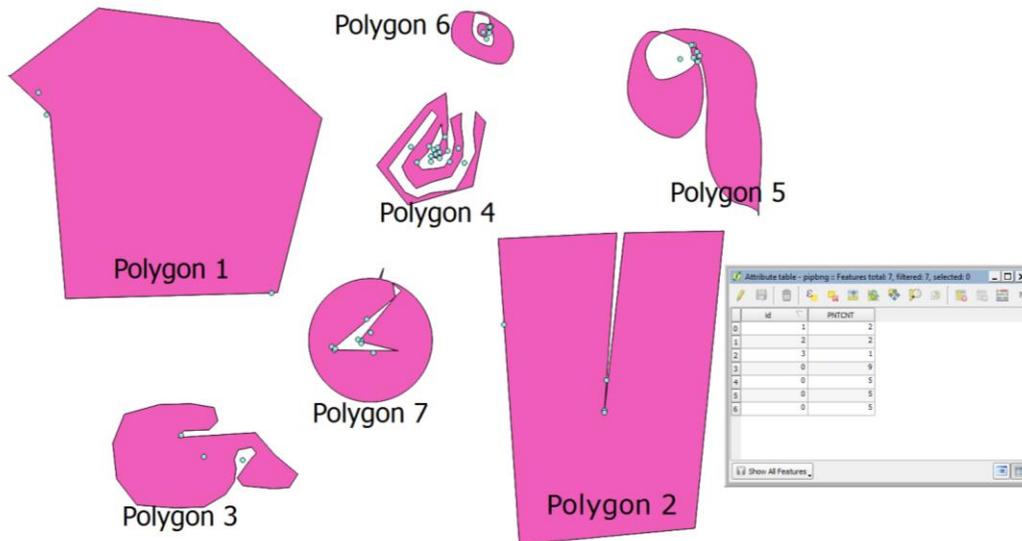

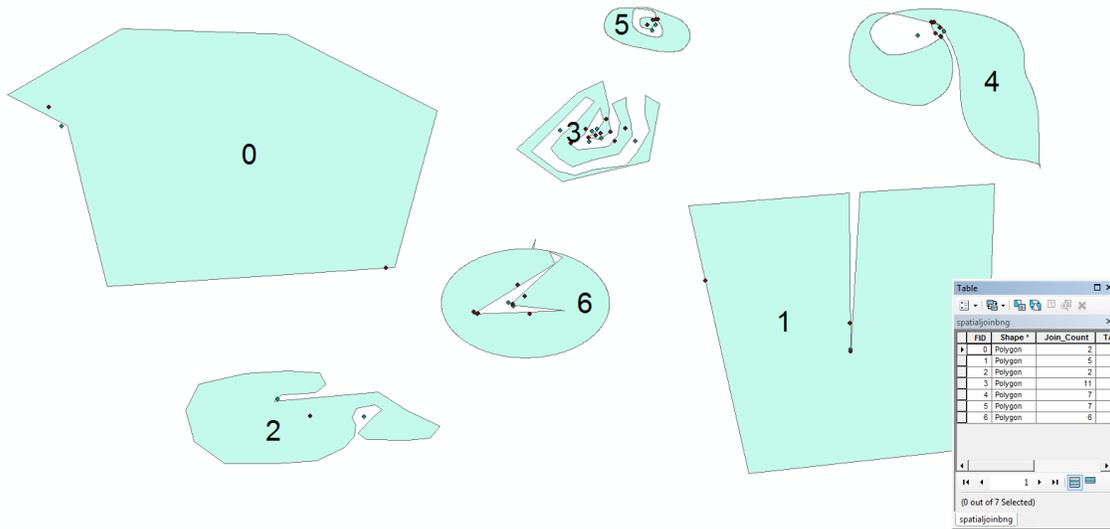*Figure 3 - QGIS 'Point in polygon' results in BNG OSGB36*

Figure 4 - ArcMap 'Spatial Join' results in BNG OSGB36

	QGIS WGS84	ArcMap WGS84	QGIS BNG	ArcMap BNG
0 or Polygon 1	2	2	2	2
1 or Polygon 2	2	3	2	5
2 or Polygon 3	1	1	1	2
3 or Polygon 4	9	9	9	11
4 or Polygon 5	5	5	5	7
5 or Polygon 6	5	5	5	7
6 or Polygon 7	5	6	5	6

Table 1 - Summary of the Point-in Polygon and Spatial Join algorithms

The results of the buffering algorithm are available in figures 5, 6, 7 and 8. Figure 5 and 6 visualise the results of the buffering algorithm in WGS84 coordinate projection system in QGIS and ArcMap respectively. The buffer distance was selected as 0.01 metres to allow a both an exterior and interior buffer. Similarly, figure 7 and 8 display the outputs of the same buffer algorithms using the BNG OSGB36 projection system, also in QGIS and ArcMap, respectively.
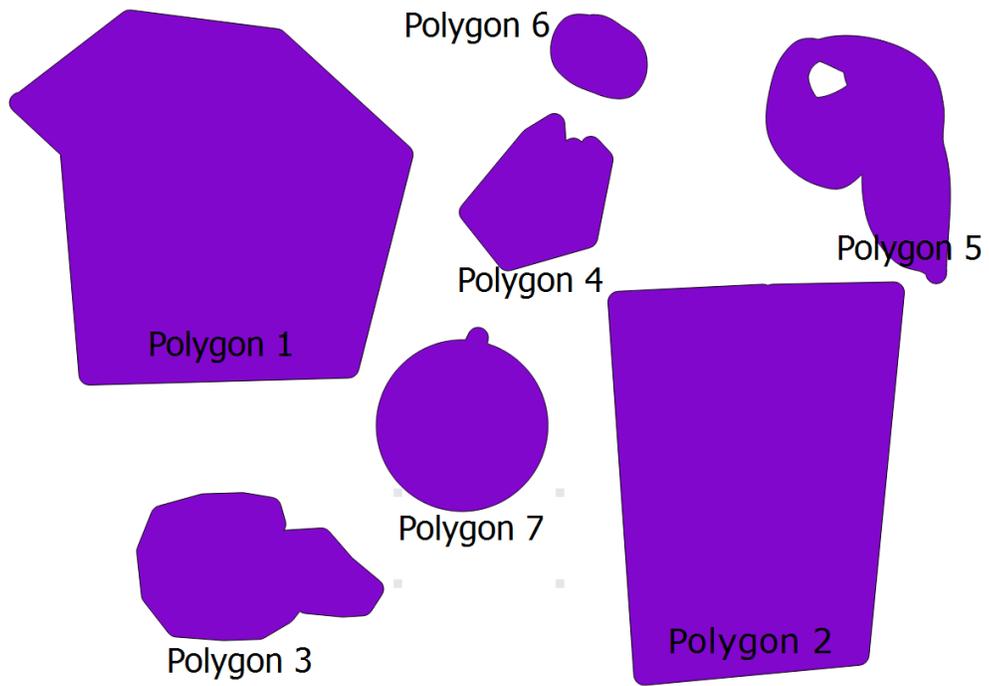
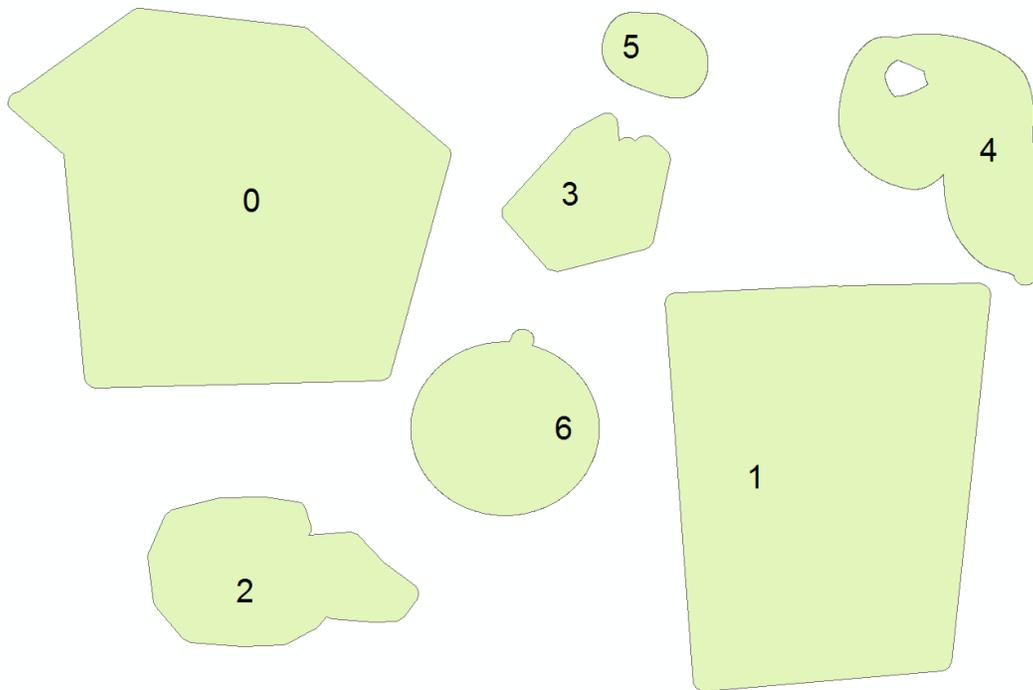*Figure 5 - QGIS Buffering results in WGS84*
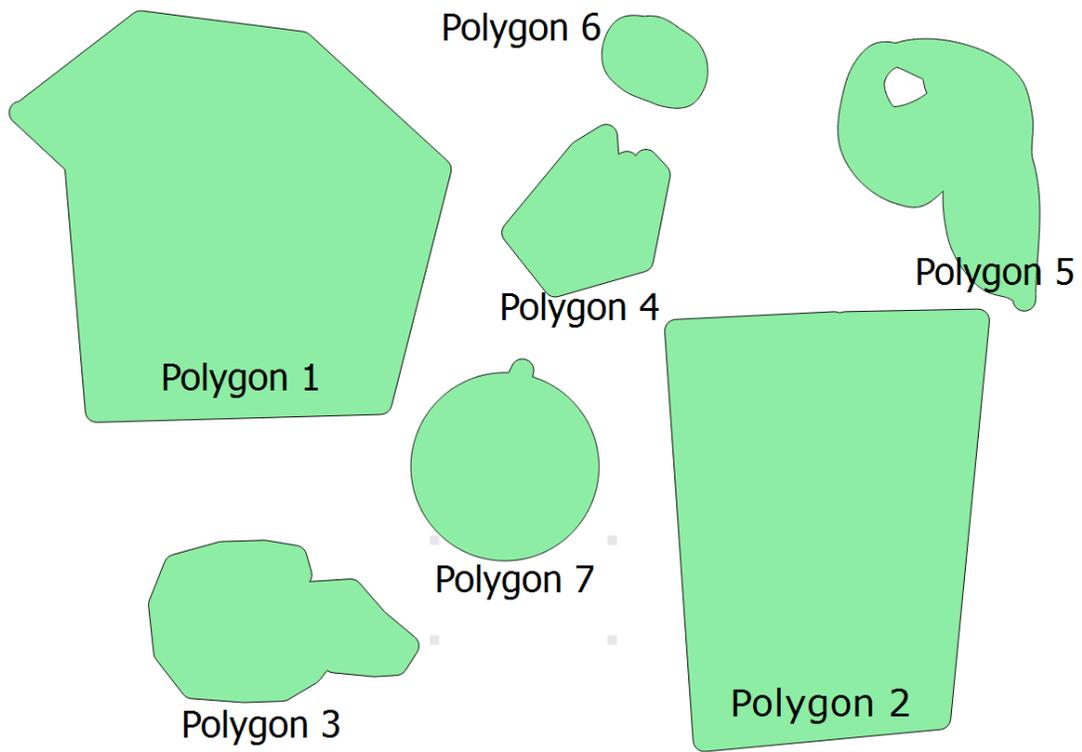

Figure 6 - ArcMap Buffering results in WGS84

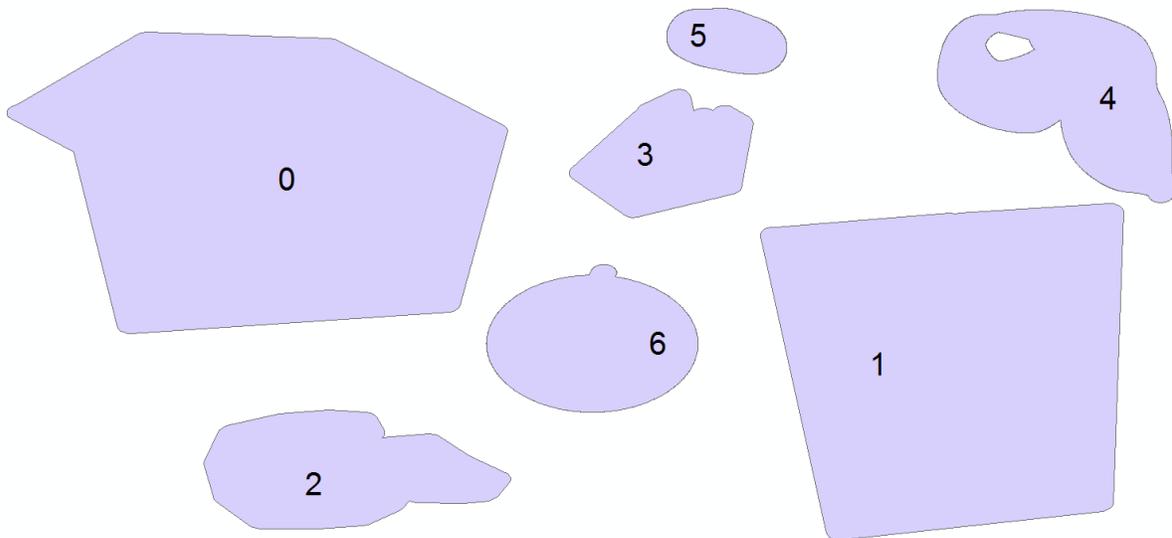*Figure 7 QGIS Buffering results in BNG OSGB36*

Figure 8 - ArcMap Buffering results in BNG OSGB

3.6 – Discussion (500 words)

The results of the study indicate that the output of a point-in-polygon algorithm can vary depending on software used and the coordinate reference system on which the data is based. In Figure 1 & 3, two point-in-polygon algorithms were executed on the same polygon and point shape files in Quantum GIS, changing only the coordinate reference system of the shape files. The resulting values were identical for every polygon, implying that the point-in-polygon algorithm used by the QGIS software package is reliable, even when the coordinate reference systems are altered. However, this could also imply unreliability as Longley et al. (2015) state:

"...others [algorithms] ignore this, on the grounds that it is never possible, in practice, to determine location with perfect accuracy..."

Thus, it could be argued that if a computer program returns the same results, it could imply that assumptions have been made in order to run the algorithm more efficiently or to overcome complicating factors. For example, the ray intersection/drop line method of detecting if a point is in a polygon has a number of possible complicating factors. One such factor could occur if the point is located directly below a vertex of the polygon as the algorithm registers the crossing of a vertex, but in reality, it has not (Longley et al. 2015, Huang & Shih, 1996). Moreover, some of the points created were located directly on the boundary of the polygons. It would be important, therefore, for any person applying the algorithm for analysis to consider the possible assumptions that the software has made, and if they affect the desired results.

In contrast, the ESRI ArcMap software produced a greater variety of results, ranging from identical, to different in three points when the projection was changed from WSG84 (Fig. 2) to BNG (Fig. 4). Further to the point made in the previous paragraph, it is possible that the variance could infer unreliability. However, it could also suggest much greater flexibility in the algorithms which could be executed and allow the user to customise the results which are produced.

The impact of changing the coordinate projection system was profound in ArcMap (5 of the 7 results changed), and non-existent in the QGIS software. The reason for this could be the algorithms which were employed; in QGIS the coordinate reference system was changed but the shape files appeared to be re-labelled rather than re-projected. However, upon employing the '*Define projection*' in ArcMap, the software appeared to re-project the shape files rather than simply change the metadata. Therefore, it is crucial for any GIS professional that is required to change the coordinate reference system of a dataset, that they confirm that the software has re-projected the data and not simply re-labelled it.

To support this claim, one should note the difference in shapes produced by the buffer algorithms in figure 8. The buffer produced in figure 8 is distorted in a variety of axes, particularly profoundly in polygons 0 and 1. Such a distortion substantiates the hypothesis that the reason for the difference in the point-in polygon results is due to the differences in the method that the software deals with a change in coordinate reference system. The distortion in figure 1 occurs as the ArcMap software has re-projected the shape file to the secondary coordinate reference system (OSGB36) and not simply adjusted the meta-data.

3.7 – Conclusion and further work (250 words)

In conclusion, both software packages were able to accurately execute the two algorithms reviewed in the study. However, a number of points were raised and should be considered by anyone that employ either algorithm:

- different GIS software packages can produce different results even when using the same data due to different algorithms, and
- it is crucial for a GIS user to ensure that the software has re-projected into the desired coordinate reference system to ensure accuracy of results.

Further work could include studies into to what extent the distortion affects point in polygon data in a greater variety of coordinate reference systems such as those for other countries and continents and using a greater variety of polygons point datasets. It is likely that higher resolution and more complex data would allow a greater understanding of the relative merits and limitations of the algorithms offered by the GIS software packages.

4 – REFERENCE LIST

Albrecht, J. (2007) Key concepts and techniques in GIS. London, United Kingdom: Sage Publications

Cheng, C. and Thompson, R. (2016) 'Application of boolean logic and GIS for determining suitable locations for temporary disaster waste management sites'. *International Journal of Disaster Risk Reduction* 20, 78–92

Dapueto, G., Massa, F., Costa, S., Cimoli, L., Olivari, E., Chiantore, M., Federici, B., and Povero, P. (2015) 'A spatial multi-criteria evaluation for site selection of offshore marine fish farm in the Ligurian sea, Italy'. *Ocean & Coastal Management* 116, 64–77

Environmental Systems Research Institute (2016a) Using select by Location—Help [online] available from <http://desktop.arcgis.com/en/arcmap/latest/map/working-with-layers/using-select-by-location.htm> [12 December 2016]

Environmental Systems Research Institute (n.d.) Definition - Esri support GIS dictionary [online] available from <http://support.esri.com/sitecore/content/support/Home/other-resources/gis-dictionary/term/intersect> [12 December 2016]

Huang, C. and Shih, T. (1996) 'On the complexity of point-in-polygon algorithms'. Computers & Geosciences 23 (1), 109–118

Kumar, S. and Bansal, V. K. (2016) 'A GIS-based methodology for safe site selection of a building in a hilly region'. Frontiers of Architectural Research 5 (1), 39–51

Longley, P., Goodchild, M., and Maguire, D. (2015) Geographic information systems and science. 4th edn. New York: John Wiley & Sons

Noorollahi, Y., Yousefi, H., and Mohammadi, M. (2016) 'Multi-criteria decision support system for wind farm site selection using GIS'. Sustainable Energy Technologies and Assessments 13, 38–50

QGIS (2016a) QGIS API Documentation [online] available from
<https://qgis.org/api/classQgsGeometry.html#a3a929249c1cc2c4cea95e81f92c6ee7f> [12 December
2016]

QGIS (2016b) FTools Plugin [online] available from
<https://docs.qgis.org/2.8/en/docs/user_manual/plugins/plugins_ftools.html> [12 December 2016]

De Smith, M., Goodchild, M., Longley, P., and Smith, M. (2006) Geospatial analysis: A
comprehensive guide to principles, techniques and software tools. Leicester: Troubador Publisher

Worboys, M. and Duckham, M. (2004) GIS: A computing perspective. 2nd edn. Boca Raton, FL:
Taylor & Francis