Representation, Structures and Algorithms - GEOGG126

Part-A Literature Review

A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales

- **Summary**

Luo, Taylor and Parker (2008) reviewed the suitability of seven different spatial interpolation algorithms when estimating daily mean wind speed surfaces. The main paper assesses which of the spatial interpolation methods presented the most accurately representative wind speed data the study area. This assessment was achieved by comparing the estimated data provided through the varying interpolation methods, with data collected independently. The study aims to contribute to the investigation of daily coverage of pathogens inducing crop diseases.

The study results from the IDW algorithm produced less desired results than many of the other spatial interpolation techniques tested. The results suggested the Cokriging method produced the most accurate estimation of continuous surface models for mean daily wind speed. However, the paper states it's limitations within this research project and that it is not feasible to assert one method superior to another unless further testing is undergone.

- **Algorithm and Software**

This review focuses on the application and analysis performed using the Inverse Distance Weighting (IDW) interpolation technique. This method uses a distance-decay approach where each data point has local influences that weaken relative to distance.

All the spatial interpolation algorithms throughout the paper were executed using the geostatistical analyst extension within the ArcGIS software platform.

– Data and Data Source

Luo, Taylor and Parker (2008) used raster wind speed measurements provided by the UK Meteorological Office (UKMO) from 189 wind stations located throughout the U.K. over the period of 1st January 1998 – 31st October 2002.

– Results

The results obtained from the IDW method were consistent throughout the data, however, had a relative poor performance for the daily mean wind speeds in comparison to other spatial interpolation methods tested. This was particularly evident in areas of unevenly distributed or sparse data, where the paper states the IDW technique produced inappropriate results due to isolated points exerting a weighted influence that weakens uniformly.

Additionally, the IDW interpolation technique produced a rougher surface model as a consequence of high variability in the data. This lead to areas with steep gradients which were believed to be unrealistic and as a result of poor surrounding station coverage. This is because the IDW method appoints similar weights for clustered points as well as sparsely located points, leading to less accurate predicted values.

The IDW method also performed poorly when comparing the results of the Mean Error calculations, showing IDW to be an algorithm which can be likely to produce more biased estimations.

– Data and Algorithm Quality Issues

The authors had to consider the parameters values that ArcGIS requires when performing the IDW algorithm, these are as follows:

- Search shape – Circle.
- Search radius – 100km.
- Number of points – minimum and maximum number of 10 and 15 respectively.
- Power – the power was automatically generated by ArcGIS

All of these parameter values must be considered carefully as they have significant impacts on the output of the algorithm. Additionally, the authors may have considered the size of their dataset as it has been recognised that the performance of a relative complex algorithm such as IDW can become very computational intensive (Mei, Xu, Xu, 2017).

Spatial characterisation, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina

– Summary

Woolard and Colby's (2002) study of coastal dune systems had the primary aim to provide supplementation and advancement of methods used to illustrate, and consequently further, comprehend coastal dune systems through the use of airborne LIDAR data. The authors were able to perform this analysis by generating digital elevation models (DEM) from LIDAR data and spatial statistic techniques.

A raster-based geographic information system (GIS) was used to generate two study site DEMs, located in Cape Hatteras National Seashore, North Carolina. The LIDAR data was interpolated to resolutions ranging from 1 x 1 to 20 x 20m, these were used to measure variances in dune volumetrics over a 1-year period (Fall 1996-1997).

The study presented 1-2m resolution provided the volumetric changes with the greatest accuracy, additionally the most accurate representation of Cape Hatteras coastal dunes. Despite results derived at other sites and varying spatial extents may fluctuate, the methods presented are functional to other study sites in the interest of deducing optimal resolutions for analyzing costal system topography using GIS.

– Algorithm and Software

The IDW (Inverse Distance Weighted) interpolation algorithm, administered on the Spatial Analyst extension within ArcView 3.0. Woolard and Colby (2002) chose ArcView 3.0, predominantly advocating ArcView's instated graphical user interface (GUI); furthermore ArcView contains designated algorithms conducted to compute directional statistics on behalf of interpreting surface complexity (Hodgson and Gaile, 1999).

– Data and Data Source

The airborne LIDAR data was acquired from flights during September and October 1996 and 1997, performed by the NOAA - CSC in affiliation with NASA, and U.S.

Geological Survey. (NOAA CSC, 1997)

– Results

The raw 1996 and 1997 airborne LIDAR data were interpolated to 1 x 1, 2 x 2, 5 x 5, 10 x 10, 15 x 15, and 20 x 20 m spatial resolutions, additionally reassigning the floating-point grid coverage's as binary files. Furthermore, each resolution of the 1996 DEM was used to generate the two research sites (100 x 200m). The DEM's illustrated that spatial resolution had a direct correlation with the volume of deposition recorded measured at both sites.

1 and 2m resolutions provided the highest consistency in measurements of volumetric change. 5m resolutions stipulated satisfactory approximation. Whereas, less reliable results were obtained from 10m and larger resolution data.

– Data and Algorithm Quality Issues

Two major considerations were to firstly use methods such as the IDW interpolation algorithm, which has functional application when processing the LIDAR data. Secondly, was using software platforms and instalments, which are easily attainable for studies within all levels of analysis.

One influential issue was found within the 1997 data, which was not present in the 1996 dataset. During 1997, the ATM II was employed in place of the ATM I. This resulted in a notable increase in density of points collected over the same spatial coverage in 1997 (post ATM II employment) than in 1996 (pre ATM II employment). As the hardware available did not allow processing of this massed dataset, in order to use the interpolation algorithm, a short C program using the Unix platform reduced the 1997-point density dataset until comparable to the 1996 dataset.

Reference List:

Hodgson, M. E. , Gaile, G. (1999): A cartographic modeling approach for surface orientation-related applications. *Photogramm. Eng. Rem. Sens.* **65**, 85–95.

Luo, W., Taylor, M. and Parker, S. (2008). A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. *International Journal of Climatology*, 28(7), pp.947-959.

Mei, G., Xu, L. and Xu, N. (2017). Accelerating adaptive inverse distance weighting interpolation algorithm on a graphics processing unit. *Royal Society Open Science*, 4(9), p.170436.

NOAA Coastal Services Center (CSC), 1997. Airborne LIDAR assessment of coastal erosion (ALACE) project. National Oceanic and Atmospheric Administration Coastal Services Center Report, Charleston, SC. http://www.csc.noaa.gov/crs/tcm/index.html.

Woolard, J. and Colby, J. (2002). Spatial characterization, resolution, and volumetric change of coastal dunes using airborne LIDAR: Cape Hatteras, North Carolina. *Geomorphology*, 48(1-3), pp.269-287.

Part-B GIS Comparison

Introduction

Spatial analysis if often used as an incredibly important form of analysis for geographic information, and allows the extraction of new and significant geographic data production. The use of spatial analyse on datasets with geospatial characteristics allows users to perform informed investigation in to many different facets of research. Geographic Information Systems (GIS) allow the spatial analyse to be performed in an accurate and efficient way, with a multitude of potential research capabilities depending on the user's requirements. Spatial analysis can be performed on GIS platforms such as ArcGIS and Q-GIS by using geospatial algorithms to perform the computation.

The aim of this study is to perform an assessment on the capabilities of geospatial algorithms when tested under varying research parameters, in particular to the competence of both the 'Buffer' and 'Spatial Join' algorithms ability when using vector datasets. The algorithms will be tested on their abilities to perform with varying size datasets, within different GIS platforms, and when performed in different coordinate reference systems (CRS). The two CRS used within this study are the Ordnance Survey Great Britain (BNG) and the World Geodetic System (WGS-84).

This assessment was undertaken by formulating and consequently answering a set of research parameters:

- How can different projections affect a geospatial algorithms?
- What impact does varying sizes of datasets have on the performance of geospatial algorithms?
- How can using a different GIS software platform affects these results?

Background

The Buffer tool is a spatial analysis Geoprocessing tool. Buffering will generally generate two areas in spatial relation to target features, separating an area containing the target features (buffer zone), and another area excluding the features. A buffer zone is an area encompassing a specified isolated object or set of objects, constructed outward from these object features to a specified distance, also known as a distance threshold (Corral and Almendros Jimenenz, 2007; Zhang, Shi and Meng, 2005). A buffer zone is exclusively represented as vector polygons.

In GIS, overlaying features is a fundamental approach to spatial analysis between map layers. A widely used form of overlay is joining multiple sets of polygonal (vector) data with spatial attributes through the spatial join method to produce a new joined output vector dataset. A

Spatial Join algorithm involves the process of connecting all objects satisfying a spatial relationship based on the multi-dimensional attributes, then joining feature classes based on this spatial relationship (Puri, 2015). A spatial join is characteristically comprised of two stages: the filter and refinement stages. Initially the filter stage provides a conservative approximation of complex objects by minimum bounding rectangles to remove objects that do not apply to the consideration. The refinement stage then filters further to eliminate any erroneous results permitting to the spatial join condition. (Puri, 2015; Jacox and Samet, 2007; Becker et al., 1999)

BNG National Grid coordinate system is a form of Projected Coordinate System (PCS) with a two-dimensional Transverse Mercator projection defined within the British National Grid, using Easting and Northings coordinates given in 'flat' units such as metres. This Coordinate Reference System (CRS) consists of the Airy ellipsoid which is closely adjacent to the GB geoid. The coordinate system features grid lines which are parallel, and the origin coming from Greenwich, U.K. The most widely used form of a Geographic Coordinate System (GCS), is the WGS-84 datum, consisting of a three-dimensional Cartesian coordinate system with coupled ellipsoid, to project a three-dimensional curvature projected grid (Graticule) on to the globes geoid. This lets geographic positions be determined by X, Y, Z Cartesian coordinates, in the form of Latitude, Longitude, and Ellipsoid Height coordinates respectively. (A guide to coordinate systems in Great Britain, 2015; Sccssurvey.co.uk, 2011)

Data

The data used to carry out this project was sourced from the Ordnance Survey OpenData platform allowing use through the Open Government License (OGL). Using this platform, both the OS Code-Point data and the OS Boundary-Line vector datasets were obtained and used to carry out this analysis of this project.

Method

As previously mentioned, research questions were derived as a foundation of analysis from which to carry out the data research methods of this study. These study parameters were chosen in order to assess the three most common form of distortion from map projects; Angular conformity, Distance, and Area. In order to conduct a fair and accurate assessment of the test parameters, when each algorithm was performed, it was performed with minimal other datasets currently 'active' within the dataset, in order to maintain continuity between different algorithm performances. In order to measure the performance of each algorithm and assess the effect of varying sizes of datasets on an algorithm, time values were recorded for the time taken to perform each algorithm using the 'results' tool in the Geoprocessing extension. Additionally, in order to assess how different size dataset's may affect the algorithm performances, three datasets of varying size where tested in the study location using polygons created from OS Boundary-Line datasets: Ashford Rural West, Kent, Great Britain (figure 1).

Figure 1: Study Area, projected with WGS84 in ArcMap.

Figure 1a: 3 different map extents, overlaid to the ArcMap World Ocean reference basemap, projected in WGS84 coordinate system.

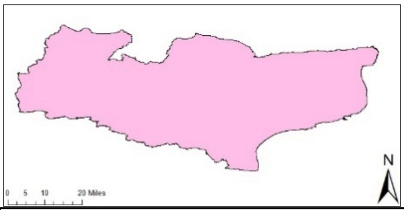

Figure 1b: Kent

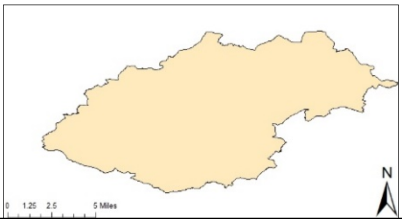

Figure 1c: Ashford Rural West

Initially, in order to assess the affect map projections can have on these research questions as outlined in the introduction. Firstly, using ArcMap all the following methods were performed by setting the dataframe to the BNG coordinate system, the OS OpenData vector data naturally comes with this coordinate reference system therefore no transformations were initially necessary. However, in order to then test the WGS36 coordinate reference system, this CRS was then set to a new dataframe and the original data sets were then transformed using the 'Project tool' extension in ArcMap to set their respective different CRS to WGS36.

Figure 2: Map showing a 5km buffer surrounding the two Railway Station points within Ash Rural West, in ArcMap. The basemap is the ArcMap World Topographic Map.

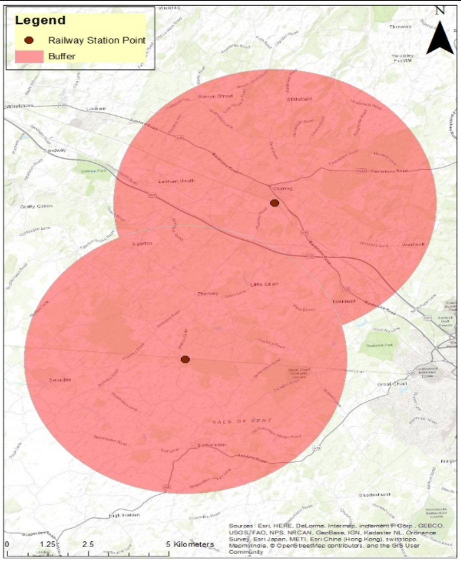

Figure 3: The figure shows the process of Spatial Join, using the Railway Stations in the Kent area,

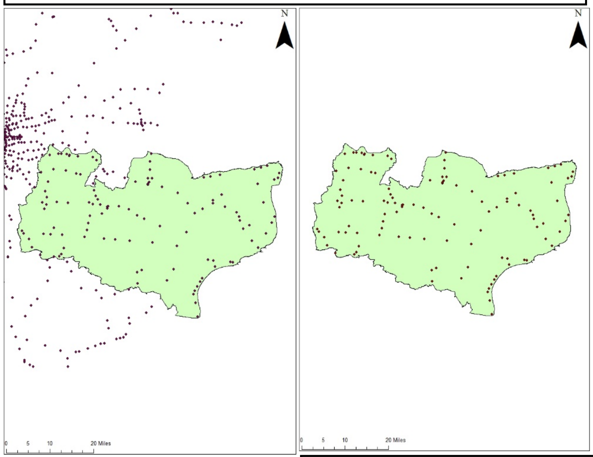

Figure 2a: Shows total GB Railway Stations distribution with Kent area.

Figure 2b: Shows GB Railway Stations post Spatial Join algorithm has excluded stations that do not lie within the

To measure the effects of projections have on angular conformity and area, the Buffer tool algorithm was used in the ArcMap platform using the Proximity toolset within the Analysis tools extension. Three different parameters of both buffer extents and size of datasets were tested. Using the 'clip' tool, the GB Railway Station dataset was clipped to polygon shapefiles of each chosen study extent (Figure 2). It was established to determine the range of capabilities for geospatial analysis uses for the buffer algorithm, tests at three different buffer extents were performed; 500m, 5km, and 100km. The Buffer algorithm provided buffered areas around the Railway Stations point dataset at varying level of extents and over numerous dataset sizes. Once the buffers had been produced, angular conformity could be visually assessed by overlaying paired buffers from each projection form.

To assess how distance can be affected from different projection CRS's, the Spatial Join tool within ArcMap's Overlay extension within the Analysis toolset. The GB Railway Station point dataset was selected to 'intersect' each varying study area extent. This outputted a polygon feature layer for each tested extent, spatially joined with the corresponding railway station which lay within this area (Figure 3). From here the total number of railway station within each extent could be calculated.

These methods were then repeated within the Q-GIS platform. The only modifications were when using Q-GIS, buffer algorithms were undergone using the 'Fixed distance buffer' tool from the vector geometry toolset, and the spatial join was performed using the 'Point in Polygon' tool, both within the QGIS geo-algorithms extension. Another difference lay when measuring the time taken to perform an algorithm, whereas ArcMap provide this information, QGIS does not have this extension already installed within the system.

Results

Tables 1: Represent the data collected from performing the algorithms in ArcMap.

Distance (metres)	500m	5km	100km
GB	2.71	2.66	2.69
Kent	1.65	1.53	1.66
Ashford Rural West	1.41	1.86	1.45

Table 1a: Time taken to perform Buffer algorithm with varying levels of buffer extents at the three different study locations. Measurements taken in using OSGB-36 in ArcMap.

	Railway Station Count	Time taken for algorithm
GB	3352	13 minutes, 53 seconds
Kent	116	1.58 seconds
Ashford West Rural	2	1.69 seconds

Table 1b: Summary of Railway Station count and time taken to perform Spatial Join algorithm at the three different study locations. Measurements taken in OSGB-36 in ArcMap.

Distance (Metres)	500m	5km	100km
GB	6.39	6.53	7.72
Kent	0.7	0.71	0.71
Ashford Rural West	0.46	0.48	0.55

Table 1c: Time taken to perform Buffer algorithm with varying levels of buffer extents at the three different study locations. Measurements taken in using WGS-84 in ArcMap.

	Railway Station Count	Time taken for algorithm
GB	3346	14 minutes, 6 secs
Kent	116	0.84secs
Ashford West Rural	2	0.9secs

Table 1d: Summary of Railway Station count and time taken to perform Spatial Join algorithm at the three different study locations. Measurements taken in WGS-84 in ArcMap.

Results from tables 1 show how both the Buffer and Spatial Join algorithm performed when tested with ArcMap. It can be seen that the significant result from the Spatial Join results (Tables 1b and 1d) is the reduction in Railway Station count by 6 in the WGS-84 projection when using the GB dataset but remained same for other extents (Figure 4). The WGS-84 projection also took less time when performing for Kent and AWR extents but similar values seen for GB between projections. Tables 1 show the Buffer algorithm had no impact from varying size of extent, whereas the dataset sized proved to cause a great impact on the time taken to perform the algorithm. This is also apparent in the Spatial Join measurements which see huge increase in time taken. Results tables for Q-GIS are not also presented due to two factors: firstly, I was unable to measure the time taken for each algorithm performance; secondly, the different projections did not affect the Spatial Join algorithm within Q-GIs, with all points remaining within the polygon, therefore additional maps representing this was not deemed necessary.

Figure 4: This figure represents the Railway Stations not included within the Spatial Join performed on the WSG projection in ArcMap.

Figure 4a: The map shows the total Railway Stations used within the study.

Figure 4b: The map shows the Railway Stations that were not included in the WSG Spatial Join

Figure 5: Map showing comparison of conformity and buffered centroid differences between the Ashford Rural West buffers for both the BNG and WGS projections in the WGS-84 CRS, in ArcMap. All maps contain the World Topographic basemap from ArcMap.

Figure 5a: Map from ArcMap on BNG dataframe, showing the difference in BNG and WSG projections for 500m buffered Railway Stations in Ashford Rural West.

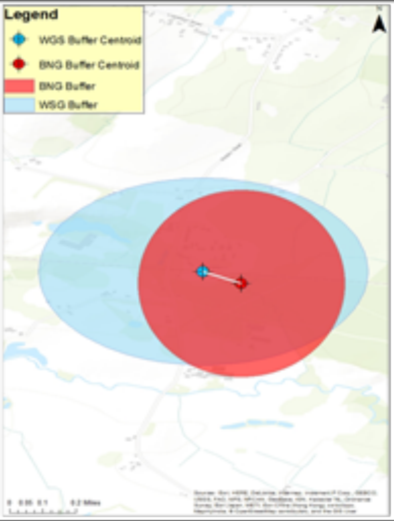

Figure 5b: Map showing measured distance between calculated centroid points for Ashford Rural West's 500m buffers for BNG and WGS projections.

Figure 5c: Map showing different in map conformity using 100km buffered Railway Station in Ashford Rural West.

Figure 6: Map showing comparisons of buffer conformity between Ashford Rural West railway station buffers for both the BNG and WGS84 projections in the WGS-84 CRS, in Q-GIS. Both maps contain the 'Google Physical' basemap from the OpenLayers plugin.

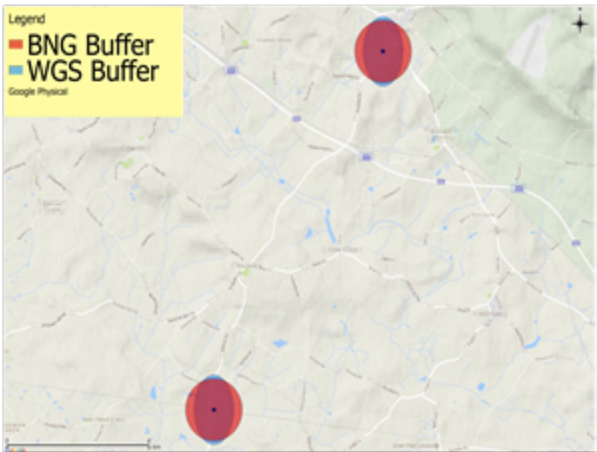

Figure 6a: Shows the difference in BNG and WGS projections for 500m buffered Railway stations in Ashford Rural West.

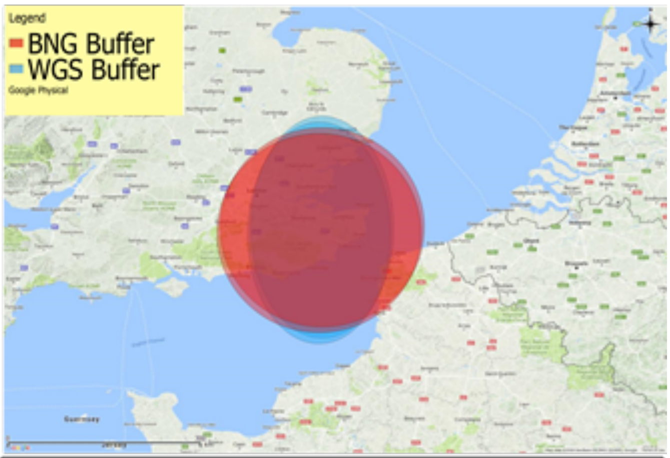

Figure 6b: Shows the difference in BNG and WGS projections for 100km buffered Railway stations in Ashford Rural West.

Discussion

The timed results from the ArcMap platforms analysis of algorithms test provided great insight in to the performance capabilities of each algorithm, particularly when the dataset size increased. This can be seen in Tables 1, where all GB dataset tests took a longer time to complete the algorithm than that from Kent or Ashford Rural West (AWR). The time delay of algorithm computation was most evident when performing the Spatial Join tests (Tables 1b and 1d), showing a very significant increase in computation time. These results agree with Puri (2015) that stated large datasets often will take longer to perform computational analysis. It can also be noted that excluding the time taken to perform GB Spatial joins, Table 1 suggests that the projection type may also have an effect on computation within ArcMap. The tests performed using the WGS-84 CRS present faster computation times than BNG tests. Significantly, the Spatial Join algorithm also suggested that greater size datasets are directed correlated with projection CRS's effect on the spatial distribution of point data. Figure 4, shows how WGS-84 calculated less Railway Station points within the GB polygon.
Despite this, the most important difference in results produced by the changing projections was the effect on angular conformity of the buffer polygons. Figure's 5 and 6 show a graphic inference into how different coordinate projections affect conformity and polygon centroids. ArcMap (Figure 5) shows the WGS CRS projected buffer displayed a less accurate representative shape – stretching horizontally - than the BNG CRS projection, which maintained the desired spherical shape and conformity. Similarly, the Q-GIS results also displayed the BNG projected buffer to retain desired conformity, however WGS projected buffers now show a distorted and stretched conformity vertically. These findings may be inferred to have correlation with the measurement of polygon centroids, which allowed the analysis of how the polygons were being transformed. Where ArcMap had a calculated centroid difference distance of 132m (Figure 5b), Q-GIS did not show any alterations in polygon centroids.

Conclusion and further work

It can be concluded the different projections within ArcMap greatly affect the buffer algorithm in respect to angular conformity of shape and distance, represented in this study by the buffer polygons, point-in-polygon's and buffered centroids. Conversely, the results (Figures 4, 5 and 6) show that Q-GIS displayed far less significant variances between CRS projections. Consequently, despite not calculating computation time in Q-GIS, it can be suggested that less correlated results between different projection systems and computation time would have been obtained.

The larger datasets had a direct correlation with the computation time of the algorithms. Therefore, future studies can continue this work produced by using larger datasets, which will consequently take more processing power in order to complete the algorithm. Similarly, testing the use of higher quality resolution data will take more computation ability and therefore may have a profound impact on the algorithm's ability. An additionally area of the further study could be to investigate the effects of different forms of data, raster and vector, point and polygon, and how this affects different algorithms.

The results from this study present clear correlations between many of the research parameters investigated. From this, it can be inferred that it is essential when performing GIS analysis, to have a clear and detailed perspective of precisely what geographic information the user wishes to use, the characteristics of this information, and the most appropriate way to display this information depending on the research purpose.

Reference List

An introduction to mapping coordinate systems and the use of GPS datasets with Ordnance Survey mapping. (2015). *A guide to coordinate systems in Great Britain*, [online] Ordnance Survey. Available at: http://www.bnhs.co.uk/focuson/grabagridref/html/OSGB.pdf [Accessed 6 Jan. 2018].

CORRAL, A. and ALMENDROSJIMENEZ, J. (2007). A performance comparison of distance-based query algorithms using R-trees in spatial databases. *Information Sciences*, 177(11), pp.2207-2237.

L. Becker, A. Giesen, K. Hinrichs, and J. Vahrenhold. Algorithms for performing polygonal map overlay and spatial join on massive data set. In R. H. Gu¨ting, D. Papa- dias, and F. H. Lochovsky, editors, Advances in Spatial Databases—6th International Symposium, SSD'99, vol. 1651 of Springer-Verlag Lecture Notes in Computer Science, pages 270–285, Hong Kong, China, July 1999.

Jacox, E. and Samet, H. (2007). Spatial join techniques. *ACM Transactions on Database Systems*, 32(1), p.7-es.

OS Code-Point [Shapefile geospatial data], Coverage: Great Britain, Updated November 2017, Ordnance Survey, GB. Ordnance Survey Service. Contains Royal Mail data © Royal Mail copyright and Database right 2017. Contains National Statistics data © Crown copyright and database right 2017. Contains OS data © Crown copyright and database right 2017.

Puri, S. (2015). Efficient Parallel and Distributed Algorithms for GIS Polygon Overlay Processing. *Georgia State University*, [online] Department of Computer Science. Available at: https://scholarworks.gsu.edu/cgi/viewcontent.cgi?referer=https://www.google.co.uk/&httpsredir=1&article=1099&context=cs_diss [Accessed 4 Jan. 2018].

Sccssurvey.co.uk. (2018). *Brief History of GB Coordinate Systems / SCCS News*. [online] Available at: https://www.sccssurvey.co.uk/news/2011/11/brief-history-of-gb-coordinate-systems/ [Accessed 4 Jan. 2018].

Wang, Y., Liu, Z., Liao, H. and Li, C. (2015). Improving the performance of GIS polygon overlay computation with MapReduce for spatial big data processing. *Cluster Computing*, 18(2), pp.507-516.

Zhang, M., Shi, W. and Meng, L. (2005). A generic matching algorithm for line networks of different resolutions. *8th ICA Workshop on Generalisation and Multiple Representation*, Department of Cartography(Technical University of Munich).