AUTOMATED MATCHING OF BUILDING FEATURES OF DIFFERING LEVELS OF DETAIL: A CASE STUDY

Patrick Revell and Benoît Antoine
Ordnance Survey® Research, Romsey Road, Southampton, SO16 4GU, UK
patrick.revell@ordnancesurvey.co.uk

Abstract

Ordnance Survey is the national mapping agency of Great Britain, supplying spatial data products across a wide range of scales. Currently all products are held in separate databases and no links are maintained between the features. It would be more efficient and flexible to maintain all products in a single database, with links between different representations of the same feature – a Multiple Representation Database (MRDB).

The first objective of this research is to develop tools that automatically match features from the current smaller scale products back to the corresponding features at the most detailed scale. If successful, the matching tools will allow automatic detection of currency problems between the datasets, which can then be manually corrected. The ultimate objective is to achieve automatic propagation of updates from the most detailed scale, to the smaller scale products.

An investigation was conducted into the suitability of existing data matching algorithms in Java, which yielded two candidates: GeOxygene and JCS. Both platforms contain tools for polygon matching and road network matching. User interfaces and data interfaces were developed to run the matching algorithms on a single research platform.

This paper concentrates on the matching of building features. A 20 × 30 km area was used for the tests, comprising large scale detailed buildings and coarser 1:25 000 scale buildings. In order to process the dataset, it was necessary to break it down into partitions using the road network and coastline.

To formalise evaluation of the results, allowing an objective comparison between the GeOxygene and JCS techniques, a building matching evaluation algorithm was developed. This assigns a score between 0 and 1 to each 1:25 000 scale matched building. Histograms were created from the building matching evaluation scores. These indicate better results from the Geoxygene algorithm.

The JCS matching algorithm proved unsuitable for the test data, due to its limitation of making only one-to-one matches. Its policy of trying to match everything from the coarse dataset caused incorrect matches. The GeOxygene matching produced better results, but it seems to be quite cautious in its approach to matching, meaning that the

matches which are made are good, but amongst the unmatched features there are many that should have been matched.

Future work should concentrate on improving the results from the GeOxygene matching. Some post-processes could be introduced that analyse the unmatched features and if appropriate include them in an existing match or make a new match. The matching tools will also be tested on other themes such as woodland, although adaptations would need to be made to the evaluation algorithm. With some improvement, the tools can be used to automatically identify discrepancies between spatial datasets and ultimately can be used for constructing an MRDB.

Introduction

Ordnance Survey is the national mapping agency of Great Britain, supplying spatial data products across a wide range of scales. The most detailed data are supplied as the OS MasterMap Topography layer (1:1250 scale urban, 1:2500 scale rural, 1:10 000 scale mountain and moorland) and the OS MasterMap ITN layer (containing a detailed road network consistent with the topography). The OS MasterMap products are maintained in the National Topographic Database (NTD). Other topographic data products include Meridian™ 2 (approx 1:50 000 scale) and Strategi® (1:250 000 scale).

In addition to the data products, there is a range of cartographic products supplied in raster format, namely 1:10 000 scale, 1:25 000 scale, 1:50 000 scale, 1:250 000 scale and MiniScale® (1:1 million scale). Of these, the 1:25 000 and 1:50 000 scale are also maintained in raster, whilst the others are maintained in vector.

Currently all of these products are held in separate databases and no links are maintained between the features. Features in the NTD have unique identifiers, but these are not referenced by features in the other databases. This means that any relevant changes in the NTD must be manually identified and transferred to all the smaller scale products. Consistency between products is maintained through good communication between the product maintenance teams. It would be much more efficient and flexible to maintain all products in a single database, with links between different representations of the same feature – namely a Multiple Representation Database (MRDB). For an overview of the concept of an MRDB see Devogele et. al. (1996).

Research Objectives

In Ordnance Survey Research, automatic generalisation of the NTD has been under investigation for at least eight years. The motivation behind this research has been to gain capabilities to recreate current products and derive new products from the NTD as automatically as possible. In recent years it has become apparent that with the current technology and source data, it is not possible to regenerate current products without a large amount of manual completion work. Such costs are hard to justify. The research is now focusing on deriving new products from the NTD, where there is much more flexibility in adjusting specifications to meet the strengths and weaknesses of the available generalisation algorithms and source data.

To address the problem of maintaining the existing products more efficiently, a new research direction has been initiated. The first objective of this research is to develop tools that can automatically match features from the smaller scale products back to the corresponding features in the NTD. If successful, the matching tools will allow automatic detection of currency problems between the datasets, which can then be manually corrected. There is a growing body of research on automated spatial data matching, see for example Cobb et. al. (1998), Sester et. al (1998), Walter & Fritsch (1999) and Mantel & Lipeck (2004). Most of these techniques have not yet been implemented mainstream GIS.

Once dataset currency issues have been resolved, change only updates can then be automatically propagated from the NTD to the smaller scale products. Initially this change information could be used to drive cartographers to places where manual update work is required. Ultimately the update task could be tackled automatically using incremental generalisation techniques, such as in Kilpeläinen & Sarjakoski (1995), Harrie & Hellström (1999), Anders & Bobrich (2004) and Bobzien et. al. (2005).

Software Platform and Test Data

Ordnance Survey Research conduct the bulk of their generalisation research using the Radius Clarity platform from 1Spatial (2009). This platform allows customisations to be made in Java, and as such is the main language used by the Generalisation Team for algorithm development. Processes developed on the platform operate on data held in the proprietary object-oriented Gothic Database format.

An investigation was conducted into the suitability of existing data matching algorithms in Java, which yielded two candidates. The first was GeOxygene (2008), which is an open framework implementing OGC/ISO specifications, designed for creating and deploying GIS applications. It was developed at the COGIT laboratory of the IGN (Institut Géographique National), the French national mapping agency. GeOxygene includes various contributions for spatial data handling tasks, including data matching.

The second option was Java Conflation Suite (JCS) from Vivid Solutions (JCS, 2003), which is a set of tools addressing many common geospatial data integration problems. This is based on the Java Topology Suite (JTS), an API of 2D spatial functions for geometric operations in an OGC-compliant spatial object model.

Both GeOxygene and JCS contain tools for polygon matching and road network matching. They are both open source, which is an important consideration for research, since it allows bugs to be quickly resolved without involving a third party. Both platforms were investigated, with the aim of gaining an understanding of data matching techniques and their applicability to Ordnance Survey data.

In order to invoke these data matching algorithms on features held in a Gothic Database, it was necessary to develop some routines to convert (in memory) between the models understood by Gothic, GeOxygene and JTS. The main work was to write a conversion between a GothicObjects and a JTS BasicFeatures, translating both geometry and attributes. Fortunately there were already routines to create GeOxygene GM_Objects from JTS geometries, so going via the intermediate format of JTS it was possible to convert between a GothicObject and a GeOxygene GM_Object.

Object identifiers are retained throughout the conversion process, allowing the results of the matching to be interpreted in terms of the original features. Using the Gothic Database it is possible to store one-to-one, one-to-many, many-to-one and many-to-many references between GothicObjects. To allow maximum flexibility, the results of the matching are stored using a many-to-many reference.

Test Data

This paper concentrates on the matching of building features. A 20 × 30 km area including Southampton and the New Forest was used for the tests. The detailed dataset comprised buildings extracted from the NTD. The coarse dataset came from a trial that vectorised data from the 1:25 000 scale raster product. Looking at the 1:25 000 scale buildings in Figure 1, it is clear that some significant manual generalisation has been carried out in order to depict the buildings at this scale.

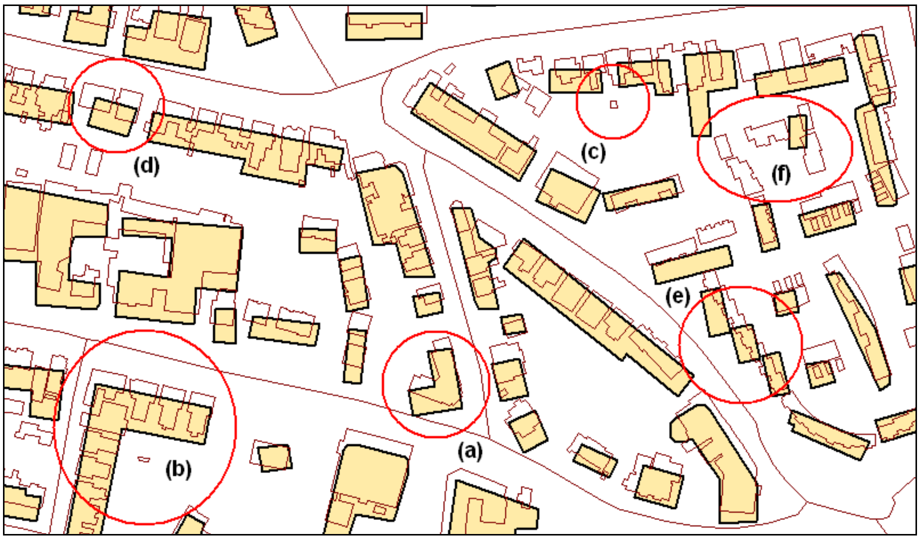

Figure 1. NTD building outlines compared with 1:25 000 scale buildings (beige).

Individual buildings have been simplified such as **(a)**, which makes a straightforward one-to-one match. More commonly, rows of buildings have been amalgamated together across gaps and then simplified such as **(b)**, which would be a many-to-one match. Small buildings have been eliminated, for example **(c)**. In some places buildings have been displaced away from their original positions to prevent overlap conflicts with the roads, such as **(d)**. Occasionally the cartographer has broken up a single NTD building into more than one 1:25 000 scale building for instance **(e)**, which would be a one-to-many match.

There are also problems of currency between the two datasets, such as **(f)**, which are due to the different product creation and maintenance regimes. A suitable data matching algorithm must be able to cope with all the differences in representation and currency between the NTD and 1:25 000 scale datasets.

JCS Polygon Matching Algorithm

The JCS polygon matching algorithm uses a set of up to six measures for deciding which features to match. The user has control over which measures are employed and can assign weights to the measures, that are then aggregated to create a score between 0 and 1. The measures are defined as follows:

- **Centroid Distance**. The distance between the centroids of the two polygons.
- **Vertex Hausdorff Distance**. The Hausdorff distance measures the greatest local deviation between two geometries. The Vertex Hausdorff distance is the Hausdorff distance restricted to the vertices of one of the geometries. This is easier to compute in terms of both algorithmic complexity and performance. It is equal to the Hausdorff distance in many cases and when not equal it still provides a useful measure of how far apart two geometries are (JCS, 2003). Before applying this measure, the geometries are shifted so that their centroids are coincident.
- **Symmetric Difference**. This is the total area of the non-overlapping portions of the two geometries.
- **Symmetric Difference Centroids Aligned**. The same as above but before calculating the symmetric difference, the geometries are shifted so that their centroids are coincident.
- **Compactness**. This is a measure of the similarity of the compactness values (area to perimeter ratio) of the two polygons.
- **Angle Histogram.** The difference between each polygon’s angle histogram, which is a histogram of the angles that the segments make with the positive x-axis, weighted by segment length. The more similar the histograms for the two shapes, the higher the score. It is possible to specify the number of bins into which to split the angle space (-p to p). For example, specifying 18 bins will create bins each of size 20° (JCS, 2003).

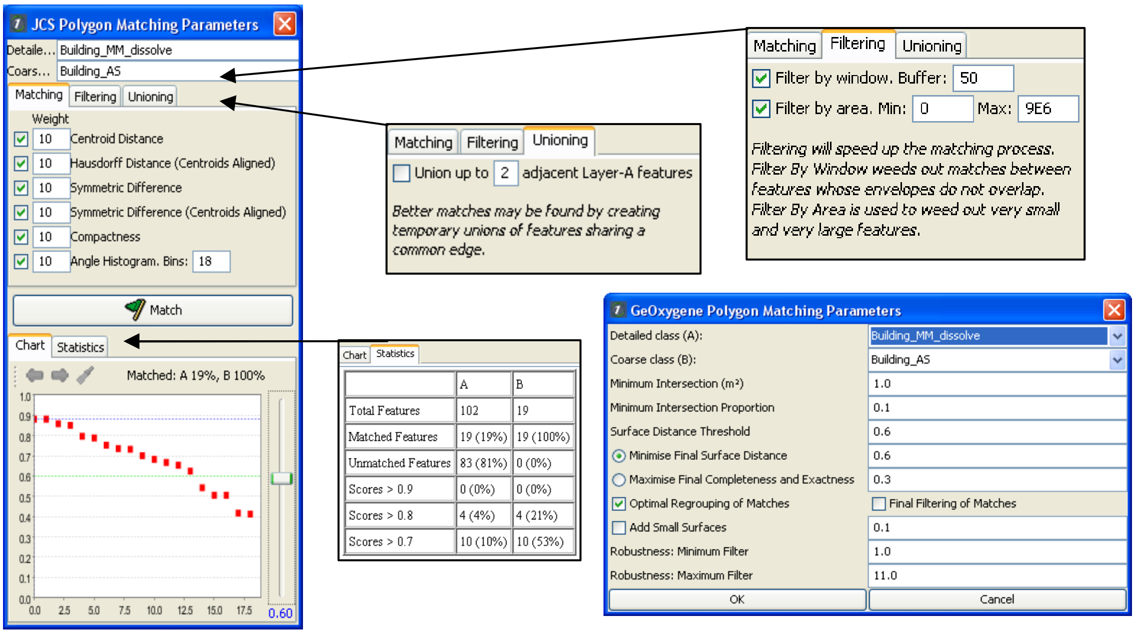

Figure 2. Polygon Matching User Interfaces: JCS (left) and GeOxygene (right)

To limit the number of comparisons, it is possible to specify a filter buffer distance, that effectively prevents comparisons between polygons that are more than a certain distance apart. It is also possible to ignore very small or very large polygons, but this option was not used, since it would be more flexible to implement as part of a separate pre-selection process, if required. The polygon matching algorithm can dissolve adjacent polygons (that share a common line segment), but it was decided that it would be better to accomplish this as a separate pre-process.

The JCS tool displays matching scores in descending order on a graph. By clicking on the data points on the graph, the corresponding features are highlighted in the map display. The idea is to visually determine a cut-off point for valid matches. The JCS interface was originally available as a plug-in for the open source GIS JUMP (JCS, 2003), but it was easily adapted to run in Clarity. This is shown on the left of Figure 2.

GeOxygene Polygon Matching

The GeOxygene polygon matching algorithm originates from the work of Bel Hadj Ali (2001). The implementation had no user interface, so the interface on the right of Figure 2 was developed to allow experimentation with the parameters:

- **Minimum Intersection and Minimum Intersection Proportion**. In order for polygons to be matched, they must overlap by at least the value of the first threshold (in square metres) and the proportion of their geometry which overlaps must be at least the value of the second threshold. A robust intersection algorithm is employed, which in the event of failure uses a simplification algorithm to filter the geometries before intersecting them (obviously with some loss of accuracy). A maximum and minimum tolerance can be given, allowing a range of different tolerances to be tried.

- **Surface Distance Threshold**. Matches are discarded if the Surface Distance = 1–area(intersection)/area(union) (Vauglin, 1997) is greater than this threshold. Note that identical polygons have a value of 0 and disjoint polygons have a value of 1.
- **Evaluation Criteria:**
    - **Minimise Final Surface Distance**. This option is recommended for data with similar levels of detail. Only matches with a Surface Distance less than the maximum threshold value are retained.
    - **Maximise Final Completeness and Exactness**, where Completeness = area(intersection)/area(coarse) and Exactness = area(intersection)/area(detailed). (Bel Hadj Ali, 2001). Only matches with both Exactness and Completeness greater than the threshold value are retained. Note that identical polygons have values of 1 and disjoint polygons have values of 0.
- **Optimal Regrouping of Matches**. This option refines the result of the initial matching by maximising the surface distance between the groups of detailed and coarse polygons. It produces a better quality matching result on the test data.
- **Final Filtering of Matches**. Option to do a final filtering of matches using Surface Distance or Completeness/Exactness criteria. This option tends to give poor results.
- **Add Small Surfaces**. Small polygons in the detailed dataset are often not matched by the algorithm. This is an optional post process that identifies small unmatched detailed polygons which are adjacent to (touching) groups of detailed polygons that have been matched. If the ratio of their areas is greater than this threshold, then the small polygon is appended to the match.

Partitioning for Building Matching

The polygon matching algorithms cannot cope with large datasets, due to the number of comparisons involved and the overhead of holding a large amount of spatial data in memory. The solution to this problem was to use the detailed ITN road network along with the NTD coastline to break the dataset down into 4393 partitions. The buildings inside each partition could then be processed independently from the other buildings.

The use of partitions was not straightforward, since there are many examples where NTD or 1:25 000 scale buildings overlap the partition boundaries. In the case of the NTD buildings, these were unusual buildings which had roads running underneath them, such as large shopping complexes. The solution was to merge together the affected partitions, until each large scale building was in one and only one partition.

For the 1:25 000 scale buildings, the overlaps were due to the higher level of detail and positional accuracy of the ITN roads. There were so many overlaps, that merging the affected partitions created partitions that were too large for the polygon matching

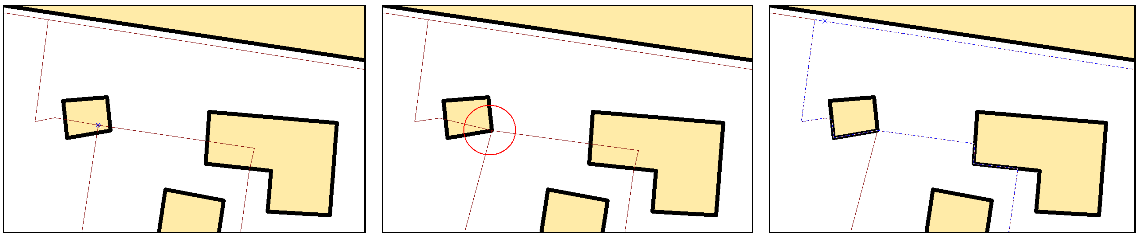

Figure 3. Adjusting partition nodes and edges to fit 1:25 000 scale buildings.

algorithms to cope with. A more focused strategy was required, which started by merging small partitions (less than 1600m²) into adjacent partitions. There were many examples of 1:25 000 scale buildings which only protrude beyond a partition boundary by a small amount. The worst case is when a building overlaps with more than two partitions. The solution here is to move the road junction node to the boundary of the building, as shown on the middle picture of Figure 3.

Cases where a building overlaps with only two partitions are easier to handle. These are resolved by reshaping the partition boundary so that the building falls totally inside the partition that it was mostly inside before, as shown on the right of Figure 3. Some cases cannot be resolved using these techniques, so any remaining conflicting partitions were merged together, creating 2436 final partitions. Theoretically buildings which should be matched could end up in separate partitions. In future the roads will be matched first, allowing the corresponding regions between datasets to be determined more accurately.

Pre-Processing for Building Matching

In the NTD, each individual building footprint is captured, so that the data includes internal building divisions. At 1:25 000 scale these building divisions are not shown. Retaining the large scale building divisions significantly slowed down the data matching process and produced poor quality results. Therefore it was decided to run a pre-process to dissolve together all adjacent buildings and store the results in a new feature class. A database reference was retained between the original and the dissolved buildings, so after matching it would be possible to trace back to the source buildings. The 285609 original NTD buildings were reduced down to 143204 dissolved buildings.

It would also be possible to do further pre-processing to facilitate the matching process. For example deleting small NTD buildings, amalgamating, and displacing the buildings away from roads. However amalgamation and displacement would be quite risky, since there is no guarantee that the automated decisions would be the same as those made when the 1:25 000 buildings were originally manually generalised.

Figure 4. Visualising GeOxygene matching results using colours (left) and lines (right).

Evaluating Building Matching

For tuning the parameters, the user interface was set up to allow the building matching to be run on selected partitions. On completion, matched and unmatched buildings are displayed in different colours (NTD matched: green, unmatched: pink, 1:25 000 scale matched: beige, unmatched: red, on the left of Figure 4). A tool was developed to visualise the match by plotting blue lines connecting the centroids of the NTD buildings with the corresponding 1:25 000 scale buildings, as shown on the right of Figure 4. The user interface allows the matching to be backtracked, allowing easy experimentation with different parameter settings.

The next stage of the project was to find a way of evaluating the results from JCS and GeOxygene for the entire test dataset. A common evaluation methodology is to first conduct a manual matching exercise, which then is used as a baseline to compare against the automated results. Given the size of the test dataset, it was not possible to do the manual matching within the timescales of the project. From the Ordnance Survey business perspective, manual matching is never an option, since the cost of this tedious, time-consuming exercise would outweigh any benefits.

Instead, to formalise the evaluation of the matching results, allowing an objective comparison between the GeOxygene and JCS techniques, it was decided to develop an automated building matching evaluation algorithm. The results of the evaluation are stored on the 1:25 000 scale buildings, since we are interested in how well these are matched with the NTD buildings, rather than the converse. This evaluation algorithm makes use of the following measures:

- **Area**. The total area of a building geometry.
- **Centre of Gravity**. The centroid of a building geometry.
- **Elongation**. The elongation is the mathematically correct Smallest Minimum Bounding Rectangle's (SMBR's) length divided by its width.
- **SMBR Orientation**. This is the orientation of the longest side of the SMBR.
- **Convexity**. This is the total area divided by the area of the convex hull.
- **Wall Statistical Weighting** (Duchêne et. al., 2003). This measures the orientation of a building, based on the orientation of each wall, weighted by the wall lengths.

For comparing the measured values between the matched buildings, the following coefficients are used. These all return a value between 0 and 1, where 1 represents a perfect correspondence. The Centroid Distance Coeff, will always return 1 if the centroids are further than 50 metres apart. The SMBR orientation is calculated modulo π, so the AngleDistance cannot be more than π/2. The wall orientation is calculated modulo π/2, so the AngleDistance cannot be more than π/4.

- **Area Coeff** = Min{area1, area2}/Max{area1, area2}
- **Centroid Distance Coeff** = Max{1 – Distance(centroid1, centroid2)/50, 0}
- **Elongation Coeff** = Min{elong1, elong2}/Max{elong1, elong2}
- **SMBR Orient Coeff** = 1 – AngleDistance(orient1, orient2)/(π/2)
- **Convexity Coeff** = Min{conv1, conv2}/Max{conv1, conv2}
- **Wall Orient Coeff =** 1 – AngleDistance(orient1, orient2)/(π/4)

If a 1:25 000 scale feature is matched with only one NTD feature, it is easy to calculate these coefficients. Otherwise if there are a number NTD features involved in the match, the comparison is more difficult. By summing the areas of the NTD features, the Area Coeff can be calculated. Similarly, the wall statistical weighting measure can be applied to the set of walls from a group of NTD buildings (Revell, 2004), allowing the Wall Orient Coeff to be calculated as before.

For the other coefficients it is necessary to first combine the NTD buildings involved in a match into a single geometry by amalgamating them (across gaps). A very basic algorithm was used for this step, but in future a better algorithm should be substituted. Once there is a single geometry representing the NTD features, it is possible to calculate the centroid, SMBR and convex hull, allowing the Centroid Distance, SMBR Orient, Elongation and Convexity Coeffs. All seven coefficients are combined together in a weighted sum, with the Centroid Distance having double the weight of the other coefficients. This final evaluation score is stored on each 1:25 000 scale building. The choices of measures and the weighting between them was determined by trial and error.

The intention is that matches with a low evaluation score will draw the operator to places where the matching is uncertain. The problem might be a commission error – i.e. the building geometries do not correspond because the wrong buildings have been

	JCS				GeOxygene			
	Num	**%**	**Area m²**	**%**	**Num**	**%**	**Area m²**	**%**
1:25k Unmatched	381	0.7	91234	0.5	5431	10.5	904810	5.0
1:25k Matched	51203	99.3	18061440	99.5	46153	89.5	17247864	95.0
NTD Unmatched	92001	64.2	5102953	29.6	60838	42.5	2728946	15.8
NTD Matched	51203	35.8	12155032	70.4	82366	57.5	14529039	84.2

Table 1. Results of JCS and GeOxygene building matching.

matched. The problem could also be an omission error – i..e. insufficient buildings from the NTD have been matched with a 1:25 000 scale building. Note that the algorithm does not indicate complete omissions – i.e. a 1:25 000 scale building that was not matched with anything, when there was one or more suitable NTD buildings with which to match it.

Results of Building Matching

After tuning parameters on some small regions, the entire dataset was processed using both the GeOxygene and JCS polygon matching, with the parameters shown in Figure 2. The matching evaluation algorithm was run as a post-process. The GeOxygene matching took 32 minutes, while the JCS matching took 12 minutes. Table 1 shows the total number and total area of matched and unmatched features in both datasets.

For JCS only 0.7% of the 1:25 000 scale buildings were unmatched, representing just 0.5% of the total building area. This seems incredibly good, but is really part of the design of the JCS algorithm, since it tries to make a one-to-one match from each coarse feature to a detailed feature, resulting in commission errors. The onus is on the user to discard matches with a low score afterwards. Only 35.8% of NTD features have been matched, which is low mainly because JCS does not make many-to-one matches,

Figure 5. Results of JCS matching. The unmatched NTD buildings are shown in pink. Ordnance Survey © Crown Copyright. All rights reserved.

causing many omission errors as shown by the pink buildings in Figure 5. The total area of matched features is high at 70.4%, mainly because when there is a choice of NTD features with which to match, the algorithm tends to choose the largest NTD building.

For GeOxygene 10.5% of the 1:25 000 scale buildings are unmatched, but this represents only 5% of the total building area. Such buildings usually have no NTD buildings in proximity, which indicates a problem of currency between the two datasets (in contrast to JCS, which nearly always finds a match, even if inappropriate). 42.5% of the NTD buildings are unmatched, but this represents only 15.8% of the total building area. Many of these are small buildings which are simply not depicted at 1:25 000 scale.

There are also many NTD buildings which should have been matched, but have not been due to large differences in building position (omission errors). These are shown on the left of Figure 4. Sometimes no match is made at all, indicated by the red 1:25 000 scale buildings. Sometimes a match is made, but one or more NTD buildings are omitted. This often happens at the end of rows of buildings, highlighted by the pink NTD buildings in Figure 4. For some NTD buildings there is simply no equivalent at 1:25 000 scale, indicating a problem of currency.

Histograms were created from the building matching evaluation scores. In Figure 6 the automated evaluation scores indicate better results from the Geoxygene algorithm, evidenced by a greater skew towards 1, with much higher frequencies of good scores.

Discussion and Conclusions

The JCS matching algorithm has proved unsuitable for the test data, due to its limitation of making only one-to-one matches. Its policy of trying to match everything from the coarse dataset caused incorrect matches. The GeOxygene matching produced better, but by no means perfect, results. The algorithm seems to be quite cautious in its approach to

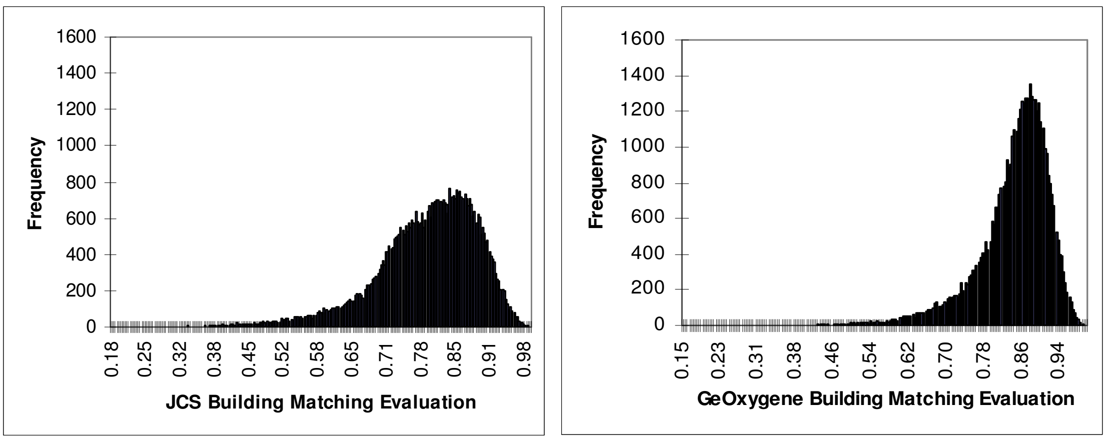

Figure 6. Histograms showing distribution of JCS and GeOxygene evaluation scores.

matching, which means that the matches which are made are good, but amongst the unmatched features there are many that should have been matched.

Future work should concentrate on improving the results from the GeOxygene matching. Some post-processes could be introduced to analyse the unmatched features and if appropriate include them in an existing match or make a new match. Any remaining unmatched features would then represent possible dataset currency problems, with the exception of the NTD buildings that are too small to depict at 1:25 000 scale.

Conducting a manual matching exercise on a large dataset for evaluation purposes is not feasible at Ordnance Survey, but it would still be worthwhile accumulating a catalogue of small sample datasets that represent common matching problems. Results of manual matching for these small samples could then be recorded and used as the basis of a testing framework for the matching and evaluation algorithms.

Two polygon matching algorithms have been tested on building data and an automated technique has been developed for evaluating them. Results from the GeOxygene matching are promising. In future the matching tools will be tested on other themes such as woodland, although the evaluation algorithm would need to be adapted. With some improvement the tools can be used to automatically identify discrepancies between spatial datasets and ultimately can be used for constructing an MRDB.

References

1Spatial, 2009. *Radius Clarity – a rules based environment for automated generalisation*. [Online] Available at: http://www.1spatial.com/products/radius_clarity/ [Accessed 14 January 2009].

Anders, K.-H. and Bobrich, J., 2004. *MRDB approach for automatic incremental update*. ICA Workshop on Generalisation and Multiple Representation, Leicester.

Bel Hadj Ali, A., 2001. *Qualité géométrique des entités géographiques surfaciques. Application à l'appariement et définition d'une typologie des écarts géométriques*. Ph.D. Paris: Université de Marne la Vallée.

Bobzien, M., Burghardt, D. and Petzold, I., 2005. *Re-generalisation and construction – two alternative approaches to automated incremental updating in MRDB*. In proceedings of the 22nd International Cartographic Conference. La Coruña, Spain.

Cobb, M, Chung, M., Foley, H., Petry, F., Shaw, K. and Miller, H., 1998. *A rule-based approach for the conflation of attributed vector data*. GeoInformatica 2(1), pp. 7-35.

Devogele, T., Trevisan, J. and Raynal, L., 1996. *Building a multiscale database with scale-transition relationships*. In proceedings of 7th International. Symposium on

Spatial Data Handling (SDH), Advances in GIS Research II, Kraak, M.-J. and Molenaar, M. (editors), pp. 6.19-6.33, Delft.

Duchêne, C. et al., 2003. *Quantitative and qualitative description of building orientation*. [Online] Available at: http://www.geo.unizh.ch/ICA/docs/paris2003/papers/duchene_et_al_v1.pdf [Accessed 14 January 2009].

GeOxygene, 2008. *GeOxygene sourceforge project*. [Online] Available at: http://oxygene-project.sourceforge.net/ [Accessed 14 January 2009].

Harrie, L. and Hellström, A. K., 1999. *A prototype system for propagating updates between cartographic data sets*. The Cartographic Journal 36(2): pp 133-140.

JCS, 2003. *Vivid Solutions: Java Conflation Suite (JCS) documentation*. [Online] Available at: http://www.vividsolutions.com/jcs/doc.htm [Accessed 14 January 2009].

Kilpeläinen, T. and Sarjakoski, T., 1995. *Incremental generalization for multiple representations of geographic objects*. In GIS and Generalization: Methodology and Practise, Müller, J. C., Lagrange, J. P. and Weibel, R. (editors). Taylor & Francis, London, pp. 209-218.

Mantel, D and Lipeck, U. W., 2004. *Matching cartographic objects in spatial databases*. In Proceedings of the XXth Congress of the International Society for Photogrammetry and Remote Sensing (ISPRS), Istanbul, July 2004, Commission IV Papers, Number Vol. XXXV, Part B4 in International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 172-176.

Revell, P., 2004. *Generalising OS MasterMap rural buildings to 1:50 000 scale*. [Online] Available at: http://ica.ign.fr/Leicester/paper/Revell-v2-ICAWorkshop.pdf [Accessed 14 January 2009].

Sester, M., Anders, K.-H. and Walter, V. 1998. *Linking objects of different spatial data sets by integration and aggregation*. GeoInformatica, 2(4), pp. 335-358.

Vauglin, F. 1997. *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires*. Ph.D. Paris: Université de Marne la Vallée.

Walter, V and Fritsch, D., 1999. *Matching spatial data sets: a statistical approach*. International Journal of Geographical Information Science 13(5), pp. 445-473.