YOUR WORLD SMARTER

Case Study

Environment Agency

Increasing data quality and performance at the Environment Agency

Background

National Flood and Coastal Defence Database
The Environment Agency (EA) is in the process of examining various aspects of the National Flood and Coastal Defence Database (NFCDD) system with respect to reliability and performance. One of the aspects examined has been the impact on the performance of their browser applications of both the quality and quantity of spatial data held within their Oracle database. The complexity of their spatial data (see Fig.1), both in size or precision, has a direct effect on both the speed of data retrieval and queries against this data by the user.

The EA has initiated a sensitivity analysis into their spatial data. The definitive objective of the work was to establish that the results of analysis on the data were comparable given the simplification of the data.

This would reduce the time taken by the Oracle database to perform spatial queries. It would also mean that less data would need to be transferred between application tiers, speeding up image rendering for the map display in the browser application.

Fig. 1: An EA dataset

Data re-engineering

The data was simplified using three different algorithms. Two algorithms were outlined by the EA (Fig. 3 Node Collinear Delineation and Angle Between Vectors) and the third algorithm was initially recommended by 1Spatial (Douglas and Peucker). Each algorithm was invoked three times on the dataset, each time using a different tolerance parameter.

At all stages of the data re-engineering exercise, all data transformations and processing had to be accounted for. Due to the complexity of the data, the processing operations required thousands of computational operations in order to generalise the information. Therefore, the algorithms chosen had to be efficient to ensure that implementing the data re-engineering was feasible. At the same time there were a number of key factors that had to be addressed to ensure that the re-engineered data still fit into the EA's business model.

In order for data analysis to produce accurate results efficiently, the data held by the database must be 'clean'. Whether the analysis is performing an area calculation or a spatial search, data containing geometric errors (see Fig. 2) will return incorrect results. In some cases features may be ignored in the analysis, if they are geometrically invalid. It is therefore important, that data is cleaned correctly before it is analysed and before any simplification processes are invoked on the data.

Initial testing demonstrated that it was possible for the smoothing algorithms to introduce gaps and overlaps between features that did not exist beforehand. This was common where polygons only shared part of their boundaries with other polygons. It was essential that the data re-engineering maintained topological consistency throughout all data processing activities. The approach was therefore taken to topologically structure the data and then smooth the resulting topology. After the topology has been smoothed, all real-world feature geometries that referenced this topology would then be automatically updated.

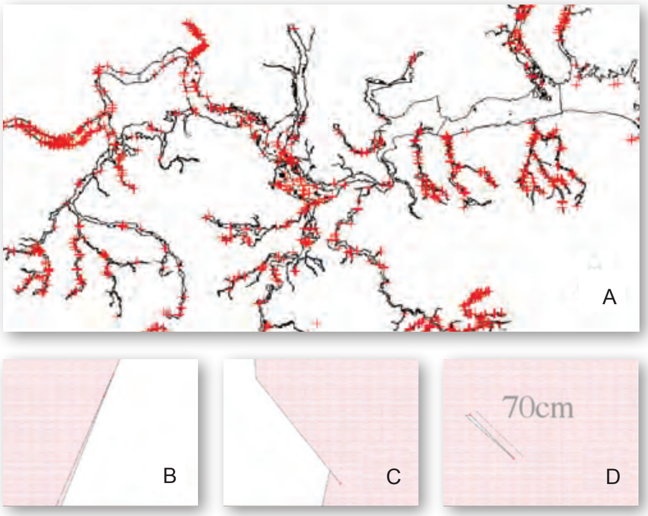


Fig. 2: Geometric errors

A - Geometric errors are represented by a red symbol

B - A kick back located in the source data

C - An inverted splice located in the source data

D - A small hole in a polygon

Results

A variety of different datasets were re-engineered by 1Spatial. Initially the datasets contained between a 1% to 100% error rate. After cleaning these datasets through the combination of geometric and topological analysis, all datasets contained between a 0% to 2% error rate. Some errors did require manual correction.

All smoothing algorithms significantly reduced the average number of vertices in each feature (see Fig. 3). Qualitatively analysing the features also showed that all algorithms maintained the shape of the features (with the exception of the Angle Between Vectors algorithm that used a 10° parameter).

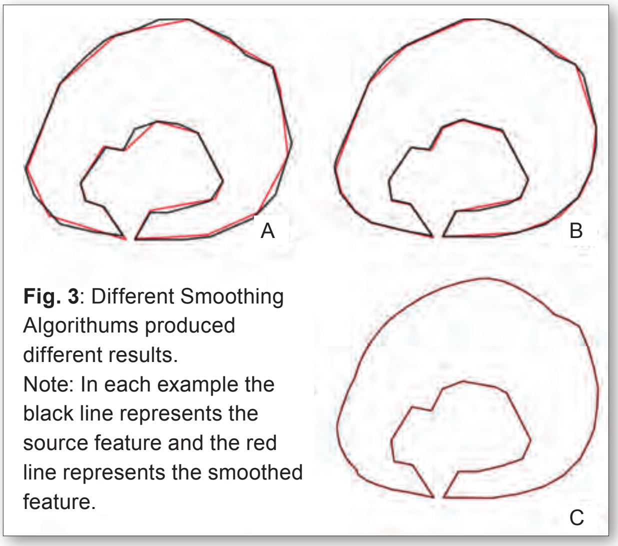


Fig. 3: Different Smoothing Algorithums produced different results.
Note: In each example the black line represents the source feature and the red line represents the smoothed feature.

All algorithms only reduced the average area and perimeter of each feature by less than 2% (with the exception of the Angle Between Vectors algorithm that used the 10° parameter, and the 3-Node Collinear Delineation algorithm that used the 5m parameter). This also supports the fact that the algorithms qualitatively maintain the shape of the original feature. The average statistics imply that either the 1m or 2m Douglas and Peucker algorithm should be applied to the data. This algorithm reduced the average number of vertices that make up each feature by 50%, whilst reducing the average perimeter length of each feature by less than 1% and by reducing the average area of each feature by less than 1%.

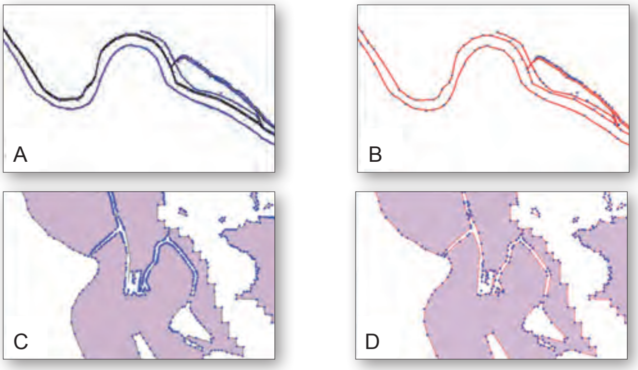


Fig. 4: An example of polygons both pre (A and C) and post (B and D) data re-engineering.

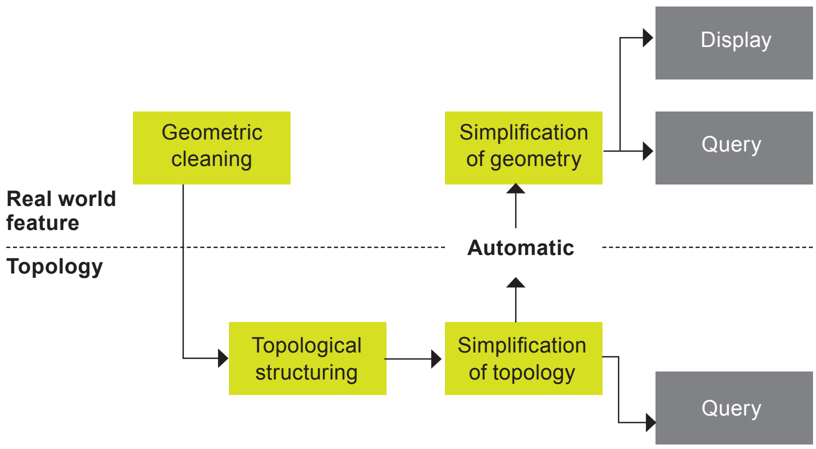
The Douglas and Peucker algorithm was finally selected because the EA only wanted to simplify but not significantly change the shape of the polygons. The data stored and maintained by the EA is very sensitive. They could not risk over-simplifying the data and in doing so move numbers of address properties in or out of the polygons. This algorithm both quantitatively and qualitatively produced the best results (see Fig. 4).

Moving forward

The EA worked with 1Spatial to evaluate this re-engineering task. This involved using 1Spatial's expertise in defining Oracle-based data models and their technology to process data. A flowline was developed to combine both geometrical and topological analysis to perform the necessary data re-engineering (see Fig. 5). The results showed that as a result of the data re-engineering exercise, rendering an image gave a 115% increase in performance, whilst spatial query using address points gave a 229% increase in performance.

Moving forward, additional investigations will be carried out to assess further performance improvements gained from querying the topological, rather than geometrical information in the Oracle database.

Fig. 5: The Data Re-engineering Work Flow.



Visit us **1spatial.com**