

Visualisation of Origins, Destinations and Flows with OD Maps

Jo Wood, Jason Dykes and Aidan Slingsby

giCentre, City University London, UK
Email: jwo@soi.city.ac.uk

We present a new technique for the visual exploration of origins (O) and destinations (D) arranged in geographic space. Previous attempts to map the flows between origins and destinations have suffered from problems of occlusion usually requiring some form of generalisation, such as aggregation or flow density estimation before they can be visualized. This can lead to loss of detail or the introduction of arbitrary artefacts in the visual representation. Here, we propose mapping OD vectors as cells rather than lines, comparable with the process of constructing OD matrices, but unlike the OD matrix, we preserve the spatial layout of all origin and destination locations by constructing a gridded two-level spatial treemap. The result is a set of spatially ordered small multiples upon which any arbitrary geographic data may be projected. Using a hash grid spatial data structure, we explore the characteristics of the technique through a software prototype that allows interactive query and visualisation of 10^5 – 10^6 simulated and recorded OD vectors. The technique is illustrated using US county to county migration and commuting statistics.

Keywords: visualisation, geovisualisation, OD map, OD matrix, treemap, flow, trajectory, migration

INTRODUCTION

There are many applications for which associations between pairs of known geographic locations provide an important data source. These associations might involve direct movements of people, for example, migration (Tobler, 1987), commuting behaviour (Chiricota *et al.*, 2008), GPS tracklog analysis (Andrienko and Andrienko, 2008) or interaction between actors in social networks and service use (Radburn *et al.*, 2009). Alternatively, movements of goods, knowledge (Paci and Usai, 2009), disease and animals (Gilbert *et al.*, 2005; Guo, 2007) may be explored. A range of techniques have traditionally been used to represent such associations, including direct mapping of geographic flow vectors (Tobler, 1987), flow density maps (Rae, 2009), origin–destination (OD) matrices (Voorhees, 1955) and statistical summaries of spatial association (Gilbert *et al.*, 2005).

We present the OD map – a new method of visualizing associations between datasets of this type that overcomes some of the problems traditionally associated with mapping geographic vectors. Our aim is to develop visualisation techniques that enable trends in OD relationships to be identified and their sensitivities explored while maintaining maximum cognitive plausibility of the representation (Skupin and Fabrikant, 2003). In so doing, we contribute a spatially embedded view of trajectories that may be combined effectively with other multiply-coordinated views.

REQUIREMENTS AND PRIOR WORK

This work was motivated by the need to analyse and understand patterns in a number of large datasets in which paired origins and destinations were key characteristics. These included collections of mobile search enquiry locations and result destinations (Wood *et al.*, 2007), spatio-temporal records of borrowing behaviour among library users relating home origin to library destination (Radburn *et al.*, 2009) and GPS records of traffic across London recorded by all vehicles from a courier company over a 1 month period that included pick-up (origin) and drop-off (destination) points (Slingsby *et al.*, 2008).

In each case, our objective was to gain insight into the nature of the journeys defined between pairs of locations. Developing an environment for the visual exploration of movement and other interactions in time and space has potential for building knowledge and understanding, but doing so has been recognized as one of the major research challenges associated with visual analytics (Hernandez, 2007; Andrienko *et al.*, 2008). When dealing with such large data volumes, the transformation into a meaningful visual representation of the spatio-temporal structure requires data reduction through selection or aggregation in a manner that suits the need of the analyst (Rae, 2009). This led to the following requirements:

1. to be able to visually represent large numbers of origin–destination vectors (of order 10^5 – 10^6);

2. to create a visual environment that provides both overview and detail on demand;
3. to emphasize representation of the origin and destination locations over the geometry of the paths that link them;
4. to use a projection that preserves the spatial configuration of the study area to allow integration of supplementary geographic data;
5. to provide a visual representation that can show both long and short trajectories with minimal occlusion;
6. to be able to distinguish origin from destination without visual clutter and so infer direction of flows;
7. to be able to compare origin–destination vectors with destination–origin vectors;
8. to be able to distinguish artefacts of the visualisation from characteristics of the data under investigation.

Existing techniques for flow mapping meet some of our requirements. Tobler (1987) provides some early examples of computer generated flow maps, Rae (2009) in his review of flow mapping methods showed how some of these techniques could be implemented in geographic information systems. The principle behind direct flow mapping is to project geographic space onto a plane and plot each trajectory as a line from origin to destination. While this approach has a long history and can produce maps that are familiar to many users, it does not scale well to large numbers of trajectories. As the number of links increases, it also becomes increasingly difficult to indicate the direction of flow visually without clutter. Occlusion of trajectories by others sharing the same space produces maps that are difficult to interpret unless some form of generalisation is applied. This problem increases as data sizes grow to orders of 10^5 or more and has resulted in a trend identified in Andrienko *et al.* (2008) towards the ‘derivation, depiction and visualization of abstract data summaries [such as] – aggregates, generalization, samples’. For example, Cui *et al.* (2008) and Holten and van Wijk (2009) have proposed aggregation of flows of high local density to overcome this problem; Guo (2009) proposed aggregation of flows according to space and attribute. However, the spatial distortion involved in combining flows may result in unacceptable loss of detail, and perhaps more significantly, still retains the problem of long trajectories occluding shorter ones that occupy the same graphic space. Even with an acceptable level of generalisation, such as the flow density surfaces of Rae (2009), density of flow lines does not necessarily indicate the density of origin and destination locations and can result in arbitrary patterns of flow density. The suitability for flow line density will depend on the importance attached to the path of the trajectory connecting origin and destination. In our case, we are more interested in the topology of connections between origin and destination rather than their geometry (requirement 3 above).

An alternative solution to the occlusion problem is to filter selected flows (Tobler, 1987). This may include only filtering to a single limited set of origins or destinations (Phan *et al.*, 2005). While this can declutter the display of flow maps, it requires some intelligent selection of the origins and destinations to filter, or dynamic interaction to

vary these on demand. It is also difficult to provide a visual overview using this approach.

Instead of plotting trajectories as vectors, another commonly used approach is to construct some form of origin–destination matrix (OD matrix) where matrix rows represent the locations of flow origins and columns the locations of destinations. Ghoniem *et al.* (2004) compared the usability of node-link views and matrix views and found that for more than 20 nodes, the matrix view was superior for most tasks, although notably, ‘path finding’ was the only exception. Path finding or similar geographic interpretation of flows may be important for geographically arranged trajectories when, for example, considering routes taken between mobile search locations and result destinations, home location and library or traffic flows in a large city to evaluate and plan the delivery of services.

Visualizing the OD matrix directly also has a long history (Wilkinson and Friendly, 2009) where the number of flows between an origin and destination is used to colour the appropriate matrix cell. Some attempts have been made to view the matrix as a three-dimensional surface (Marble *et al.*, 1997), although the cognitive plausibility of the results is highly questionable. To enhance the utility of OD matrices, some form of sorting and aggregation is frequently applied. However, there are many ways to aggregate and reorder (Wilkinson, 1979; Guo and Gahegan, 2006). This approach has been applied for computational efficiency, especially for sparse matrices, but for visualisation, the main benefit of matrix sorting is to make clusters more apparent, famously illustrated by Bertin (1983). Reordering can be achieved in a number of ways – for example, by clustering strongly connected nodes (Jarvis and Patrick, 1973; Karypis and Kumar, 2000), to reveal spatial clusters (Andrienko and Andrienko, 2008; Guo *et al.*, 2006; Guo, 2007) or to preserve spatial proximity relations (Marble *et al.*, 1997).

While cluster detection has obvious benefits when analysing matrices, the reordering required fails to meet requirement 4 outlined above as the original spatial configuration of the matrix cells is lost during reordering. Guo (2007) and Guo *et al.* (2006) suggest overcoming this problem by providing dynamic linking between the transformed matrix space and a conventional geographic view. Alternative strategies for preserving some of the spatial properties of the matrix cells include reprojecting two-dimensional space into one-dimensional sequences using space-filling Morton ordering (Marble *et al.*, 1997) and aggregation into identifiable spatial regions (Andrienko and Andrienko, 2008). These approaches, which we might term ‘quasi space-preserving’, allow some meaningful geographical interpretation of results, but are limited in their ability to integrate with other geographic data in the same projected space.

The approach we propose here attempts to retain as much of the geographic space as possible by dividing it into a regular grid nested at two levels. As with the work of Andrienko and Andrienko (2008), this is a form of $S \times S$ aggregation, but unlike their regional clustering, the regular grid is a property of the geographic space, not the clustering algorithm or dataset. This has a significant benefit in interpretation and integration with supplementary spatial

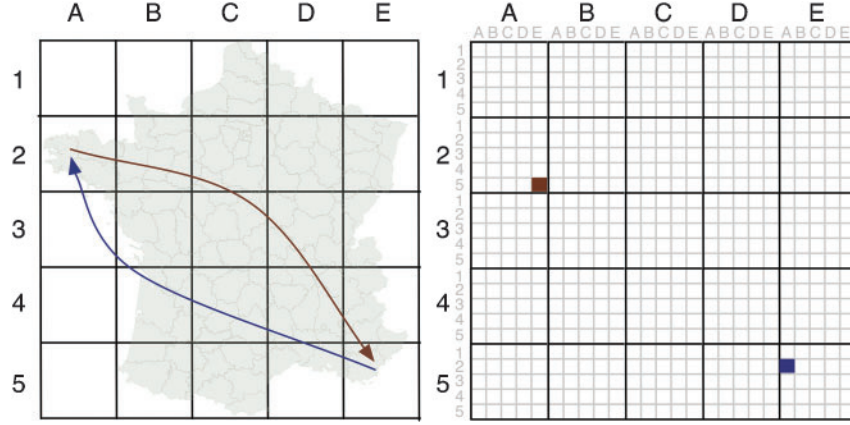


Figure 1. Left: geographic space partitioned into a regular grid. Right: OD map space. Origin space uses identical gridding of geographic space (dark grid lines); destination space consists of nested small multiples of the geographic space (light grid lines). The red trajectory from geographic location (A,2) to (E,5) is represented by the single OD map cell with coordinates (AE,25). The reverse blue trajectory is shown as a cell with coordinates (EA,52)

information which may support spatial tasks (such as the analysis of spatio-temporal pattern across a city) more effectively (Ghoniem *et al.*, 2004). Because of the spatial autocorrelation of most geographic phenomena, retaining spatial structure can also reveal clustering in the data.

THE OD MAP

To overcome problems of occlusion inherent in node-link representations of origin–destination vectors, we propose a transformation of graphical space whereby each two-dimensional vector \vec{v} from origin (p_o) to destination (p_d) is represented by a cell p_{od} in a two-dimensional matrix. But unlike a conventional OD matrix, we order cells to reflect their original two-dimensional geographic location. This is achieved by dividing geographic space into a regular coarse grid. Each trajectory can therefore be referenced by two grid cell locations – the grid cell in which the trajectory’s origin lies, and the grid cell in which its destination lies. By nesting the destination grid cells within the origin grid cells, we can preserve the spatial relationships between origin cells and the relationships between destination cells in a single matrix (Figure 1).

Assuming that both OD map columns and rows are numbered from 0, the transformation from OD map to geographic space involves a simple rounding and modulus arithmetic operation

$$p_o = \text{floor}(p_{od}/n) \quad (1)$$

$$p_d = p_{od} \bmod n \quad (2)$$

where n is the number of cells along one side of the geographic space. The transformation from geographic to OD space simply involves adding the two locations, scaling origin space by n

$$p_{od} = np_o + p_d \quad (3)$$

The result is a set of small multiples of destination space each arranged in their geographic position in origin space. This is a

special case of a treemap (Shneiderman, 1992) in which the top two levels of the hierarchy are an identical set of spatial nodes. These may be consistently sized cells that partition space as a regular raster or the result of a spatial treemap (Wood and Dykes, 2008). Indeed, the arrangement is a spatial treemap, denoted as `sHier(/, $origin, $destination); sLayout(/, SP, SP); sSize(/, FIXED, FIXED)` using the hierarchical visualisation expression language (HiVE) to describe the representation (Slingsby *et al.*, 2009).

The recursive layout has similarities to the Map² arrangement of trajectories proposed by Guo *et al.* (2006) in their VIS-STAMP system, but unlike their system, we guarantee that the projection of origin space is an identical, but scaled, projection of destination space. This consistent spatial arrangement between original and destination maps is likely to reduce cognitive load in comparison to arrangements that use a different spatial projection for origin and destination spaces.

It also has the advantage that asymmetry of flows can be shown by swapping O space and D space in the tree, thus meeting requirements 6 and 7 above. In HiVE, this is a move to `sHier(/, $destination, $origin)` through a swap operation denoted as `oSwap(/, 1, 2)`. Dynamically swapping O space and D space provides a visual indication of flow asymmetry that can be identified without the need for extra symbolisation to represent flow direction (e.g. Rae, 2009).

In addition to the removal of the line occlusion problem, a cell-based representation of OD patterns allows us to colour each OD cell according to some attribute of the aggregated trajectories between a pair of origin and destination locations. The obvious symbolisation is to colour according to total flows, but other attributes may be represented such as difference maps or the signed chi statistic (Census Research Unit, 1980; Dykes and Brunson, 2007; Wood *et al.*, 2007) when comparing actual flow magnitudes with those expected based on some underlying model.

Simulated OD vectors

To test the ability of the OD map to discriminate between different structures of geographic vectors, various simulated flow sets were created. In particular, the origin location

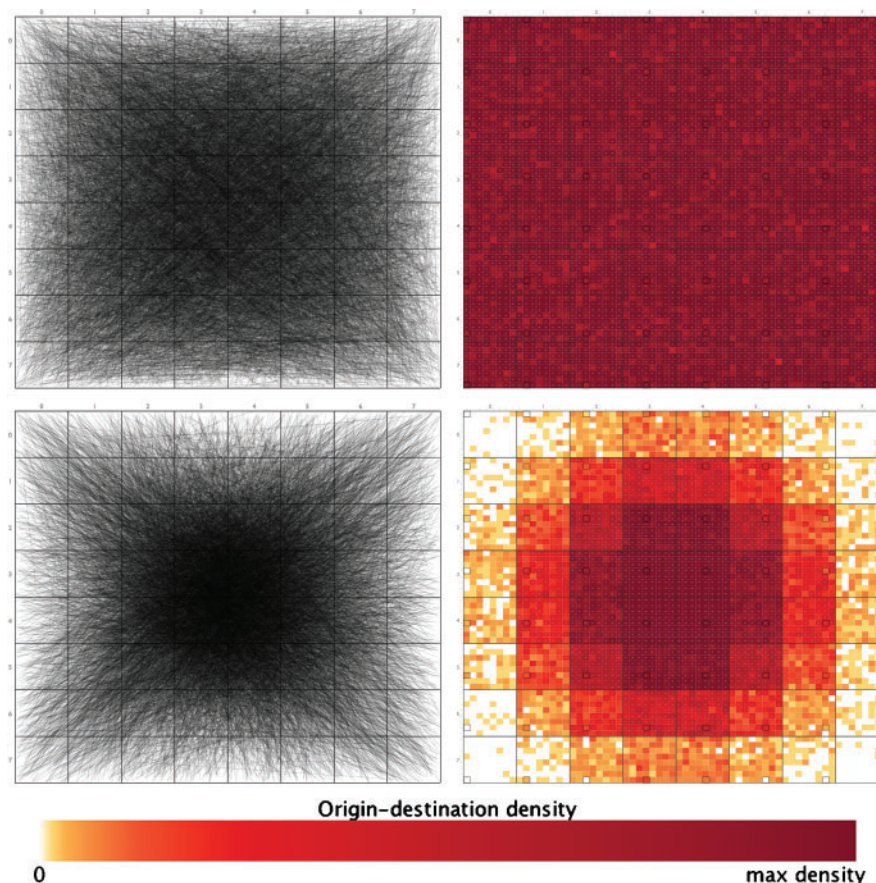


Figure 2. 100,000 simulated trajectories shown as a vector flow map (left) and an OD map (right). Destination locations are uniformly randomly distributed. In the upper example, origin locations are also uniformly random, while in the lower example, origin locations have a Gaussian random distribution about the centre. In all cases, the OD map colour uses a Brewer 'YlOrBr' colour scheme (Brewer, 2002) exponentially scaled between 0 and the maximum OD density

distribution, the destination location distribution, the length of and direction of flow were randomly generated under a range of assumptions. The aim was to see whether the OD map was adequately discriminating between different forms of flow distribution in large datasets (requirements 1 and 2 above); to identify any visual artefacts produced by the OD map or conventional flow map (requirement 8).

In the first set of simulations, 100,000 vectors were generated each with a uniformly distributed random origin and destination (Figure 2 top row). They were represented as a conventional flow map (Figure 2 left) using alpha blending to reveal the density of flow vectors. The same vectors were shown as an OD map (Figure 2 right). As expected, the OD map indicates an approximately uniform density of origin-destination cells. In contrast, the flow map suggests an apparent increase in density away from the edges. This is purely an artefact of the probability of line overlap, and not a true variation in OD density. For studies where the geometric path between origin and destination is not considered, it is important to appreciate that a conventional flow map or flow density surface (e.g. Rae, 2009) can produce such arbitrary variations in line density that do not reflect the true distributions of OD locations.

To distinguish arbitrary edge effects from true variations in OD density, a second set of 100,000 trajectories was

generated, again with a uniformly random set of destination locations, but containing a Gaussian distribution of origin locations around a centre point and standard deviation of 20% of the map width (Figure 2 bottom row). In this case, the concentration towards the centre is clearly evident in both the flow map and OD map. The OD map further reveals the uniform random nature of the destination cells as the approximately homogeneous density in each large grid square is apparent in this representation. This would not be detectable from the flow map directly.

Many real geographic flows have a tendency to show spatial autocorrelation (Cliff and Ord, 1973), i.e. shorter flows between origins and destinations are more likely than longer flows. To simulate this effect, a further set of simulations was created where each destination was generated in a uniformly random direction and Gaussian distance (with standard deviation of 20% of the map width) from each origin location (Figure 3). For both a uniform (Figure 3 top row) and Gaussian (Figure 3 bottom row) distribution of origins, the OD map representation shows the Gaussian flow length distribution clearly. The flow maps shown on the left of the figure, while distinguishable from each other, do not clearly reveal the positive spatial autocorrelation of the trajectories, nor do they strongly distinguish themselves from their uncorrelated equivalents in Figure 2.

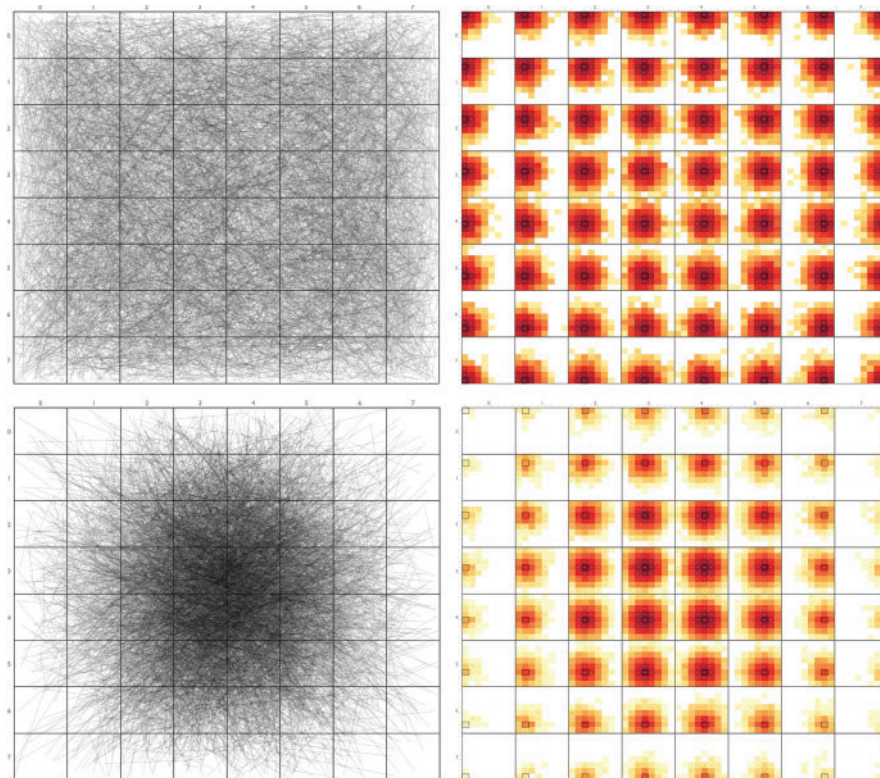


Figure 3. 100,000 simulated trajectories shown as a vector flow map (left) and an OD map (right). In the upper example, origin locations are uniformly random, while in the lower example origin locations have a Gaussian random distribution about the centre. In both cases, destination locations are a random direction with Gaussian random distance from the origin

To investigate the ability to reveal directional and distance bias, a final set of simulations was created where flows from left to right were excluded (Figure 4). This represents an extreme case of bias, but provides a useful benchmark with which to compare output. In Figure 4 upper row, the length of each flow is Gaussian (with standard deviation of 20% of the map width). Here the directional bias is visible only around the margins of the flow map where densities are sufficiently low. In contrast, the OD map shows the bias as a clear edge in each origin grid cell. In Figure 4 lower row, the length of each flow is fixed at 25% of the map width, thus precluding short flows. Visually, this flow map is indistinguishable from the Gaussian distribution, but the OD map clearly shows ‘hollow’ OD densities where close origin–destination flows are not permitted. It is important that the visualization is able to reveal these types of structure as many real geographic vectors will possess a tendency to have a minimum length (e.g. vacation trips or courier deliveries) or directional bias (e.g. animal migration).

It is recognized that there are many other structures that might exist in real geographic OD interactions (e.g. polycentric commuting patterns, impact of geographic barriers to movement, etc.) that are not tested with these simulations. The case study (see the section on ‘CASE STUDY’) demonstrates how such structures may be identified, and the interaction enabled by our OD map software (see the section on ‘Interaction’) allows their characterisation to be explored.

Rasterisation artefacts

Aggregation is often necessary when analysing large datasets and geography provides a meaningful framework for doing so. This enables the exploration of large datasets while identifying trends and structure in the data (Andrienko and Andrienko, 2008). However, any process that involves geographic aggregation can lose information that may have some value. Equally it is important to recognize that any trends or structure revealed could be an artefact of the aggregation undertaken rather than a property of the data. In particular, if the spacing between origin or destination locations is at approximately the same scale as the OD grid spacing, aliasing effects can be introduced. Figure 5 shows a flow map for travel-to-work flows for the US counties of Ohio. In this dataset, all home and work locations are aggregated to the county/counties in which they occur. In turn, these locations are mapped to the county centroid. Dividing the state into an 8×8 grid yields cells of approximately the same size as the counties. In most cases, the distribution of counties results in one centroid per grid cell, but in some cases, two centroids can coincide with a cell (e.g. Union and Logan in cell [3,2]), or even three (e.g. Geauga, Portage and Summit in cell [1,6]). The resulting OD map may therefore show higher OD densities not necessarily due to a regional control, but simply the coincidence of data points and arbitrary grid boundaries. This is one type of example of the modifiable area unit problem (MAUP) (Openshaw, 1984). Dynamic OD maps allow us to explore the sensitivities and effects of the MAUP

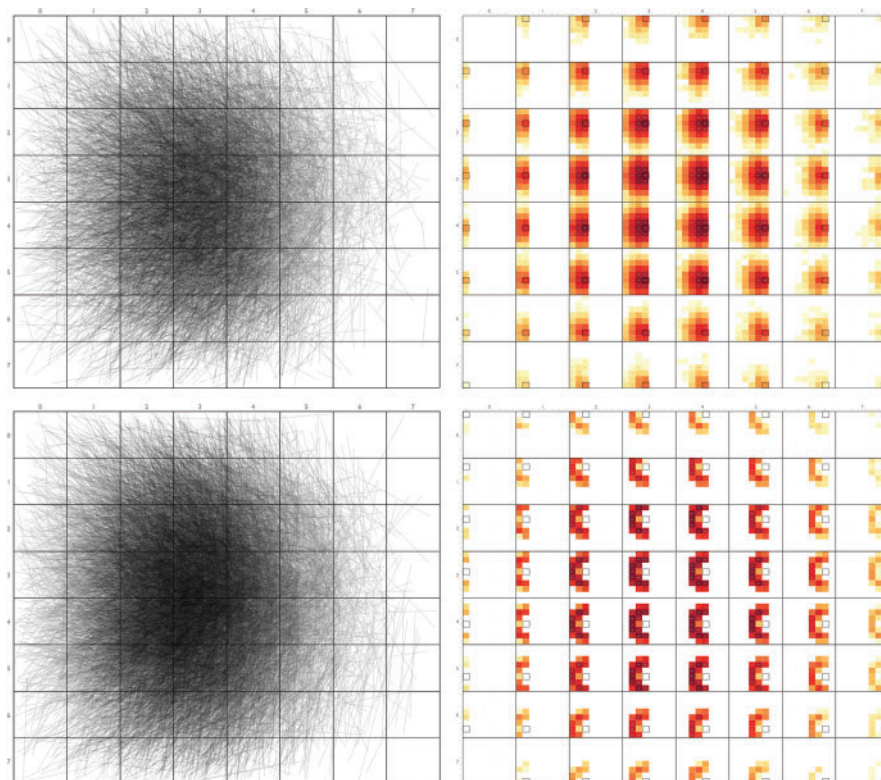


Figure 4. 100,000 simulated trajectories shown as a vector flow map (left) and an OD map (right). Origin locations have a Gaussian random distribution about the centre and a directional bias that prevents any left-to-right flows. In the upper example, length of the trajectory is Gaussian, while in the lower example, length of the flow is fixed at 25% of the map width (two grid cells)

and we have implemented two interactive approaches to do this.

The first is to permit dynamic variation of the number and location of the grid cells used to aggregate the data. This allows a visual spatial sensitivity analysis to be undertaken where the user interactively changes the grid aggregation to see if this has any significant effect on the trends evident in the OD map. A sample set of OD map excerpts with differing grid aggregations is shown in Figure 6. Offsetting the grid location (Figure 6 top row) has relatively little effect on the flow trends in this example, but changing the grid resolution (Figure 6 bottom row) has a more significant impact. In particular, increasing the number of grid cells also increases the number of blank origin and destination cells as fewer county centroids coincide with the smaller grid cells. This suggests that selecting an appropriate grid size is important if origins or destinations are clustered in space.

The choice of appropriate grid size can be facilitated through interactive control over grid size and location. In our optimized implementation (see the section on ‘Optimisation with hash grids’ below), immediate visual feedback is given on the effect of grid changes for OD maps with up to order 10^6 OD vectors.

An alternative solution to the aliasing problem can be applied when grid cells are approximately the same size as the spatial units under investigation. Here we tessellate the county centroids to form a grid arrangement. This is in effect a special case of the spatial treemap (Wood and Dykes, 2008), where all centroids are forced to their nearest

unique grid location. Where spatial units have some meaning, this can produce a more interpretable OD map (Figure 7). This quasi space-preserving solution may be more spatially consistent than others proposed and thus has advantages over existing alternatives.

A spatial treemap may not result in a regular grid tessellation, and while this would still allow an OD map to be created, its cognitive plausibility would be reduced as each coarse grid cell could potentially be a rectangle of a different aspect ratio. As a result, the destination cells within each origin cell would be subject to an inconsistent scaling. To overcome this problem, we ensure that the number of tessellated grid cells is a perfect square, by if necessary, adding some blank ‘dummy’ grid cells where no flows occur. The location of these cells is selected at the points furthest away from any known origin or destination cells. Typically this will be around the edge of non-rectangular study regions (see bottom corners and top central portion of the spatial treemap shown in Figure 7).

The disadvantage of regular grid tessellation is that for geographic regions with elongated aspect ratios, the nested cells become even more elongated (the aspect ratio is squared). Geographic integrity is maintained, but at the cost of very thin cells that can be difficult to interpret when coloured. This can be overcome by either reprojecting geographical space to give a squarer aspect ratio or by inserting dummy cells along the ‘thin’ edges of each cell. The former approach has the cost of transforming geographic space to something that may be unfamiliar to users (e.g. Chile projected to a square), while the latter can



Figure 5. Ohio travel-to-work flows showing county aliasing. Flows are located at each county's centroid. Where the scale of the county is approximately equal to the scale of the OD grid, aliasing effects can occur

result in inefficient use of graphical space (dummy cells repeated for each nested cell). A balance between geographic familiarity, cell interpretability and space efficiency must be struck by the analyst when constructing an OD map space.

Optimisation with Hash Grids

The OD map provides a visual interface for filtering of a set of vectors or trajectories. By selecting any given cell in the OD map, a query may be made of just those trajectories that originate from the given origin cell and end in the given

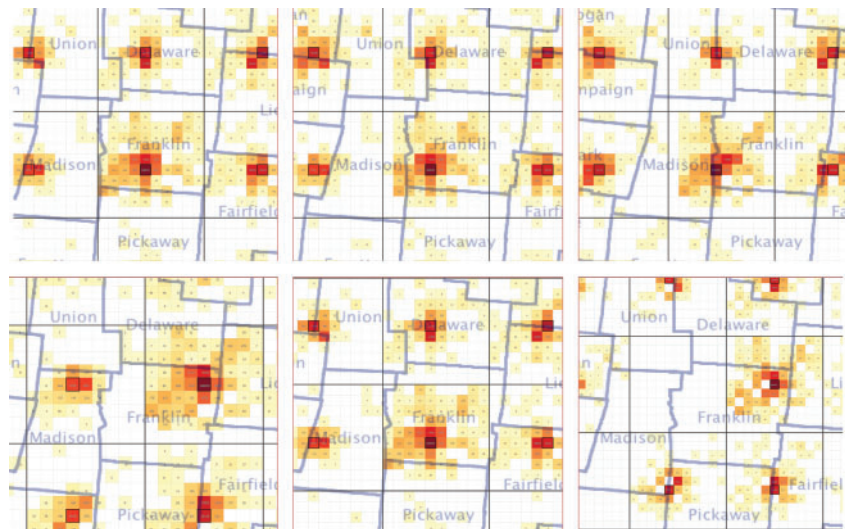


Figure 6. Effect of changing grid resolution and position on OD map densities of travel to work flows from Franklin county Ohio. Top row shows effect of moving the 10×10 grid in an east-west direction. Bottom row shows the effect of changing grid resolution from 9×9 (left) through 10×10 (middle) and 11×11 (right)

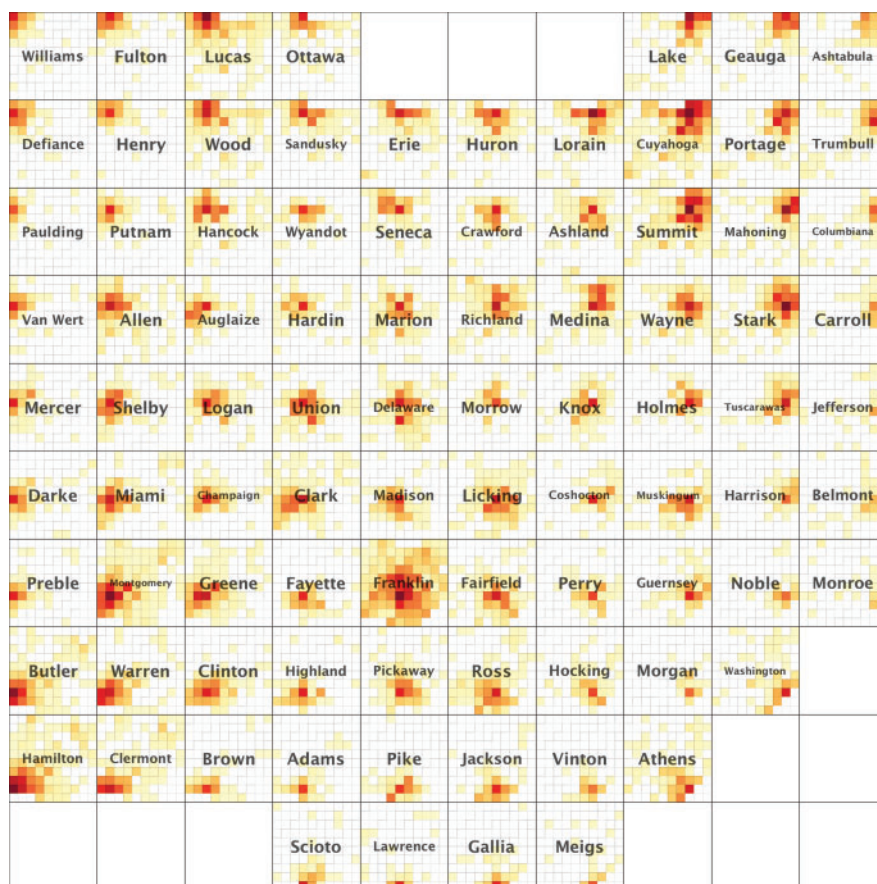


Figure 7. Ohio travel-to-work OD map. Here counties are tessellated into a grid using a spatial treemap (one county per grid cell plus 12 'dummy' grid cells)

destination cell implied by the OD cell location. Since the OD map itself retains the geographic projection of the original data, such a query could be used to project the full geographic path of the selected trajectories over the OD map by plotting lines or polylines. To allow this to happen interactively, and to facilitate rapid brushing over the OD map, an efficient data structure is required to store the set of 10^5 – 10^6 OD trajectories. Here we use the spatial hash grid, more commonly deployed for rapid collision detection (Eitz and Lixu, 2007).

The hash grid divides space into a regular grid, where each grid cell is accessible via a one-dimensional hash code. Each cell then stores a collection of references to spatial objects associated with the region within the grid cell boundaries. This has an obvious mapping to the OD map that also uses a regular gridding of space. In this case, we construct two hash grids, one that stores references to all the trajectories that originate from any given cell, and the other that stores references to the trajectories that end in each cell. The cell hash value is easily constructed in such a coarse two-dimensional grid as

$$\text{hash} = \text{row} \times n + \text{col} \quad (4)$$

where *row* and *col* are the OD map coarse row and column values and *n* is the number of rows or columns in the OD map. Given that *n* is likely to be of order 10, there is no

danger of overflow in the hash value if stored using at least 16 bit integers.

To select the set of trajectories between given origin and destination cells, the origin hash grid is first queried to retrieve only those trajectories that originate from the given cell. The destination hash grid is likewise queried to find only those trajectories that end in the given cell. Finally, the intersection of these two sets filters all but those vectors between the queried origin and destination cells. If the collections within each hash grid cell in turn are sorted using an optimisation structure (implemented here using Java's TreeSet structure), queries are sufficiently fast to allow interactive brushing of trajectory sets of order 10^5 – 10^6 trajectories.

CASE STUDY

To explore the validity of OD mapping, we created a prototype environment for visual exploration that could show flow maps, OD maps and OD matrices, and provide interactive control and linking between these views. It was developed in Processing (<http://www.processing.org>; Reas and Fry, 2007; Fry, 2007), an extension to Java for rapid visually orientated application development.

We selected county to county migration flows for the conterminous USA from the US 2000 Census (US Cenus

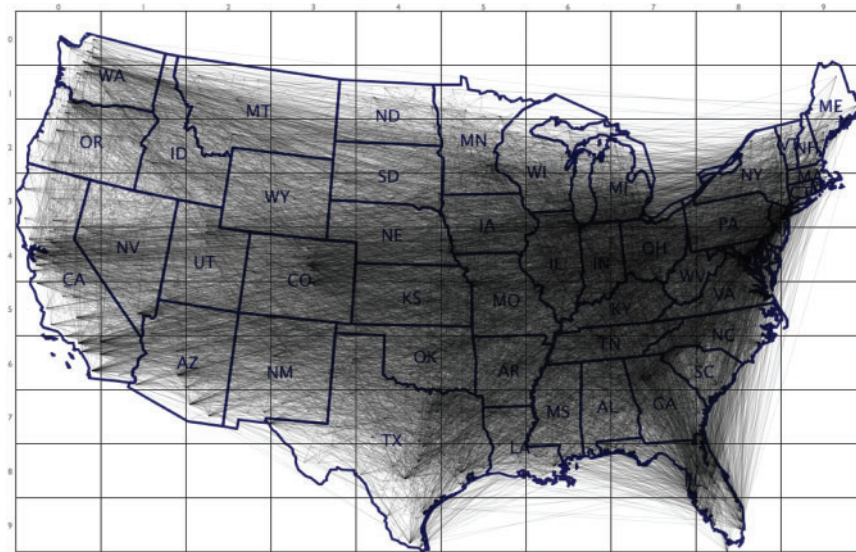


Figure 8. 20,000 US county–county migration vectors (3% random sample). Vectors rendered using transparency and anti-aliasing to allow ‘occlusion density’ to be seen

Bureau, 2008). This dataset recorded the numbers of changes in home address between 1995 and 2000 aggregated to the county of origin and destination. Locations were added by combining the dataset with the 2000 census county gazetteer and projecting from latitude/longitude to an Albers Equal Area projection. In total 721,432 separate migration vectors representing the inter-county movements of 46.6 million people were recorded. These vectors are shown as a conventional flow map in Figure 8. Evident in this view of the data is the ‘population footprint’ showing some of the major population centres, most clearly where they contrast with regions with fewer migration paths. However, from the experiments shown in the section on ‘Simulated OD vectors’ above, caution must be exercised in interpretation of artefacts. For example, it is not always evident whether the higher flow line densities in

the central states are due to migrants at those locations or simply because they happen to be placed on the path between origins and destinations to the west and east.

In contrast, Figure 9 shows the OD map of the same data. As in the previous OD maps, each OD cell is coloured according to the absolute numbers of movements between the origin and destination using an exponentially scaled colour scheme, and the ‘home cell’ for each origin is highlighted. The spatial autocorrelation inherent in migration is evident in that the highest densities of migration are where origins and destinations are close (darker red cells clustered around the home cell). There appears to be significant migration along the west coast of the US, both within California (e.g. LA and Orange County) and further north in Oregon and Washington. Some grid cells appear to show a much more homogeneous set of destination cells

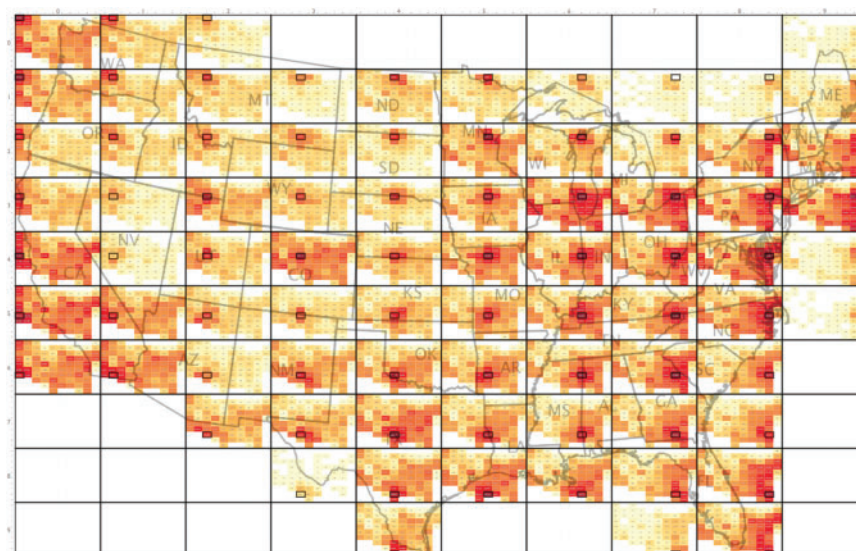


Figure 9. All 721,432 US county–county migration vectors shown as an OD map. Each large grid cell represents origin location, within which the map of destination densities using the same grid is shown

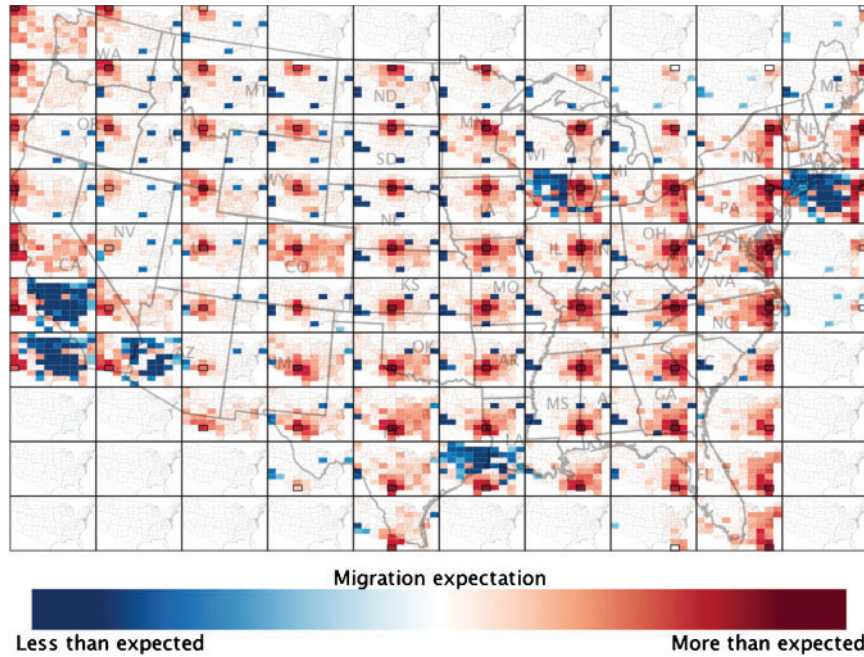


Figure 10. Chi statistic OD map showing the difference between observed county-county migration and that expected based on the populations of origin and destination counties (100% sample). Geographical context is provided by both a large origin map of US states and small multiples of destination maps. Chi values are coloured using an exponentially scaled diverging Brewer ‘RdBu11’ scheme

than others. For example, Southern California (0,6) and Colorado (3,5) have destinations evenly spread, while New York and Florida show much more concentrated (and reciprocal) flows. The non-coastal northwest states show some inter-state migration between them, but relatively little to other large parts of the USA.

The grid cells containing large urban populations (e.g. Southern California, Chicago and New York) inevitably show higher migration across the USA, but caution should be exercised when interpreting these patterns, especially with an exponential colour scale.

The signed chi statistic

Analysis of Figure 9 demonstrates that it is not always possible to separate high frequency of migration from the underlying population footprint. It is inevitable that on the whole areas of higher population density will have more migration, simply because there are more people available to move. The cell-based symbolisation of the OD map allows us to substitute more discriminating measures than absolute counts, such as the signed chi statistic for comparing observed with expected values

$$\chi = \frac{obs - exp}{\sqrt{(exp)}} \quad (5)$$

In this case, we define the observed value as the numbers of migrations from any given origin to a given destination and the expected value is weighted according to the mean population of the observed and expected counties

$$exp_{od} = \frac{\sum m \ pop_o + pop_d}{\sum pop \ 2(n-1)} \quad (6)$$

where $\sum m$ is the total number of people who have moved from one county to another, $\sum pop$ is the total population of all counties, pop_o and pop_d are the populations of origin and destination counties, and n is the total number of counties. In other words, this particular measure of expectation assumes that people migrate to all other counties in proportion to their respective populations. This simple model makes one of a number of different possible assumptions about expectation, but serves to illustrate how the OD map can be used to summarize such statistical measures. Other more sophisticated measures that could be quantified using the chi statistic might incorporate socioeconomic status, geographical permeability or trends over time. The results of mapping the population-based chi value is shown in Figure 10.

The chi OD map confirms that there is indeed greater than expected migration along the west coast (dark red destination cells). Cells that contain significant numbers of darker blue cells show greater ‘selectivity’ in migration destination given the size of the origin populations. The OD resolution and spatial origin were varied interactively in order to examine the persistence of patterns with scale and aggregation. It is apparent that there is less migration from Southern California to most of the USA, with the exception of the Pacific coast, Chicago, the large cities of Texas and New York. Similar patches of blue can be seen at these cities, suggesting a population less inclined to migrate away from big cities than our simple model suggests. New York, Chicago and Southern California show a relative lack of east-west migration, whereas Houston shows a resistance for north-south movement. Dark red cells that are not close to the home cell indicate larger than expected movements between geographically distant locations. For example, there are greater than expected flows from the NE coast to Florida and from Florida to Colorado.

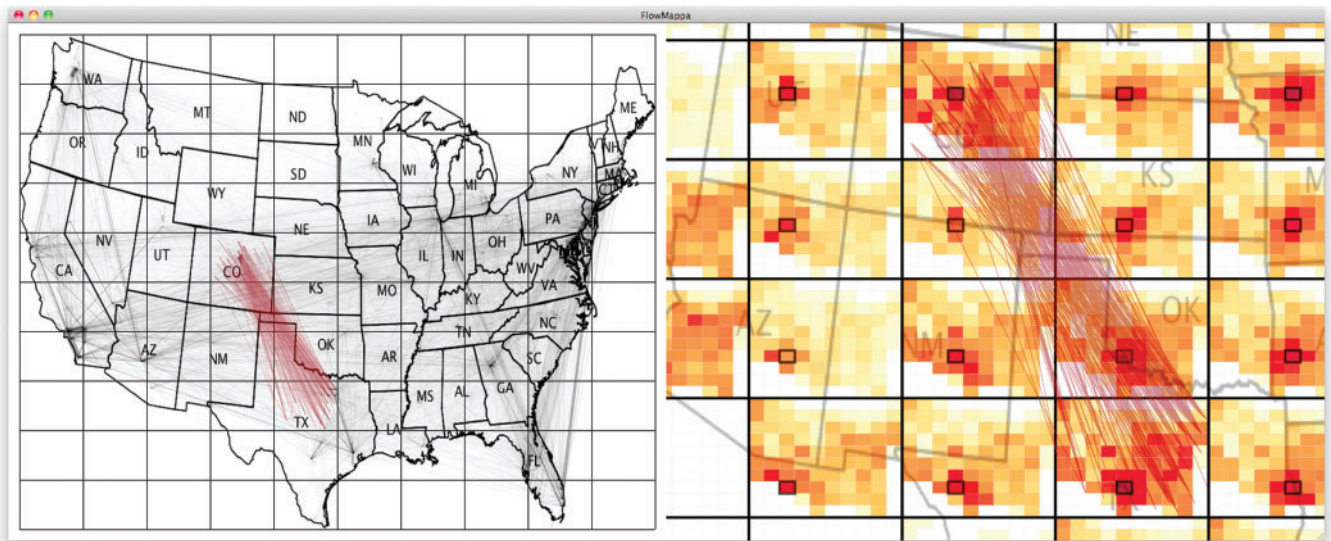


Figure 11. Example of interaction query and linked views. The application shows a flow map view on the left hand side and an OD map view on the right. Both can be independently zoomed and panned allowing both overview and detail-on-demand. By brushing over the various destination cells in the North Texas origin cell, the trajectories between the locations are overlaid on both views. Trajectories are stored at the resolution of the county centroids

Interaction

Viewing static OD maps provides some insight into the structure of the geographic vectors under investigation, particularly in overview. Adding interactive features to the prototype supports the exploratory process in a number of ways. Various interactions are possible (Yi and Stasko, 2007) and provide access to alternative layouts (Slingsby *et al.*, 2009), transformations between them (Heer and Robertson, 2007) and details to be accessed on demand (Shneiderman, 1996).

We found the following interactions useful in visually exploring spatial interactions in a number of datasets and have implemented them in our flowMappa environment for visual exploration (<http://www.flowmappa.com>):

- zooming and panning;
- toggling of numeric indicators of OD densities;
- toggling of context maps in both O space and D space;
- dynamic changing of grid cell size and offset;
- ability to swap O space for D space and *vice versa*;
- brushing to overlay selected OD vectors;
- linked views between flow map, OD map and OD matrix;
- varying colour scheme between ColorBrewer alternatives;
- varying colour scaling between log and linear scales.

Some selected examples of the effects of this interactive control are shown in Figure 11. While in this example, selected OD vectors are shown as straight lines, the projection into geographic space would allow the full geometry of the trajectories between selected origins and destinations to be overplotted.

CONCLUSIONS

Revisiting our list of requirements (see the section on ‘REQUIREMENTS AND PRIOR WORK’), it is apparent that none of the existing techniques for showing large sets of

trajectories are able to meet all eight of these. The OD map offers several advantages over the more commonly used flow mapping and OD matrix representations. Owing to aggregation of OD vectors into a regular grid, the OD map is scalable to large collections of vectors (requirement 1). The cost of this aggregation is twofold. First, like any form of aggregation, a potential loss of detail in origin and destination location results as the geographic grid resolution of the OD map is limited to geographic grids 20×20 or so (each cell having to contain a further 20×20 cells). This limit can be partially overcome by interactive zooming to reveal detail on demand (requirement 2). Second, aggregation into grid cells that do not reflect the underlying geographic structure can give rise to aliasing effects. These can be partially overcome through the use of a spatial treemap to aggregate into meaningful geographic units along with dummy cells to preserve key geographic properties.

For analysis where the vector between a pair of two locations has greater importance than the geometric path between them, the OD map provides this detail with minimal loss of information and little visual clutter (requirements 3 and 5). Where the geometry of the path is important, such as transportation infrastructure management, the geographic projection of the OD map allows this geometry to be overlaid (requirement 4). This works well for spaces with relatively square aspect ratios, but would be more problematic when nesting long thin regions. Methods for transforming geographic spaces, such as the spatial treemap algorithm used in Figure 7, provide some solutions but with some loss of spatial coherence and possible impacts on cognitive load.

All forms of trajectory mapping are liable to visualisation artefacts that do not reflect true properties of the data. Despite their common use, flow maps of large collections of vectors seem particularly vulnerable to this. Occlusion effects are removed by the OD map, although the partitioning of space into a grid introduces possible aliasing

and MAUP artefacts. The effect of these on the stability of the visualisation can be explored though dynamic change in gridding parameters (requirement 8). If the resolution of the vector data is much finer than the grid resolution, this effect is minimal, but aggregation is greater. Where grid resolution is closer to the data resolution, the spatial treemap provides one way of removing aliasing effects, but at the cost of some spatial distortion.

The asymmetry between flows in both directions between pairs of points is explicit in the OD map (requirement 6). It can be further explored by interactive swapping of O space and D space as well as through brushing over OD cells (requirement 7).

We do not propose the OD map as replacement for other forms of trajectory exploration, but suggest that it offers a new spatial view of large collections of geographic vectors that may be integrated with existing systems to help reveal and consider the geography of associations between pairs of locations. As the cells in the OD map are identical to those of the OD matrix, it provides a supplementary spatial ordering of this much-used aspatial representation of geographic information to which spatial cognition can be applied as we endeavour to explore geographic interactions and processes.

BIOGRAPHICAL NOTES



Jo Wood is a Reader in Geographic Information at the giCentre, City University London, with interests in new methods for visualisation of information with a spatial component. Jason Dykes is a Senior Lecturer in Geographic Information at the giCentre with interests in developing and using interactive graphics for geovisualisation. Aidan Slingsby is a Willis Research Fellow at the giCentre with

interests in spatial modelling and geovisualisation for applications relevant to the global insurance industry.

REFERENCES

- Andrienko, G. and Andrienko, N. (2008). 'Spatio-temporal Aggregation for Visual Analysis of Movements', in *IEEE Symposium on Visual Analytics Science and Technology (VAST 2008)*, pp. 51–58, Columbus, OH, Oct 19–24.
- Andrienko, G., Andrienko, N., Dykes, J., Fabrikant, S. and Wachowicz, M. (2008). 'Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research', *Information Visualization*, 7, pp. 173–180.
- Bertin, J. (1983). *Semiology of Graphics*, University of Wisconsin Press, Madison, WI.
- Brewer, C. (2002). Selecting Good Color Schemes for Maps, <http://www.colorbrewer.org>
- Census Research Unit (1980). *People in Britain: A Census Atlas*, HMSO, London.
- Chiricota, Y., Melançon, G., Quang, T. T. P. and Tissandier, P. (2008). 'Visual Exploration of (French) Commuter Networks', in *Geovisualization of Dynamics, Movement, and Change AGILE'08 Satellite Workshop*, Girona, May 5.
- Cliff, A. and Ord, J. (1973). *Spatial Autocorrelation*, Pion, London.
- Cui, W., Zhou, H., Qu, H., Wong, P. C. and Li, X. (2008). 'Geometry-based edge clustering for graph visualization', *Transactions on Visualization and Computer Graphics*, 14, pp. 1227–1284.
- Dykes, J. and Brunsdon, C. (2007). 'Geographically weighted visualization – interactive graphics for scale-varying exploratory analysis', *IEEE Transactions on Visualization and Computer Graphics*, 13, pp. 1161–1168.
- Eitz, M. and Lixu, G. (2007). 'Hierarchical Spatial Hashing for Real-time Collision Detection', in *IEEE International Conference on Shape Modeling and Applications (SMI 2007)*, pp. 61–70, Lyon, Jun 14–16.
- Fry, B. (2007). *Visualizing Data*, O'Reilly, Cambridge.
- Ghoniem, M., Fekete, J. and Castagliola, P. (2004). 'A Comparison of the Readability of Graphs Using Node-link and Matrix-based Representations', in *10th IEEE Symposium on Information Visualization (Infovis 2004)*, pp. 17–24, Austin, TX, Oct 19–21.
- Gilbert, M., Mitchell, A., Bourn, D., Mawdsley, J., Clifton-Hadley, R. and Wint, W. (2005). 'Cattle movements and bovine tuberculosis in Great Britain', *Nature*, 435, pp. 491–496.
- Guo, D. (2007). 'Visual analytics of spatial interaction patterns for pandemic decision support', *International Journal of Geographic Information Science*, 21, pp. 859–877.
- Guo, D. (2009). 'Flow mapping and multivariate visualization of large spatial interaction data', *IEEE Transactions on Visualization and Computer Graphics*, 15, pp. 1041–1048.
- Guo, D., Chen, J., MacEachren, A. and Liao, K. (2006). 'A visualization system for space-time and multivariate patterns (VIS-STAMP)', *IEEE Transactions on Visualization and Computer Graphics*, 12, pp. 1461–1474.
- Guo, D. and Gahegan, M. (2006). 'Spatial ordering and encoding for geographic data mining and visualization', *Journal of Intelligent Information Systems*, 27, pp. 243–266.
- Heer, J. and Robertson, G. (2007). 'Animated transitions in statistical data graphics', *IEEE Transactions on Visualization and Computer Graphics*, 13, pp. 1240–1247.
- Hernandez, T. (2007). 'Enhancing retail location decision support: The development and application of geovisualization', *Journal of Retailing and Consumer Services*, 14, pp. 249–258.
- Holten, D. and van Wijk, J. (2009). 'Force-directed edge bundling for graph visualization', *Computer Graphics Forum*, 28, pp. 983–990.
- Jarvis, R. and Patrick, E. (1973). 'Clustering using a similarity measure based on shared near neighbours', *IEEE Transactions on Computers*, 22, pp. 1025–1034.
- Karypis, G. and Kumar, V. (2000). 'Multilevel k-way hypergraph partitioning', *VLSI Design*, 11, pp. 285–300.
- Marble, D., Gou, Z., Liu, L. and Saunders, J. (1997). 'Recent advances in the exploratory analysis of interregional flows in space and time', in *Innovations in GIS 4*, pp. 75–88, Taylor & Francis, London.
- Openshaw, S. (1984). 'The modifiable area unit problem', *Concepts and Techniques in Modern Geography*, 38, p. 41.
- Paci, R. and Usai, S. (2009). 'Knowledge flows across European regions', *The Annals of Regional Science*, 43, pp. 669–690.
- Phan, D., Ling, X., Yeh, R. and Hanrahan, P. (2005). 'Flow Map Layout', in *IEEE Symposium on Information Visualization (Infovis 2005)*, pp. 219–224, Minneapolis, MN, Oct 23–25.
- Radburn, R., Dykes, J. and Wood, J. (2009). 'vizLib: Developing Capacity for Exploratory Data Analysis in Local Government – Visualization of Library Customer Behaviour', in *GIS Research UK 17th Annual Conference*, pp. 381–387, Durham, Apr 1–3.
- Rae, A. (2009). 'From spatial interaction data to spatial interaction information? geovisualisation and spatial structures of migration from the 2001 UK census', *Computers, Environment and Urban Systems*, 33, pp. 161–178.
- Reas, C. and Fry, B. (2007). *Processing: A Programming Handbook for Visual Designers and Artists*, MIT Press, Cambridge.
- Shneiderman, B. (1992). 'Tree visualization with tree-maps: a 2-d space-filling approach', *ACM Transactions on Graphics*, 11, pp. 92–99.

- Shneiderman, B. (1996). 'The Eyes Have it: A Task by Data Type Taxonomy for Information Visualizations', in **IEEE Symposium on Visual Languages**, pp. 336–343, Boulder, CO, Sep 3–6.
- Skupin, A. and Fabrikant, S. (2003). 'Spatialization methods: a cartographic research agenda for non-geographic information visualization', **Cartography and Geographic Information Science**, 30, pp. 99–119.
- Slingsby, A., Dykes, J. and Wood, J. (2008). 'Using treemaps for variable selection in spatio-temporal visualization', **Information Visualization**, 7, pp. 210–224.
- Slingsby, A., Dykes, J. and Wood, J. (2009). 'Configuring hierarchical layouts to address research questions', **IEEE Transactions on Visualization and Computer Graphics**, 15, pp. 977–984.
- Tobler, W. (1987). 'Experiments in migration mapping by computer', **The American Cartographer**, 14, pp. 155–163.
- US Census Bureau. (2008). County to County Migration Flow Files, <http://www.census.gov/population/www/cen2000/citytoctyflow>
- Voorhees, A. (1955). 'A general theory of traffic movement', in **Institute of Traffic Engineers Past Presidents' Award Paper**, ITE, New Haven, Proceedings, Institute of Traffic Engineers, pp. 46–56.
- Wilkinson, L. (1979). 'Permuting a matrix to a simple pattern', in **Proceedings of the Statistical Computing Section**, pp. 409–412, American Statistical Association, Washington, DC.
- Wilkinson, L. and Friendly, M. (2009). 'The history of the cluster heat map', **The American Statistician**, 63, pp. 179–184.
- Wood, J. and Dykes, J. (2008). 'Spatially ordered treemaps', **IEEE Transactions on Visualization and Computer Graphics**, 14, pp. 1348–1355.
- Wood, J., Dykes, J., Slingsby, A. and Clarke, K. (2007). 'Interactive visual exploration of a large spatio-temporal dataset: reflections on a geovisualization mashup', **IEEE Transactions on Visualization and Computer Graphics**, 13, pp. 1176–1183.
- Yi, J. S. and Stasko, J. (2007). 'Toward a deeper understanding of the role of interaction in information visualization', **IEEE Transactions on Visualization and Computer Graphics**, 13, pp. 1224–1231.

Copyright of Cartographic Journal is the property of Maney Publishing and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.