Estimating secondary school catchment areas and the spatial equity of access

Alex D. Singleton\textsuperscript{a}, Paul A. Longley\textsuperscript{b,*}, Rebecca Allen\textsuperscript{c}, Oliver O’Brien\textsuperscript{b}

\textsuperscript{a}Department of Civic Design, University of Liverpool, The Gordon Stephenson Building, 74 Bedford Street South, Liverpool, L69 7QZ
\textsuperscript{b}Department of Geography and CASA, University College London, Pearson Building, Gower Street, London WC1E 6BT, United Kingdom
\textsuperscript{c}Department of Quantitative Social Science, Institute of Education University of London, 20 Bedford Way, London WC1H 0AL, United Kingdom

\textbf{A R T I C L E   I N F O}

Article history:
Received 12 February 2010
Received in revised form 21 September 2010
Accepted 26 September 2010
Available online 30 October 2010

Keywords:
GIS
Percent volume contour
Catchment analysis
Decision support system
Education
School

\textbf{A B S T R A C T}

Following the Educational Reform Act of 1988, families in England and Wales have been free to identify a preferred school for their children’s secondary education. However, as part of this open selection, the demand from parents opting to send their children to the best performing schools far outstrips the supply of available places at them, and consequently many schools ration places using entry criteria that favour those pupils domiciled close to the school. Through this geographic selection process, choice is spatially sorted and access to the best schools is often crucially dependent upon where parents live. After illustrating this problem, this paper develops an automated modelling technique that can be used to define and map school catchment areas based on the home locations of pupils attending every publicly funded school in England. It then develops this framework to create a web based decision support tool to aid parents seeking secondary school places.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Spatial assignment rules such as allocation to nearest or catchment area school have always played an important role in the assignment of pupils to UK secondary schools at the age of 11. However, the binding nature of these rules changed in 1988 with the Education Reform Act (ERA), which introduced a “quasi-market approach to the allocation of resources” (Bartlett, 1993, p. 125), giving parents greater powers as service consumers. In a general sense these changes were also designed to render educators more accountable to parents, and to stimulate a process of competition through which over-all educational standards would be raised (Ranson, 1994). Integral to these reforms was monitoring the apparent efficacy with which a National Curriculum was taught, and publication of results from annual assessment tests in league tables. In this way, parents were to be given free choice of school for their children rather than being allocated a place based on where they lived. For a detailed account of these changes, see Taylor (2001).

Today all Local Authorities (LA) in England operate an application procedure where parents can express a ranked preference for between three and six (depending on the LA) schools anywhere in England. There is much written on the processes which inform these choices (Ball & Vincent, 1998; Lacey & Reay, 2002), and indeed, the aggregate outcomes of choice strategies in terms of the social (Allen, 2007; Ball, 1993; Gewirtz, Ball, & Bowe, 1995) or ethnic (Harris, Johnston, & Burgess, 2007) stratification of schools or neighbourhoods. Once the parent has selected a set of schools the LA allocates prospective students into available places using pre-specified admissions criteria, with the exception of Academies, foundation and voluntary-aided schools which control their own admissions. A consequence of this choice mechanism is that there are inevitably more applications to popular schools, such as those with particularly high average attainment, than places at them. Thus, for example, Haberdashers’ Aske’s Hatcham College (now an Academy) in the London Borough of Lewisham receives approximately eleven applications for every place available. A variety of mechanisms have evolved to manage these rates of oversubscription, the most widely used of which is to favour pupils living in close proximity to an oversubscribed school.

When parents disagree with the allocation of school places the ERA gives them a right of appeal, and there are formal procedures that govern these challenges. The number of appellants averages approximately 8% of subsequent admissions in the state sector, although this is a misleading statistic insofar as some unsuccessful applicants subsequently choose private sector education for their children. The uneven 2006–2007 geography of appellants is shown in Fig. 1, and this illustrates that, locally, up to one third of eventual admissions may be the outcome of the appeals process. A variety of hypotheses have been raised in relation to this variability, including the number of applications, school capacity and diversity of
A common difficulty for parents when choosing a school is how to discern a realistic expectation of securing a place under proximity based admission criteria. This issue can be exasperating, and indeed form the grounds for successful appeal if the information given to parents by the LA is inaccurate or misleading. For example, Fig. 2 shows the “areas of primary responsibility” which Bristol LA used to define the catchment of an expensive new school facility that opened in 2007. Given that the locations of all state educated primary school pupils are known (and small area census aggregations could be used to model the distribution of the minority educated in the private sector), it should be straightforward to model the likely demand for a new facility. Yet the grey buffer shows the c. 1 km limit beyond which children without special circumstances were not offered places until after the LA’s appeal procedure was exhausted. When siting the school, the LA proved incapable of anticipating the geography of demand for places, and it proved impossible for parents to form a realistic view of their chances of securing a child’s place at the new facility. Additional measures thus became necessary to shore up the admissions procedures (BBC, 2008), including redefining the intended area of primary responsibility into two zones (dotted line) – the same de jure catchment over-all, but a smaller de facto catchment, which is unlikely to have been acceptable when the school was planned, in view of the stark under-provision of state schooling for children in the west of the area.

In this particular instance, the mismatch between parental expectations and reality could have been managed by competent modelling of likely demand for a new school using available demographic sources, including primary schools census data. This information might have informed a more intelligent school location in the first place, or a more realistic initial demarcation of the area within which parents had a realistic expectation of securing places once the decision had been made. Although in this particular case the school was new, the more general point is that parental choice needs to be informed with better information about de facto school catchment areas, in order to shape realistic expectations. However, to date, there is very little usable information in the public domain.

2. Dynamic modelling of school catchment areas

The Department for Children, Schools and Families (DCSF) collects a pupil level annual schools census each academic year. It records a number of attributes of individual students within English state schools, including residential addresses. These data are incorporated into the (English) National Pupil Database (NPD) that records attainment and provides a unique identifier for each pupil to enable linkages through different educational key stages. For purposes of the following analysis, unit postcode address data are extracted to georeference the residential addresses of pupils using the National Statistics Postcode Directory (NSPD).

Numerous methods have previously been used to assess catchment areas of schools independent of actual information about the location of pupils. These have included the use of radial buffers (Brunsdon, 2001; Conduit, Brookes, Bramley, & Fletcher, 1996), census geographies (Herbert & Thomas, 1998), autoregressive models (Martin & Atkinson, 2001) and location allocation techniques (Pearce, 2000). These approaches are usually based solely upon the locations of schools, sometimes augmented by averaged or modelled information of the likely distribution of pupils. We do not consider these methods here, instead focusing upon methods that utilise the actual distributions of pupils, now identifiable...
given the wider availability of NPD data (see Harland & Stillwell, 2007). Thus, for example, Harris et al. (2007) used a convex hull procedure to define catchment areas based on the spatial locations of the 50% nearest pupils resident to a school, measured by Euclidean distance. Fig. 3 illustrates a typical output from this procedure for a school in Leeds, shown as an angular polygon. Convex hull polygons can provide a useful delineation of the smallest area that encloses a set of points (de Smith, Goodchild, & Longley, 2009). However, they have a series of disadvantages when used to define catchments areas for schools. First, there are issues of disclosure control, given that the nodes of the polygon boundary identify the approximate residential locations of some pupils. This becomes a serious issue if this information is to be placed in the public domain as part of an information service. Second, the convex hull can be radically distorted by outliers, and therefore affecting the overall shape of the polygon. In Fig. 3 this is illustrated by the extension of the convex hull to the south of the school into an area of low population density (as indicated by lighter shading on the map) where there are few recorded pupils attending the school. These effects could potentially be limited by adjusting the percentage of pupils from a school included within the convex hull analysis. Although any catchment definition is vulnerable to such effects, the angular nature of the resulting catchment presents a counterintuitive representation that is not weighted in any way by the over-all distribution of points.

An alternative representation of school catchment areas can be derived using kernel density estimates (KDE) of the areas where a local population of interest is spatially concentrated. Percent volume contours (PVCs) may be used to bound pre-specified cumulative percentages of the densities of registered pupils for any given school. Thus, for example, PVCs might be used to identify the areas within which approximately 50%, 75% or 95% of the pupils live, and who attend a given school. The KDE technique does not divulge the point locations of any pupils, and the technique has been successfully used to maintain confidentiality in health research (Guagliardo, 2004; Petersen et al., 2009). KDE is specified using bandwidth and spatial resolution parameters for the output raster: the map shown in Fig. 4 was created by passing a kernel over the home locations of the pupils with a bandwidth of 500 m, and the output raster with a pixel size (spatial resolution) of 100 m. This map was created using the quadratic kernel function (de Smith et al., 2009) that is implemented in ArcGIS. The darker areas of the density surface in Fig. 4 indicate where more pupils are located. In some sense this creates a reasonable approximation of the school catchment, and indeed mirrors hotspot type visualisations that are commonly used in other domains such as crime analysis (Chainey & Ratcliffe, 2005). However, for modelling purposes it is often useful to have a simplified binary measure of the areas that the user can readily interpret as lying within or outside a de facto school catchment. As such, the PVC at 50% and 90% PVC were extracted from the density surface; thus, pertaining to the locations of around 50% and 90% of the pupils respectively. These PVCs are further illustrated in Fig. 5, where the undulating terrain relates to the distribution of pupils as extracted from the density surface in Fig. 4. The white line that delineates the densest part of the distribution of pupils is here defined as accommodating 35% of the

---

Fig. 2. The geography of realistic expectations in Redland Green, Bristol: areas of primary and secondary responsibility and the area (grey buffering) within which applicants without special circumstances secured places.
pupils attending the school. Any other convenient interval can be used as desired. Fig. 4 also shows the convex hull catchment extracted as using 50% of the pupils, and again it is clear that this catchment zone does not capture the core location or those pupils attending the school.

One obvious limitation of PVC based representations is the discontinuous nature of the catchment areas (see Fig. 5). In order to make the visualisation most intuitive, a representation of a catchment needs to be continuous, that is the catchment area must join up to create a single and unified boundary. Discontinuities occur in PVCs for two reasons. First, a PVC can be sensitive to outliers which do not necessarily fit the main geographic distribution of the pupils. Such outliers may manifest the effects of spatially clustered but non geographic admissions criteria, such as priority admission...
for siblings that were admitted at a time when the household lived closer to the school. A second issue effecting spatial contiguity of the PVC relates to the selection of an appropriate bandwidth for the KDE. Fig. 6 illustrates how, by maintaining a boundary enveloping the residences of 70% threshold of all pupils and specifying three different bandwidth parameters, it is possible to change the shape and contiguity of the output catchment area.

It would be a very time consuming process to manually create catchment areas for all schools given the numerous possible bandwidth and pupil percentage threshold combinations. Additionally, it would also be preferable to update the catchment areas annually, in response to changes in the residential composition of the school roll as manifest in successive PLASC releases. In order to accommodate these problems, a set of procedures were written in the statistical programming language R (www.r-project.org) which utilised a series of heuristics to generate catchment areas based on automatically selected pupil percentage thresholds and bandwidth sizes. R was chosen over scripting in ArcGIS as it was found to be significantly quicker at generating results and offered more flexibility for the specification of functions. The exact implementation in R that was developed used an unbounded, Normally distributed kernel with a fixed bandwidth which was initiated at 200 m. After initially passing this kernel over the pupil point data to generate a KDE surface, the algorithm attempted to draw a contiguous PVC with an initial 70% threshold value, i.e. a single zone which enclosed the residential locations of 70% of the pupils attending each school. The two initial threshold values (200 m and 70%) were used following discussions with education professionals and potential end users, in which these values emerged as being likely to demarcate the envelopes within which parents could have a realistic expectation of their children securing a place, where distance based criteria were paramount. A higher PVC value, say 95%, might be distorted by small numbers of pupils who had moved locally since gaining a place. The algorithm then proceeded to count the number of polygons created by the PVC procedure.

If only a single polygon was created, then the algorithm moved onto the next school. If multiple polygons were created, the KDE was run again on the same pupil locations with a new bandwidth that was increased by 100 m. A new 70% PVC was calculated and the polygon count test repeated.

This procedure continued using successively larger bandwidths until a maximum bandwidth of 3500 m was reached. At this point the threshold was reduced by 10% and the bandwidth reset to 200 m and the incremental procedure run again.

If the algorithm reached a minimum threshold (specified as 50%) and a single polygon still could not be created with a 3500 m bandwidth, then these values were used and the algorithm moved onto the next school. This effect was found most likely to occur in schools serving populations dispersed across very large

---

**Fig. 5.** KDE of the pupils within Carr Manor High School visualised as a 3D slope.

**Fig. 6.** Bandwidth tests for a 70% PVC.
geographic areas, most typically in rural locations. The KDE based algorithm is not sensitive to changes in underlying settlement structure, and as such, can produce multiple polygon catchment areas at the maximum bandwidth specification. It is envisaged that the algorithm could be adapted in the future to increase bandwidths for those schools that are located in low density or rural areas, and as such force the creation of contiguous catchment areas.

There may, however, be other explanations for the emergence of non-contiguous catchment areas where, for example, a school has been rebuilt in a different physical location. As such, the use of the default maximum bandwidth parameters may create misleading results. A range of possible diagnostic checks might be used instead: the school roll postcodes might be checked against some measure of urbanity such as those derived from national urban–rural classifications; or a density measure based on a fixed buffered radial distance from each school might be used to examine the population of students within the school’s immediate locale. We have not yet attempted these refinements for the school catchments that are modelled using the maximum values, but emphasise that these and other adaptive procedures could be used to ensure contiguous results for schools in all areas.

3. Online decision support

Having set out the automated process through which school catchment areas can be estimated using KDE and PVC, this section seeks to identify a suitable means of serving this information to parents seeking to make school choices for their children under the UK government’s parental choice agenda. This entails the creation of an online decision support tool.

Neogeography is a term that has recently re-emerged in academic parlance to describe the use of new web mapping techniques to create online interfaces which are intuitive and offer more fluid navigation of spatial data (Singleton, 2010). In previous work, ‘neogeography’ interfaces have been constructed using the commercial application programming interfaces (APIs) of the major web mapping suppliers (e.g. Google, Yahoo, Microsoft: see Giblin, Singleton, Mateos, & Longley, 2008), and although these interfaces can be straightforward to use, a number of disadvantages have emerged from their implementation. First, they lock the developed into a specific set of background data with fixed cartographic style, and as such this can create visualisations which may not be most appropriate to specific applications. Second, the APIs used to generate the web mapping are under constant development, and this creates potential maintenance problems when functionality is upgraded or otherwise changed. Third, the terms and conditions of some APIs when used to visualise third party data imply that the ownership of the data are transferred to the API provider. This is problematic in a UK context given the complexities of boundary data licensing or indeed the visualisation of other information derived from public sector sources. For these reasons, we describe here the development of a more flexible alternative using predominantly open source software and data (see Fig. 7).
The two main data formats which are visualised by this application are a series of raster map tiles which present a variety of open source contextual background data and the school catchment areas themselves, which are stored as vectors. The contextual data (e.g. the Digital Elevation Model) were designed to enhance the general appearance of the Website, and to foster ease of use. The data which were used to create the background raster tiles comprise a series of choropleth maps which were either derived from the NPD or were pre compiled third party classifications or indicators (see Table 1). All of these data were imported into ArcGIS and merged with appropriate Shapefile polygon boundaries.

OpenStreetMap (www.openstreetmap.org) data were extracted for the UK and used to create a neutral coloured street network overlay. All contextual data, including the street networks, were rendered from the Shapefiles using the Mapnik (www.mapnik.org) tile server. This process essentially creates a number of geographically extensive maps for each zoom level and then divides these up into small square images for display in the map interface. Mapnik enables the customisation of all elements of the cartography from colour and line style to label placement. Additionally, the visual appearance of the maps was improved using a terrain layer created using a Digital Elevation Model (DEM) provided by the NASA Shuttle Radar Topography Mission (SRTM). These data were rendered by importing into the Geographic Data Abstraction Library (GDAL) tools (www.gdal.org) which converted the SRTM DEM data into raster tiles. A further process was run on the tiles to create a hill shaded effect using PerryGeo tools (www.pemmy-geo.networdpress/?p=7).

The interactive display of the school catchment and contextual data was created through a customized interface built using OpenLayers (openlayers.org: see Fig. 8). This JavaScript library enables the display of a variety of spatial data in similar formats to those used in a commercial API. The interfaces have a “slippy map” (Crampton, 2009) usability akin to the commercial API and enable both panning and zooming. Further control and search functions were built with JavaScript to enable map layer visibility switching and the display of information boxes detailing a variety of contextual data about the school that are extracted from the pupil level data. These include the average school attainment and a series of profiles for IDACI, POLAR and OAC. Finally, the graphs which are generated on these information tabs were created using the Google Chart API (code.google.com/apis/chart/). An illustrative screenshot is shown in Fig. 8 and online at www.educationprofiler.org.

### Table 1

<table>
<thead>
<tr>
<th>Data</th>
<th>Source</th>
<th>Geography</th>
</tr>
</thead>
<tbody>
<tr>
<td>% 5 &amp; GCSE A–C grades</td>
<td>NPD/NSPD</td>
<td>Lower level Super Output Area (LSOA)</td>
</tr>
<tr>
<td>% Free school meals</td>
<td>NPD/NSPD</td>
<td>LSOA</td>
</tr>
<tr>
<td>Average A-level points</td>
<td>NPD/NSPD</td>
<td>LSOA</td>
</tr>
<tr>
<td>Income Deprivation Affecting Children Index (IDACI)</td>
<td>Communities and Local Government (CLG)</td>
<td>LSOA</td>
</tr>
<tr>
<td>Output Area Classification (OAC), Vickers and Rees (2007)</td>
<td>Office for National Statistics (ONS)</td>
<td>OA</td>
</tr>
</tbody>
</table>

The interactive display of the school catchment and contextual data was created through a customized interface built using OpenLayers (openlayers.org: see Fig. 8). This JavaScript library enables the display of a variety of spatial data in similar formats to those used in a commercial API. The interfaces have a “slippy map” (Crampton, 2009) usability akin to the commercial API and enable both panning and zooming. Further control and search functions were built with JavaScript to enable map layer visibility switching and the display of information boxes detailing a variety of contextual data about the school that are extracted from the pupil level data. These include the average school attainment and a series of profiles for IDACI, POLAR and OAC. Finally, the graphs which are generated on these information tabs were created using the Google Chart API (code.google.com/apis/chart/). An illustrative screenshot is shown in Fig. 8 and online at www.educationprofiler.org.

### 4. Concluding comments

Geographers have long been somewhat preoccupied with the application of location–allocation type techniques to problems of school catchment definition (Pearce, 2000; Sutcliffe & Board, 1986), yet it is difficult (and possibly invidious) to attempt to square the outcomes of such exercises with the revealed preference outcomes under the parental choice agenda. Moreover, it seems clear from the outputs of analysis such as the Bristol example in the introductory section that the de jure catchments, or areas

![Fig. 8. The beta interface for the school catchment profiler.](image-url)
of primary responsibility, may bear little correspondence with either public sector service plans, or the underlying preferences of parents. We have not dwelt at length on the specifics of school catchment definition in the UK, particularly as there is no uniformity of approach across the country. For example, the Bristol authority that provides the setting for one of our case studies is amongst the minority of jurisdictions that does delineate geographic ‘areas of primary responsibility’: other areas use any of a range of ad hoc procedures, including ballots and selective entrance examinations. The overarching methodological aim of this paper has been to demonstrate a new approach to the geographical representation of de facto school catchments in England, that may be used to give parents realistic expectations of the schools that are available for their children. The KDE/PVC method presented builds cumulatively on previous catchment estimation techniques using convex hulls with DCSF data. The decision support tool that we have developed represents a considerable improvement upon what has previously been available. The KDE/PVC method offers significant improvement in terms of disclosure control, which is essential if this information is to be placed in the public domain.

Kernel density estimation is a method of proven validity in a technical sense (de Smith et al., 2009), but our research to date falls short of a full and geographically extensive evaluation of its provenance for school catchment modelling throughout England. Such evaluation is by no means a trivial task. The processes by which parents secure places for their children in desirable UK schools are multifaceted and often intrinsically uncertain. For example: different local education authorities adopt different practices, ranging from strict catchment based analysis to lotteries; faith and language based schools enjoy considerable flexibility in setting their admissions criteria; and increasing numbers of other public sector schools are empowered to devise their own admissions criteria – for example, the new generation of so called ‘free’ schools with ‘academy’ status.

In view of local variations in school admissions criteria, and the sensitivity of the extent of school catchments to small numbers of pupils in particular years, we do not think that convex hulls or other comprehensive tessellations provide appropriate alternative solutions – since user interpretation is guided to infer distinct break points. Our percent volume contours seek to provide a clear statement of the de facto contours that bound 70% of the catchment areas of schools, with variable bandwidth accommodating the sparsity of settlement in rural areas. Informal consultations with education professionals familiar with different school settings leads us to hypothesise that these provide a useful delineation of the areas within which parents have a realistic expectation of securing a place for their children – subject of course to fulfilment of other (faith, linguistic, etc.) criteria. The emphasis upon ‘useful’ is cognisant that we are attempting an England wide representation, that is not sensitive to the specifics of local education policy.

Validating this hypothesis is non-trivial, because we only have data pertaining to choice outcomes, which may be different to parental preferences. Parental revealed preferences may themselves be grounded in the perceived likelihood of successful applications to specific schools. Moreover, our incremental adjustment of bandwidth in less populous areas is a pragmatic decision guided by the priorities of visual interpretation. Bearing all of this in mind, we contend that the contours are a successful attempt to achieve consistency and intelligibility. As such, they provide a useful ‘what is’ counterweight to the minefield of individual school and local authority admissions policy documents. Whilst we have not conducted any extensive usability studies, we believe that the results of the exercise provide a valuable picture of admissions outcomes across England, as well as an intelligible tool to guide parental decision making. It remains for detailed usability studies to investigate the value of this approach in different educational settings across the country.

The online visualisation interface for the catchment area data was enabled through a variety of open source libraries and tools, as well as novel use of public domain street mapping and topographic data. The coupling and customisation of this software enabled the creation of a user interface that displays the locations of catchments for secondary schools against a series of contextual choropleth data. To aid interpretation, road network data and labels were generated from OpenStreetMap data and are used as an overlay to the choropleth maps. A feature of our system is that the PVC level can change – so some catchment areas identify a different proportion of students to those delineated by others. In future work we will consider ways in which this could be made more apparent to users through graphical representation.

With distance based admissions criteria being the determining factor behind access of most pupils to the vast majority of state (publically-funded) schools in England, the availability of reliable information on the location of those students who currently attend schools is valuable for a variety of reasons. Most obviously, the catchment data provide a valuable resource for parents seeking to ascertain the schools that are in their de facto choice set under prevailing government policy, and enable them to shape their expectations accordingly. This information is of great potential interest to parents, particularly those contemplating residential relocation. Seen from the perspective of public administrators, the same information might usefully be used by local authority officials to align their published areas of ‘catchment’ boundaries with the actual allocations of pupils within (and without) their jurisdictions. This latter application might allow them to understand the true pattern or potential and actual demand for their services.

Acknowledgements

This work is based on data provided through EDINA UKBORDERS with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown. In addition, schools data were provided by the Department for Children Schools and Families. Street network data were provided by OpenStreetMap. This research was conducted under an ESRC First Grant (RES-061-25-0303) and the HEFCE Centre of Excellence in Teaching and Learning (CETL) Initiative (Spatial Literacy in Teaching: SPLiNT).

References


