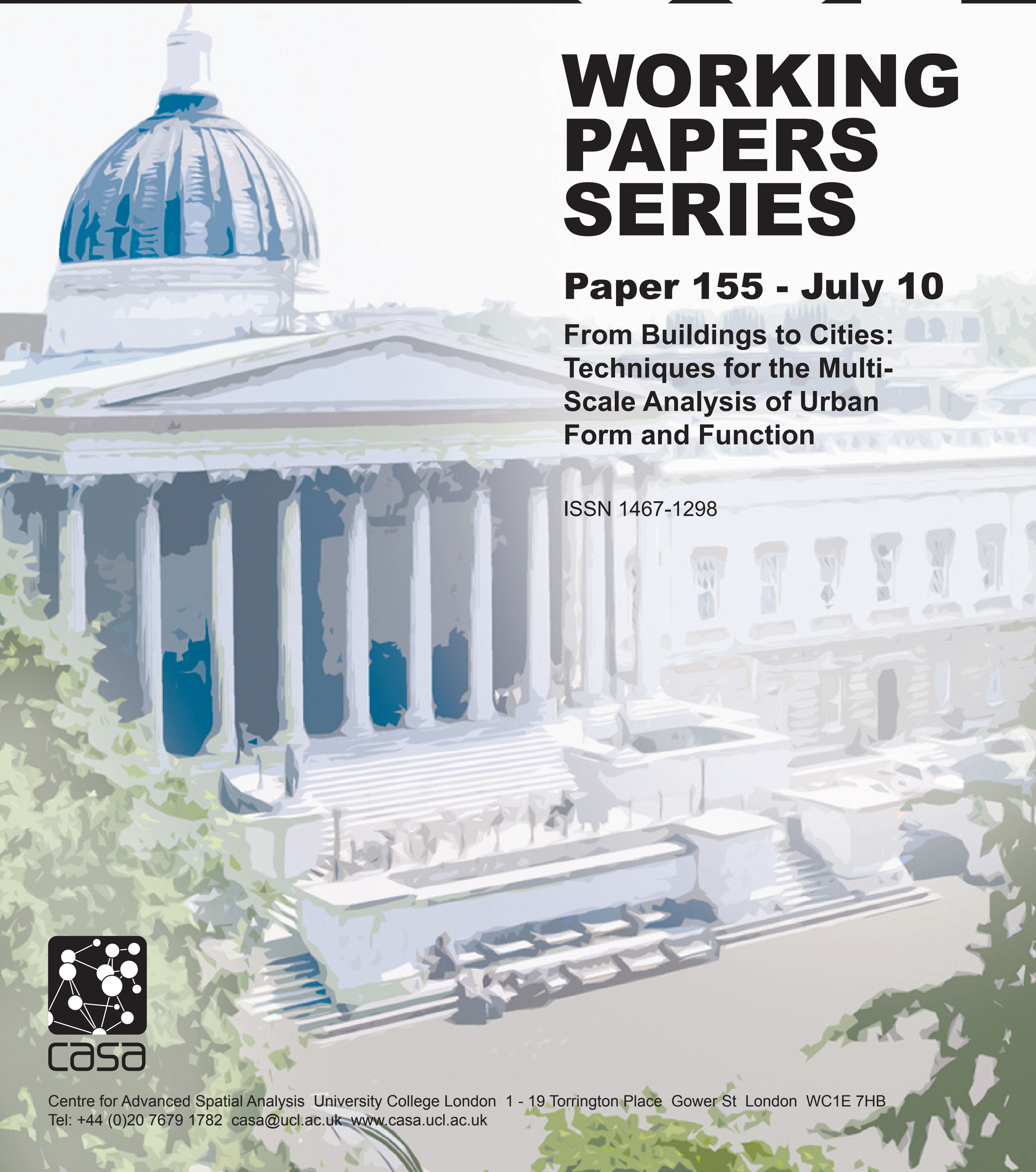
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL



WORKING PAPERS SERIES

Paper 155 - July 10

From Buildings to Cities: Techniques for the Multi-Scale Analysis of Urban Form and Function

ISSN 1467-1298

casa

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

From Buildings to Cities: Techniques for the Multi-Scale Analysis of Urban Form and Function

Duncan A. Smith and Andrew T. Crooks

duncan.a.smith@ucl.ac.uk, acrooks2@gmu.edu

Centre for Advanced Spatial Analysis

University College London, 1-19 Torrington Place, London, WC1E 7HB, England

Tel. +44(0)207 7679 1782 Fax +44(0)207 7813 2843

6th July 2010

Abstract

The built environment is a significant factor in many urban processes, yet direct measures of built form are seldom used in geographical studies. Representation and analysis of urban form and function could provide new insights and improve the evidence base for research. So far progress has been slow due to limited data availability, computational demands, and a lack of methods to integrate built environment data with aggregate geographical analysis. Spatial data and computational improvements are overcoming some of these problems, but there remains a need for techniques to process and aggregate urban form data. Here we develop a Built Environment Model of urban function and dwelling type classifications for Greater London, based on detailed topographic and address-based data (sourced from Ordnance Survey MasterMap). The multi-scale approach allows the Built Environment Model to be viewed at fine-scales for local planning contexts, and at city-wide scales for aggregate geographical analysis, allowing an improved understanding of urban processes. This flexibility is illustrated in the two examples, that of urban function and residential type analysis, where both local-scale urban clustering and city-wide trends in density and agglomeration are shown. While we demonstrate the multi-scale Built Environment Model to be a viable approach, a number of accuracy issues are identified, including the limitations of 2D data, inaccuracies in commercial function data and problems with temporal attribution. These limitations currently restrict the more advanced applications of the Built Environment Model.

Keywords: Urban Form, Function, Land Use, GIS, Housing, Residential Type, Dwelling Type, Multi-Scale, Visualisation, MasterMap, Address Layer 2, Greater London

Table of Contents

List of Figures

List of Tables

1: Introduction

We begin this paper by introducing the aim of integrating urban geographical analysis with iconic urban representations used in planning and 3D digital city modelling (Section 2). This includes a discussion of the importance of scale in geographical analysis and the advantages of a multi-scale approach (Section 2.3). Advances in urban data infrastructure that underlie this approach are then described in Section 3. The study area for this research is Greater London, and the aims of the research in the context of urban change and planning in London are outlined in Section 4. The subsequent sections of the paper cover more technical aspects of the research including the methods used to create the data model and classifications (Section 5) and validation of the model classification accuracy (Section 6). In the penultimate section, potential applications of the Built Environment Model are illustrated (Section 7), featuring visualisations of urban function and residential clustering at multiple scales for Greater London. Finally conclusions of the research are discussed in Section 8.

2: Integrating Geography and Geometry

Interactions between socio-economic processes and the built environment are relevant to many aspects of urban geographical research. For example, research into urban demographic and economic spatial structures link locational decision-making to buildings through property markets and land ownership. Many significant processes of change in cities such as gentrification and urban renewal involve specific transformations in urban form[1] (see Davidson and Lees, 2005). Urban sustainability is another research area strongly linked to the built environment, as key aspects of sustainability include the energy efficiency of buildings (see Steemers, 2003; Bruhns *et al.*, 2000), and the functional integration of urban activities to reduce travel distances (see Urban Task Force, 1999; Banister, 2005). Despite these relationships, the direct measurement and analysis of urban form and function is rather limited in geographical research, particularly for city-wide studies. Where urban form is considered in geographic research

[1] The term 'urban form' is used here to refer to all physical aspects of the city, its buildings, streets, and all other elements that make up the urban realm (Talen, 2003).

it is often through aggregate proxy measures such as population and employment density (Talen, 2003). While these measures are a useful starting point, they do not consider any physical properties of built form and leave many relationships unexplored. In this research we argue that the inclusion of urban form and function analysis can provide new insights and empirical grounding for a number of fields, particularly for urban planning and development research, property market and housing analysis, and urban sustainability studies.

There is also a contrast within the fields of architecture, planning and geography in the sense that architects and planners commonly use iconic representations of urban form such as geometric plans and physical models. This approach contrasts with geographical research in two aspects. Firstly the focus is on the physical properties of the built environment rather than the socio-economic interests of urban geography. Secondly the extent of study is generally restricted to buildings and localities, in contrast to the city-wide (or larger) scope of geographical studies. There is a growing interest in linking these *geographical* and *geometrical* approaches to provide an improved understanding of cities (Batty, 2007). Over the last decade there has been a continuing development of geographic information (GI) technologies and the emergence of rich fine-scale digital data sources (Longley, 2003). These new detailed datasets have enhanced spatial and attribute information, and are sufficiently *intensive* to analyse detailed form and function relationships and also sufficiently *extensive* to enable patterns to be generalised across entire metropolitan areas. It is increasingly possible to link the socio-economic focus of geographical analysis to the geometric built environment approach that is employed in local urban planning. Batty (2000; 2007) has termed this linking process '*geography to geometry*', the merging of iconic and symbolic urban models[2], and it opens up many possibilities for research.

[2] One of the most widely recognised classifications of urban models is that by Lowry, 1965, who defined models on a continuum between the iconic and the symbolic. Iconic models are physical versions of the 'real' thing, normally scaled down. Typical traditional examples include the architects' block model and 2D cartographic maps. Symbolic models represent systems in terms of the way they function, often through time and over space. Such models replace the physical or material system by some logical and/or

New data models and analysis techniques are required to achieve this goal. This paper is intended to further this research agenda: firstly by describing a methodology to combine geometric and socioeconomic datasets for a large city in a single spatial database; and secondly by implementing spatial analysis techniques to provide data and indicators for urban research: principally urban function and residential dwelling type. Fine-scale relationships between urban form and function can then be explored to provide an evidence base for research topics such as urban sustainability, residential property analysis, gentrification, land-use change and neighbourhood definition (Galster, 2001).

2.1: Built Environment Models

In this research we link urban geography to the built environment by integrating socio-economic spatial data with iconic urban models. As noted earlier, an iconic model is a geometrical representation of a feature or set of features, typically taking the form of a physical scale-model (such as an architect's miniature building model) or a digital model. 3D digital city models have become increasingly widespread and sophisticated with the development and integration of computer-aided design (CAD) software, geographical information systems (GIS), computer graphics, web and aerial sensing technologies (see Zlatanova and Prosperi, 2006, van Oosterom *et al.*, 2008 for further reviews). An earlier research project at the Centre for Advanced Spatial Analysis (CASA) developed a 3D digital city model of London called 'Virtual London' (Batty and Hudson-Smith, 2005b) and this acts as a foundation for this research.

Potential applications of 3D digital city models range between urban planning, telecommunications, architecture, facilities and utilities management, property analysis, marketing, tourism and entertainment (see Batty *et al.*, 2001 for a review). The development of web and virtual globe technologies has given a massive boost to digital urban models, enabling widespread access and interaction by the public through geo-browsers such as the popular Google Earth.

---

mathematical formula, often in the form of algebraic equations within a digital form (e.g. a computer) such as in the case of land use transport models (e.g. Batty, 1976).

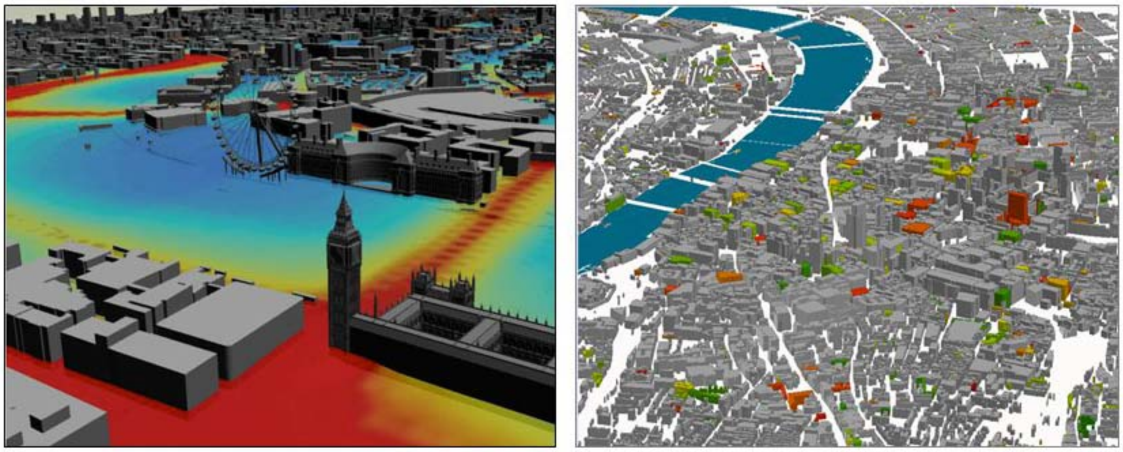
Whilst the visualisation capabilities of 3D digital city models are clear, their analytical functionality is often underdeveloped (Batty and Hudson-Smith, 2002). Significant advances have been made in increasing the geometrical sophistication of 3D digital city models, but many models remain 'empty shells' without any socio-economic data associated with the buildings or the capability to analyse the role of the built environment in urban processes. We believe that future advances will explore how such models can be populated with socio-economic data and linked to transportation networks thus moving from visualisation to focus on policy applications and analysis. This would essentially mean enhancing digital city models to become Planning Support Systems (PSS), i.e. tools to aid and enhance planning tasks (see Brail and Klosterman, 2001).

We define four levels of integration between iconic urban models and geographical analysis. Starting with the most basic, these techniques are:

i) the spatial overlay of thematic data on a city model for visualisation;
ii) the creation of a city model database with building geometry linked to address/cadastral geography for socio-economic attribution (i.e. a Built Environment Model);
iii) the combined spatial analysis of socio-economic attributes with urban form geometry;
iv) the integration of Built Environment Models with urban symbolic mathematical models (for example, land-use transport models).

These methods range from description to analysis, and from static representations to the potential for dynamic modelling. This research focuses on the second and third methods of this typology where the benefits for urban research and PSS are most immediate. The integration of Built Environment Models with symbolic urban modelling is not explored in this paper, though we believe that improved accuracy in the representation of urban form and function would be beneficial, particularly for land-use and urban growth models, and should be investigated in future research.

The first method of combining datasets is the overlay of thematic data on a city model. The visualisation of built form and landmarks can help users to navigate and identify familiar urban locations, thus improving data legibility (Tuan, 1977). Figure 1A illustrates this with an air pollution surface combined with the Virtual London model. Insights can emerge from seeing patterns which may not have been evident without the context of urban form (Batty, 2007). With this method relationships between urban geometry and geography are visual and lack explicit spatial relationships.



A **B**

Figure 1: Virtual London 3D City Model with Different Data Layers Overlaid on Top.

A: Nitrogen Dioxide Layered onto the Street System of London, and **B:** Querying Buildings Developed from 2001 to 2004 (Source: Batty, 2007).

A more advanced approach is to develop a spatial database of the built environment- a Built Environment Model, and store socio-economic attributes associated with buildings. These attributes of buildings can be queried and visualised within a GIS as demonstrated in Figure 1B where buildings in Central London of a particular age are identified. This type of functionality is relevant to planners querying building stock, and to researchers studying property markets and relating socio-economic trends to the built environment.

To create a Built Environment Model, the geometry of buildings has to be linked to socio-economic attributes through a postal or cadastral geography. Depending on data availability, the development of the data model can be a non-trivial task. In the United Kingdom (UK) context there has been much recent innovation in spatial address infrastructure, yet a full description of building addresses and sub-building geometry of premises remains incomplete. Data modelling techniques to link topographic and spatial address datasets to try to fill in information gaps are thus pursued in this research.

Once the Built Environment Model has been created then it is possible to perform spatial analysis linking geometric and socio-economic properties- the third stage in our typology of integration. A simple example is to calculate building density measures by dividing a quantity (e.g. the number of residential properties) by building footprint or volume. Any address-based data can be incorporated, from business function data, to geodemographic and neighbourhood characteristics. Several examples of these techniques are provided in Section 7 of this paper.

Overall with a data model that successfully integrates urban form and socio-economic data, there are a wide range of potential applications relating to planning practice and geographical research. These applications vary in terms of data inputs and analysis methods, and so flexibility is a useful quality of the data model to enable a range of applications. A common characteristic is the variation in the scale of data inputs and of the required output, as discussed further below.

2.3: Scale and Aggregation in Built Environment Models

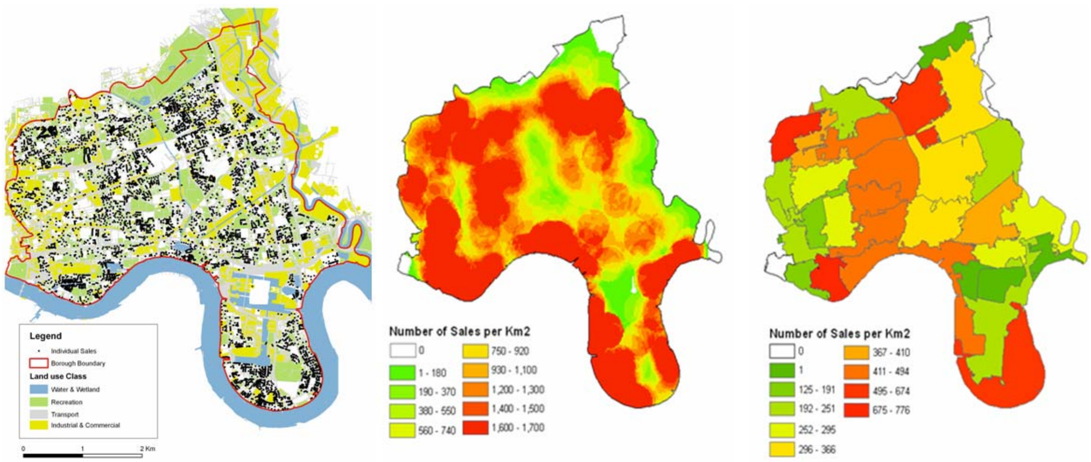
In geographical contexts the term scale is used to refer to both the level of detail and extent of spatial data and analysis. This dual meaning derives from the common association between data resolution and study area size. This balance is significant both for spatial analysis and visualisation. For spatial analysis large high-detail datasets increase methodological complexity and computation demands; while in visualisation there is a limit to the density of information that is intelligible to the viewer on a page or screen (Skupin, 2000).

The association between large scale and coarse resolution is readily seen in aggregated types of urban measurement, where information on the form and pattern of the built environment is frequently missing (Talen, 2003; Moudon, 2002). Built environment studies ideally require analysis that is both fine-scale, to include premises, buildings and streets, and large extent, to allow the study of city-wide processes. To achieve this, techniques are needed to integrate datasets at fine-scales, make calculations for large urban areas computationally feasible, and to visualise data at multiple scales.

The most common quantitative approach geographers use to explore urban data is aggregate zonal analysis. Aggregate analysis is a powerful method of simplifying the complexity of urban data, smoothing over local variation and enabling patterns to be identified. It complements the structure of core datasets such as censuses of population and businesses. Additionally using aggregate data reduces computational demands for spatial analysis operations.

On the other hand, aggregation is a source of uncertainty and error in geographical research. Datasets and zonations can vary widely in terms of geographical scale, aggregation rules and spatial resolution, and this influences research outcomes. All aggregate units are spatially modifiable, i.e. they can be partitioned geographically in many different ways to generate different types of results. This is known as the modifiable areal unit problem (MAUP) (Openshaw, 1984).

Of concern in built environment research is that socio-economic zonations typically ignore urban form, and overlook variation in physical features. For example, physical urban barriers can be influential in urban processes such as segregation (see Rabin, 1987), yet are often not represented using zone boundaries. To highlight the impacts of data aggregation on the results and interpretation of geographical research we provide a simple example. If we create a point map of housing sales data over several years from a district in London (Tower Hamlet's, for example) we see there is a detailed pattern of clusters spread over the entire borough (Figure 2A). This detailed pattern can be simplified through aggregation. Figure 2B we show the point density of individual sales[3], which captures the general distribution of points whilst losing some of the detail. This contrasts with Figure 2C, where zonal aggregation (in this instance postcode sectors) has changed the distribution, with the impact of built environment features such as water and parks now masked within the zonal geography. From this simple example it is clear that the scale and aggregation impacts on the overall results and interpretation.



A B C

Figure 2: Masking Variation with Data Aggregation.

A: Individual Addresses **B:** Point Density of Sales, and **C:** Density of Sales per Postcode Sector.

These issues can be minimised by using disaggregate data. Greater flexibility is possible as fine-scale data can be aggregated into any chosen zonation. Therefore the impact of the MAUP can be tested and quantified. Zone boundaries can be tailored to particular studies, and built environment features can be considered. The effect of the trade-off between level-of-detail and extent can be reduced.

---

3 Point densities - calculates a magnitude per unit area of the point/sale feature that fall within a neighbourhood around each point. A 500m radius was used.

Figure 3 illustrates aggregation techniques designed for built environment analysis. Data is first integrated at the disaggregate scale of building footprints and addresses. This can be aggregated into intermediate geographies based on physical features, such as street block geography for density analysis, or street network geography for accessibility analysis. For large urban areas grids at various resolutions can be used. A regular grid should have less inherent bias than socio-economic zones tailored to particular administrative purposes.

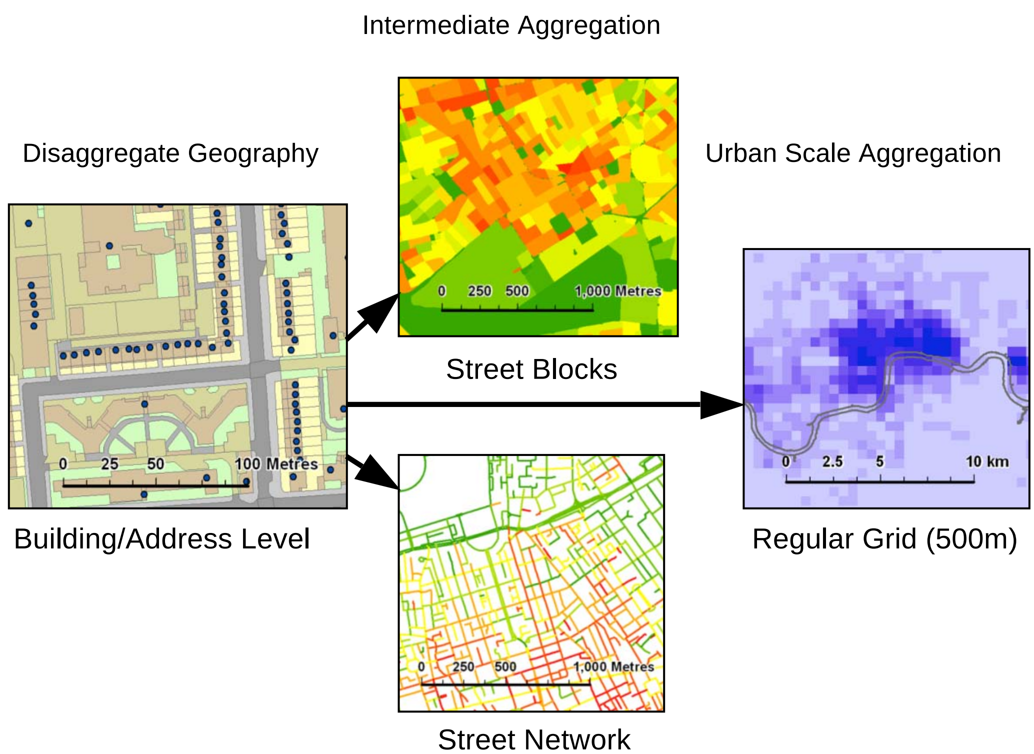


Figure 3: Aggregation Methods for Varied Scales of Built Environment Analysis.

This approach relies on the availability of fine-scale geographical data, and increasingly this is becoming available, as will be discussed in Section 3. Privacy is an important consideration for demographic data, and this prevents data such as the census being released at address level. Attempts to disaggregate zonal data are likely to result in ecological fallacy errors, where attributes based on aggregate data are applied to individuals that form the aggregate group (de Smith *et al.*, 2007). Microsimulation methods are an appropriate tool to mitigate this (see Clarke and Holm, 1987). In this study we base our analysis on anonymous address-based data and so microsimulation approaches are not pursued, but we believe this could be a fruitful avenue for further research especially in terms of linking iconic to symbolic models in the form of agent-based models, for example.

However, the fine-scale large-extent approach is computationally demanding. For a large city such as London, millions of features must be joined to integrate data at address and building scales. The continued advances in computer hardware and Spatial Database Management System (SDBMS) software make this approach feasible. Suitable data models that fully represent relationships between addresses and building geography are necessary. Once the data model is implemented, then analysis becomes an address matching process (i.e. attribute joins) as opposed to more computationally intensive spatial joins.

2.4: Summary

This section has argued that geographers and planners have often studied cities differently, looking at the geographical aspects or the geometrical aspects respectively. Geometrical built environment information is generally overlooked in geographical analysis because of the lack of fine resolution and extensive urban datasets, and because of the aggregate methodologies commonly used. These challenges are beginning to be solved as a new kind of fine-scale urban geography is emerging, using datasets which are sufficiently intensive to detect detailed patterns and morphologies and also sufficiently extensive to enable these patterns to be generalized to entire metropolitan areas. We have highlighted an alternative to the zonal approach that much of quantitative geographical research relies on. However, this requires fine-scale large-extent datasets. Nevertheless, one can potentially build new theories and models which were not possible in the past, and provide a stronger evidence base for planners and researchers.

3. Urban Spatial Data Advances

The field of urban spatial data infrastructure has seen continuous development and innovation in recent years, both in the domains of urban geometric data and socio-economic geographical data (see Onsrud, 2007; Masser, 2005). Improvements have included higher spatial resolution, richer attribution, improved integration, and entirely new data sources emerging from new technology. These advances towards detailed and extensive urban GI make this research feasible.

3.1: Urban Geometric Data Sources

Geometric urban data can be used in the analysis of the location of features, in terms of context and accessibility, and the geometric quantification of features, such as areas, heights and volumes. To measure these properties there are two main sources of urban geometric data: topographic mapping and aerial/remote sensing data. Topographic mapping data continues to improve as GI applications have expanded, markets have matured and the economic value of GI data is increasingly being recognised (see Longhorn and Blakemore, 2007). Detailed spatial features such as building outlines and roads are more widely available. Furthermore, major providers can now offer topographic data closely integrated with other layers such as road network data and spatial address data. Advances in spatial address infrastructure in particular enable integration with socio-economic data sources.

In the UK context, the Ordnance Survey's (OS) MasterMap product suite (Ordnance Survey, 2007b)[4] is the most comprehensive source available and it continues to advance in content and detail. MasterMap provides a UK-wide detailed vector topographic layer with building outlines, internal property divisions, and street infrastructure mapped to an accuracy of 1m. This provides the core building geometry data in this research. The other key feature of MasterMap is the spatial address product Address Layer 2 (Ordnance Survey, 2007a), which can be used for address matching. Addresses are geocoded at the

[4] http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/

level of building polygons, and some functional attributes are also included to identify residential and commercial properties (this is discussed further in Section 3.2.1).

Topographical data sources are by their nature two dimensional, and so cannot adequately represent three dimensional urban features such as multi-storey buildings, bridges and subways. Advances, however are being made towards data standards that fully represent the three dimensional nature of the built environment, for example, the City Geography Markup Language (CityGML) standard for 3D digital city model exchange (Gröger *et al.*, 2008). The challenges and costs of gathering and integrating widespread 3D built environment data however are high. Large quantities of 3D data relating to specific buildings lie in CAD models and drawings, but there are several technical and data ownership hurdles still to be overcome if these are to become a common feature of city models (see Zlatanova and Prosperi, 2006).

So lacking three dimensional data from topographic sources, we look to aerial and remote sensing for building height data. In particular Light Detection and Ranging (LIDAR) aerial data provides information on the external geometry of buildings and is available over large urban areas. While the automatic and semi-automatic derivation of 3D models from LIDAR is a very active research area, internal property divisions cannot be derived, and these are needed for this particular research[5]. Therefore we use topographical mapping data as the geometric base and this is augmented with LIDAR height data aggregated across building footprints. This was the method used in the Virtual London project (Batty and Hudson-Smith, 2005a).

The combination of 2D topographic data and LIDAR essentially creates a 2.5D block model of the city. This method is effective for adding urban texture and a 'sense of place' to visualisations. The lack of true 3D data does however mean that vertical geometry is missing, so the distribution of functions between floors in a multi-storey building is not known. This undermines the accuracy of direct geometric measures of premises,

[5] However, attempts are being made to overcome this issue but tend to be error prone. See Orford (2010) for more information.

particularly for mixed use buildings (see Section 6.2). However, this shortfall can be minimised by linking to address-based sources of property quantification (e.g. property taxation surveys) as discussed further in the following section.

3.2: Urban Socio-Economic Data

Several trends have come together to expand the availability of address-based socio-economic data. The increased recognition of the importance of GI has led to initiatives to standardise and integrate GI data, including the improvement of spatial address information (e.g. Department for Communities and Local Government, 2005). Another significant trend has been e-government initiatives to increase the availability of government services and information online. This includes services such as business rates, property sales and planning permissions.

There are two basic categories of address-based socio-economic data relevant to this research. The first category is information that relates to real estate, such as function, ownership, market transactions, size, age and so forth. The second category covers information about residents or businesses who occupy that real estate, for example, demographic information and business classifications. Here we focus mainly on the first category of real estate attributes, with the priority to analyse building function and produce a residential classification (see Section 5). The methodology developed can also incorporate the second category of demographic and business information, and we intend to explore this in future research.

3.2.1 Property Function Data

Function[6] is a core attribute for understanding urban structure. From the description of basic urban districts such as centres and suburbs, to detailed studies of mix-of-uses and urban agglomeration, functional information is crucial. The mapping of building use and

[6] The phrase 'land use' is often used to describe urban function. Since multiple functions are often combined on a single piece of land in cities, 'land use' can be ambiguous and so function is chosen here as the preferred term.

function has several applications in visualising patterns of land use and urban texture, and investigating mix of uses and local service provision.

Up until recently property function data has not been available in a detailed comprehensive format for the UK. The OS has been working to overcome this through its Address Layer 2 product (Ordnance Survey, 2006, 2007a)[7]. This is the first attempt at a complete address-based functional data source for the UK and it is a significant advance. There are however shortcomings with the current (2008) release (see Section 5.4). An example of the potential output from the dataset generalised into basic categories is shown in Figure 4. Useful features of the data include detailed residential address information, with multiple dwellings within buildings represented as coincident points. The number and location of these dwelling addresses can be used to derive a classification of residential types, as discussed further in Section 5.3.

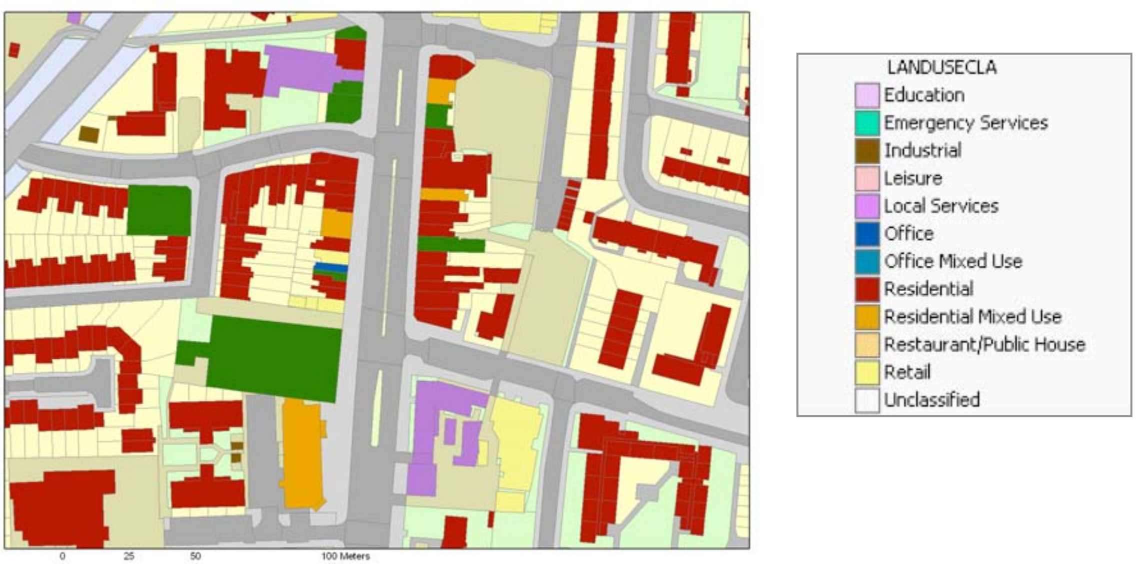


Figure 4: Building Function and Land Use Classification.

A significant limitation to the Address Layer 2 functional data is the incomplete classification of commercial properties (see Sections 5.4.2 and 6.1 for further

[7] http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/

information). The major source of commercial property information in England and Wales is the local taxation administration body, the Valuation Office Agency (VOA). VOA surveys are carried out every five years and record detailed business function classifications and property size measurements. The OS are working to integrate the VOA classifications into their Address Layer 2 product, but at present this process is far from complete. It is possible to access the VOA independently to analyse commercial real-estate geography (Smith, 2009), though the data is not geocoded and must be address matched by the user.

Overall significant progress is being made in provision and integration of detailed spatial data on urban function in the UK. The data infrastructure is still in development and there are some shortcomings, but we believe the datasets are at a sufficient stage to illustrate their application to city model research.

3.2.2 Real Estate Transaction Data

Real estate is bought, sold and rented in a series of markets, with profound results for the geography and structure of cities. With data at address level, analysis is possible at the scale of property transactions. This is an active area of geographical research, including residential valuation models (Pryce and Evans, 2007) and studies of property uplift from infrastructure development (Atisreal, 2005). Built Environment Models have the potential to enhance these research efforts and integrate further data sets to improve results, as well being a means of communicating market trends to researchers and planners.

Property transaction data sources include residential property sales and mortgage databases from banks. Business rental information is more restricted in the UK as it is judged to be commercially sensitive. The UK government now publishes residential property sale information through the Land Registry, beginning in January 2000 with the most up-to-date records 3 months old. Attributes include the transaction price, date, and a classification into basic housing types. The data set does not include local authority housing transactions.

The Land Registry data does not record all significant real-estate attributes, in particular the size of the property (which is often the most influential factor in the transaction price, see Fotheringham *et al.*, 2002). It is possible that digital city modelling techniques could be used to estimate property size and augment this dataset (e.g. Orford, 2010), although the 2.5D nature of topographic and aerial data hinders this task.

Banks and building societies providing mortgages are an alternative source of residential housing information. These tend to be more detailed than the land registry information and have been used successfully in hedonic price models (see Pryce and Evans, 2007). The main shortcomings of this source is that, unless all mortgage providers supply information, then the number of properties covered will be much less comprehensive than the Land Registry data[8].

3.2.2 Residential Type Data

Residential or dwelling type data has a number of applications in urban research including density and morphological analysis (Longley and Mesev, 2003), socio-economic segregation, environmental quality, and residential property analysis. The main source of residential type data for the UK is the 2001 census, which records general categories of housing and is available for aggregate zonal geographies. The spatial and temporal resolution of the census limits some of the more interesting applications related to fine-scale and dynamic residential processes.

Recent research has been investigating whether topographic and address-based data sources can improve or at least complement the census dwelling data (Orford and Ratcliffe, 2007). Micro scale classifications are possible at building level, and there is the potential to link this data to other information such as property transactions or geo-demographic data. This research pursues a micro-scale residential classification using these methods as discussed in Section 5.4.

[8] Additionally not all property transactions are financed through mortgages, therefore making such data even less comprehensive.

4. Project Aims and Greater London Context

The data advances discussed in Section 3 enable a range of new analytical possibilities for urban research linked to the built environment. This particular study is in collaboration with the planning authority of Greater London, the Greater London Authority (GLA), and the structure of the city model is directed towards the analysis of current planning issues in Greater London (henceforth referred to as London).

As a major world city with a resident population of over 7.5 million, urban planning challenges in London are multiple and complex. London's population is projected to increase by over a million by 2026 (GLA, 2006a) whilst employment is projected to increase by over 900,000 jobs between 2006 and 2026 (Spooner and Cooper, 2006). The current economic crisis has increased the uncertainty of economic and demographic forecasting and may well curb these predictions, but significant growth remains likely. Recent and projected rises in employment and population translate into major development and built environment change. For example, the London Plan has set targets of building over 30,000 new homes per year to address the increase in demand (see GLA, 2004; GLA, 2006b for further information). Planners managing this growth must consider the various and often conflicting goals of business needs, social equality, sustainability and conservation.

A Built Environment Model with analytical capabilities would improve the quantitative evidence base for managing urban change. We do not wish to imply a city model is a panacea for planning- clearly for issues such as deprivation and migration such a model is of little relevance. But for planning issues linked to the built environment we believe this approach can improve the current evidence base. The urban form and function approach pursued here can be used in several planning tasks, such as the visualisation of urban structure and development, the measurement of density and mix-of-uses, and the classification of building forms such as residential housing types. These tasks all relate to characterising and analysing urban texture, and managing growth. There has been much debate in London surrounding the significantly higher densities of recent developments and whether they integrate with the existing urban fabric (GLA, 2002). There is a lack of a quantitative base from which to make these assessments and to analyse what change is occurring. Mix-of-use and density measures are also highly relevant to urban sustainability research, and this too would benefit from improved evidence.

Measurements of urban form, density and function vary with scale, and comprehensive analysis should include measurements at multiple scales. This flexibility is therefore a key feature of the Built Environment Model, by including fine-scale data that can be aggregated to larger scales depending on the questions being asked. This feature also complements planning practice, which considers fine-scale issues at local government level, and larger scale trends at strategic regional government levels. At strategic regional scales, planning is concerned with larger scale land use patterns such as the efficiencies of monocentric and polycentric structures (see Aguilera, 2005; Gordon and Richardson, 1996). The degree of monocentricity or polycentricity in form and function can also be explored using built environment models (see Smith, 2009 for such an application).

While we believe there is great potential in developing a Built Environment Model to enhance urban analysis, the current stage of data infrastructure (in the UK at least) is limited. The lack of a comprehensive property cadastre in the UK leaves data gaps for significant real estate attributes such as property function and size. The methodology for creating the Built Environment Model must then begin with some basic structures to relate address data to topographic mapping, and with algorithms to classify built form functions.

5. Methodology for Creating the Built Environment Model

To create a Built Environment Model we need the relationships between building geometry and address geography to be fully represented. This section describes the data model that allows the topographic building data to be integrated with the spatial address data. First, we present the data model in Section 5.1 that allows groups of address objects to be associated with multiple building polygons. The algorithm to identify which building polygons belong to which addresses is discussed in Section 5.2. Section 5.3 concerns how residential type is inferred from the address and geometric data. Finally Section 5.4 discusses the functional classification of non-residential addresses.

5.1 Address Points and Building Polygons

Spatial address data is generally point based, and this is the case with the OS data used in this research. The points are abstractions of the areal extent associated with a particular address, derived from the UK postal service database, the Postal Address File (see Longley and Mesev (2000) for a discussion). In Address Layer 2, each address is linked to a single building polygon and has a single x and y coordinate, as shown in Figure 5. There are many building polygons that do not contain an address point. In reality each address can relate to a larger area of buildings and land than the coincident building polygon defines (for example, the large building at the top of Figure 5). This is not recorded within the OS data model.

Adding another layer of complication, each building polygon can contain multiple addresses (as for example, a block of flats). So the MasterMap data model has a one-to-many relationship between building polygons and addresses, excluding many-to-many relationships, as illustrated in Figure 6A. The number of building polygons that do not have coincident address points is significant. In the London study area, there are 3.49 million building polygons from OS MasterMap data, of which 1.46 million are not directly addressed via Address Layer 2. Linking the unclassified building polygons to addresses would be useful for several reasons, the first is for visualisation, providing a more complete picture of how address attributes are related to the built form, and secondly for analysis, ensuring that the full extent of the building geometry is known. A modified data model to enable many-to-many relationships between addresses and building polygons is illustrated in Figure 6B. Multiple addresses within the same polygon share a common group reference, forming an Address Group object.

Figure 5: Addresses (Circles) and Building Polygons in MasterMap.

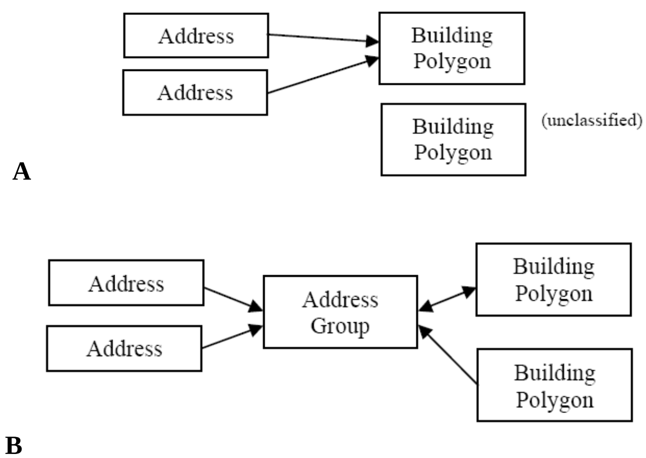
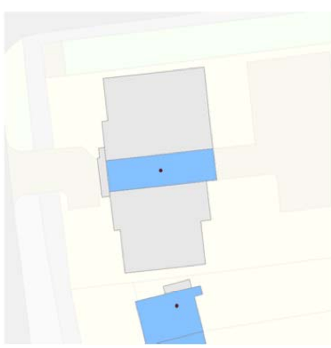


Figure 6: Address and Building Geometry Relationships.

A: MasterMap One-to-Many Data Model, **B:** Modified Many-to-Many Data Model.

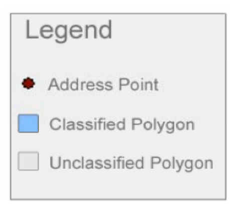
5.2 Classifying Building Polygons Using Spatial Relationships

The task for the classification algorithm is to associate unclassified polygons with Address Group objects based on spatial relationships. Different types of relationships between classified and unclassified building polygons are illustrated in Figure 7. Adjacency relationships are central to the classification algorithm as contiguous polygons in MasterMap often comprise a single building which in reality relate to the same address. Adjacency relationships vary between basic topology as shown in Figure 7A, involving first order neighbours of single classified polygons, to more complex topology, for example, in Figure 7B where unclassified polygons are related to multiple classified polygons at various orders of adjacency. We assume here that basic adjacency relationships, particularly first order neighbours of one classified polygon, relate to the same Address Group, i.e. the address point marks the delivery point within a single multi-polygon building. For complex adjacency relationships the degree of ambiguity and uncertainty is higher. Unclassified polygons with more than one classified neighbour of the same order remain unclassified by the algorithm.



A: Basic Adjacency

B: Complex Adjacency



C: Non-contiguous Minor Outhouses

D: Non-contiguous Multi-building Group

Figure 7: Unclassified Building Polygon Relationship Types.

In addition to unclassified polygons related by adjacency, there are also a high number of non-contiguous building polygons in the data. The vast majority of these are minor buildings such as sheds and garages as highlighted in Figure 7C. Classification of these minor buildings is not essential for this research, as they do not greatly affect function or property size. There are however other configurations of non-contiguous buildings that do have a significant bearing on property size as shown in Figure 7D. This type of arrangement is generally found in institutional buildings such as hospitals and university campuses. Automatic identification of these configurations is challenging as it requires geometric and morphological analysis to estimate the association between nearby building polygons. The scope for classification errors is high. For this study the volume of building polygons to be processed and limited time for detailed validation has led to the decision not to include non-contiguous building classification in the algorithm.

5.2.1 Classification Algorithm and Block Features

For the purposes of assessing adjacency relationships, and to increase efficiency by minimising the number of spatial join operations, it is useful to create the concept of Blocks in the data model. Blocks are groups of contiguous building polygons as highlighted in Figure 8. Each block has a number of directly classified polygons from the address data (e.g. Address Layer 2), and a number of unclassified polygons. Based on these two properties, the classification algorithm selects how to process each Block, as shown in Table 1. The simple steps highlighted in Table 1 limit the number of Blocks that need to be processed with the adjacency algorithm. The adjacency algorithm loops through the Block polygons, assigning unclassified polygons to neighbours (classified polygons) and gives them the same address until no unclassified polygons remain. Polygons that share multiple closest neighbours at the same order of adjacency are left unclassified as highlighted in Figure 8. The method of classification and order of adjacency is stored in a classification confidence attribute which can be used for further analysis if needed.

Classified Polygons on Block Feature	Unclassified Polygons On Block Feature	Algorithm Procedure
0	-	Polygons remain unclassified.
1	> 0	Link all unclassified polygons to the Address Group of the single classified polygon.
> 1	> 0	Run adjacency algorithm to assign unclassified polygons to closest classified polygon Address Group.

Table 1: Algorithm Steps Relating to Block Properties.

Table 2 shows the results of the adjacency analysis. An additional 20% of the total building polygons for the London study area are classified through this method. A large number of unclassified polygons still remain (32%) so there is still much scope to improve the process and include non-contiguous polygons in the classification.

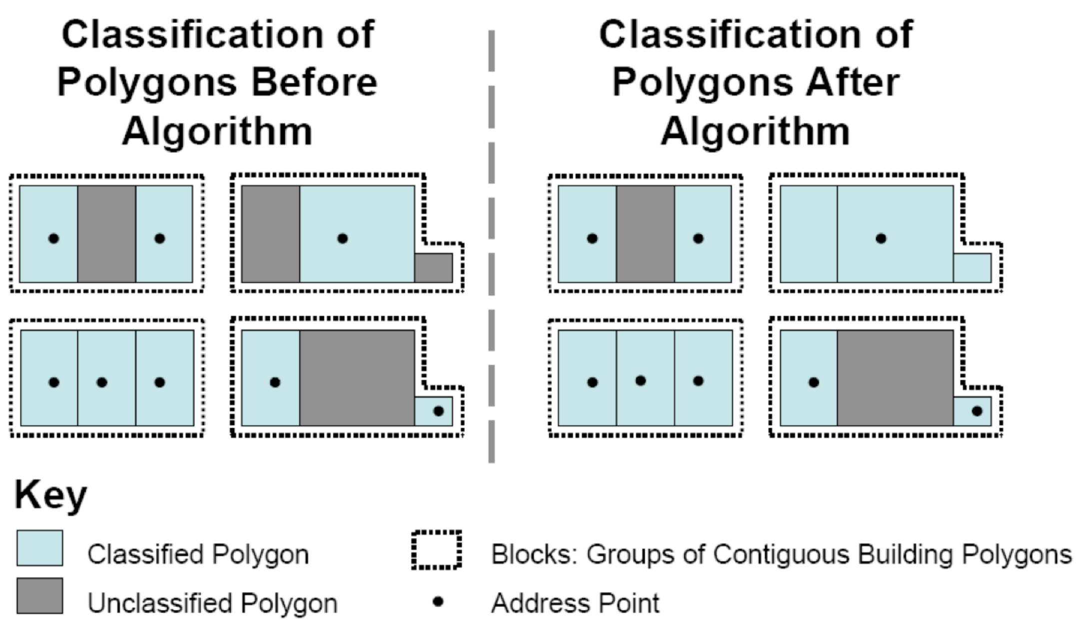


Figure 8: Classification of Polygons Before and After Running the Classification.

Total Building Polygons	Before Adjacency Classification		After Adjacency Classification	
	Directly Classified Polygons	Unclassified Polygons	Polygons Classified Through Agency Analysis	Remaining Unclassified Polygons
3485711 (100%)	1683349 (48%)	1802362 (52%)	689243 (20%)	1113119 (32%)

Table 2: Adjacency Classification Totals for the London Study Area.

5.3 Residential Type Classification

The relationships between building geometry and addresses can be used to classify residential buildings into dwelling types. There are various means of defining housing types, and classification schemes vary between data providers and between nation states across the world (see Orford and Ratcliffe, 2007 for a detailed discussion). Here we pursue a dwelling address and geometry-based functional classification method. With this approach, houses are classified using the number and location of residential addresses. So for example, a large house that has been subdivided into flats would be classed as flats in this method. Other methods, for example, a historical approach, may be more interested in the original building type.

The residential classification is applied to residential only (single use) classified polygons, identified using Address Layer 2 attributes (see Section 5.4). The algorithm relies on two key properties: the number of residential dwellings within each polygon, and the number of classified polygons on the parent Block feature. This is a simplification of the classification method developed by Orford and Radcliffe (2007). However, this was deemed acceptable due to time constraints of the project, computational demands and the size of the study area which is a significant order of magnitude larger. With these two properties, some basic rules can be set out to derive housing types as shown in Table 3.

Residential Addresses within Polygon	Classified Polygons on Block	Residential Classification Result
1	1	Detached House
1	2	Semi-detached House
1	> 2	Terraced (Row) House
2	-	Maisonette (Upper/Lower Villa)
> 2	1	Flats (single block)
> 2	> 1	Flats (contiguous blocks)

Table 3: Residential Classification Rules.

An example of the output from this classification is shown in Figure 9, with local scale clustering of housing types clearly highlighted. There is a degree of arbitrariness in the exact definition of some of the housing types. In the case of end-terrace housing these could be classed as ‘Semi-detached’ or ‘Terraced’ depending on the classification method used. The Block-based algorithm used here classifies end-terraces as ‘Terraced’. This issue is more a question of classification semantics than error *per se*. Of greater concern is where the algorithm results in a clear classification error. The issue of trivial building polygon links, such as minor extensions that join together detached houses, is one such source of error (see Orford and Radcliffe, 2007). Identifying this error requires some form of building roofline analysis. This process is computationally demanding for such a large study area and has not implemented at this stage of the research.

Figure 9: Example of the Residential Type Classification.

Generally as an automatic classification method the algorithm performs consistently and produces useful results for analysis with strongly clustered housing groups as highlighted in Figure 9. More complex classifications could be devised, for example, relating houses to gardens, or breaking down classes into sub-categories. In this research we intend to begin with basic categories for mapping out general relationships. More sophisticated classifications could be added in future work.

5.4: Building Function Classification

The Address Layer 2 data includes several functional attributes sourced from OS surveys, VOA surveys and the National Land Use Database (NLUD) information. The OS survey data, referred to as the 'field surveyor's allocation' or 'Basefunction' attribute, is included for all addresses and is the basis of the functional classification developed in this research. Unfortunately the VOA and NLUD attributes are highly incomplete in the 2008 Address Layer 2 data and are not used here for classification.

This OS Basefunction attribute allows residential and non-residential addresses to be distinguished. For non-residential addresses, there are over 1,500 functional types- a highly detailed dataset. The classes come from raw surveyors' data and are a somewhat jumbled assortment of features ranging from farming structures, to energy infrastructure and urban functions. These classes lack any hierarchy and need to be simplified to be intelligible in any kind of visualisation and analysis.

5.4.1: Functional Classification Scheme

With over 1,500 functional types from the OS survey data, these need to be summarised into general categories to be intelligible in visualisation and analysis. The classification scheme developed here is based around core urban functions, such as residential, office and retail uses. The intention of the scheme is that, when the data is mapped, basic aspects of urban structure can be grasped quickly by viewers. Therefore the scheme has a limited number of classes, with the intention that these are distinctive in their function and are also likely to share similar locational patterns. The ten classes created are described in Table 4. Note that the final class of 'General Commercial' is forced by data issues (see Section 5.4.2).

Classification schemes are to an extent arbitrary. There are several function types that have several roles and fall between the classes in Table 4. For instance, high street banks are included in the 'Local Services' category, but banks also incorporate elements of retail and office functions and an alternative classification is possible. Libraries are also classed as 'Local Services' though they have a strong educational role. Hotels are classed as "Hotel" but they often include bars and nightclubs which are found in the "Leisure-Restaurant" category. Alternative classification schemes can be designed and implemented reasonably quickly. Future applications may well require a more detailed classification of functions than is provided here, and so the scheme could be tailored to these requirements. Fuzzy classification methods would also be useful to handle the kind of ambiguities discussed.

Classification Group	Major OS Basefunction Classes Incorporated Within Group
Residential	DWELLING
Office	OFFICE, COUNCIL OFFICE, GOVERNMENT OFFICE, BUSINESS PARK, FINANCIAL AND PROFESSIONAL SERVICES
Retail	SHOPPING, SHOPPING CENTRE, SUPERMARKET, SUPERSTORE, RETAIL PARK, MARKET
Local Services	BANK, CHEMIST, COMMUNITY CENTRE, POST OFFICE, CHURCH, MOSQUE, SYNAGOGUE, LAUNDRETTE, LIBRARY, YOUTH CENTRE
Leisure-Restaurant	ART CENTRE, BINGO HALL, CAFÉ, BAR, MUSEUM, RESTAURANT, PUBLIC HOUSE, TAKE AWAY, NIGHTCLUB, PUBLIC BATHS
Education	SCHOOL, NURSERY, PRIMARY SCHOOL, SECONDARY SCHOOL, UNIVERSITY, FURTHER EDUCATION COLLEGE, HIGH SCHOOL
Health and Emergency Services	HOSPITAL, GENERAL HOSPITAL, DENTAL HOSPITAL, FIRE STATION, POLICE STATION, CHILDRENS HOSPITAL
Hotel-Accommodation	HOTEL, GUEST HOUSE, HOSTEL, YOUTH HOSTEL, MOTEL, INN
Industrial	FACTORY, MANUFACTURING, INDUSTRIAL ESTATE, ELECTRICITY SUB STATION
Transport	BUS STATION, COACH STATION, RAILWAY STATION, FILLING STATION, GARAGE
General Commercial	GENERAL COMMERCIAL

Table 4: Functional Classification Scheme from OS Basefunction Attribute.

It would be useful for the classification scheme to conform to an official standard. The most likely candidate in the UK is the Use Class codes[9] that are an integral part of the planning permissions process. The Use Class scheme classifies office and retail space by the quality of the real-estate. Unfortunately this kind of property specification is not possible with the OS data at present. It is likely the OS will change the format of the functional data in future releases of Address Layer 2 as the current structure lacks any classification hierarchy and is difficult to manage. Use Class codes would be a good basis for providing an improved structure, or indeed the VOA functional classification scheme, Special Category Codes (SCATs).

[9] Use Class codes refer to classes of use for England which were set out in the Town and Country Planning (Use Classes) Order of 1987 and its subsequent amendments. Readers wishing to know more are referred to the UK Planning portal http://www.planningportal.gov.uk/england/genpub/en/1011888237913.html.

5.4.2: Commercial Classification Limitations

There are a substantial number of non-residential addresses that lack attributes from either OS cartographic surveys or the VOA data. These addresses are given the provisional category of ‘General Commercial’ in the OS data (although there also a small number of non-commercial addresses that are also included in this category). In the London context, 51% of non-residential addresses have the General Commercial classification. This is a very significant shortcoming for any analysis that needs to distinguish between commercial functions such as retail and office activities. In contrast to the commercial data, the functional classification of non-commercial services, such as educational facilities, is much more complete.

5.4.3: Classification of Mixed-Use Buildings

It is common in dense urban environments for multiple functions to be combined in a single mixed-use building. In the study area approximately 2% of building polygons have multiple functions, and these are strongly clustered in the Central London. The data sources used in this research do not include any sub-building information and so these multiple functions need to be combined into some form of mixed-use classification.

The majority of mixed use buildings (82%) combine only two functions, and of these functions, residential is by far the most common (74%). Therefore a simple method of classifying mixed-use buildings is to prioritise non-residential functions, creating classes such as ‘Mixed-Use Office’ and ‘Mixed-Use Retail’. For mixed-use buildings that combine multiple non-residential functions some form of comparison is required to estimate the dominant function. For this purpose a script was written to calculate the function with the greatest number of addresses. In cases of two or more functions with the same number of addresses the classification simply remains ‘Mixed Use’. This is not an ideal method of estimating the most prominent function, as the number of addresses does not include the size of the premises. For example, a large office of many floors may be above two small retail units. This type problem requires sub-building geometry or detailed real-estate floorspace data at premises level to be tackled adequately.

6: Built Environment Model Validation

The previous section detailed our methodology for creating a Built Environment database. To test its usefulness we need to be able to explore sources of error and validate the data model. It is to this, that we now turn. There are potential sources of error at each stage of the analysis process, from the accuracy of the original datasets, to the matching of addresses to building polygons, and the classification of buildings into functional and residential types. As the datasets used in this research are somewhat unique there is a lack data that can be used to validate the results against. One exception is the residential classification data which can be compared to the dwelling type data from the 2001 census (see Section 6.2). For the other data sources the most suitable validation method is to ground truth the data against manual surveys. Section 6.1 compares the functional classification outputs against manual surveys of two commercial streets in London. This is a brief validation survey, but is sufficient to highlight sources of error in the functional data.

6.1: Functional Classification Validation Survey

The output of the functional classification was compared against manual surveys of two streets in London. These were a high density city centre high street- Tottenham Court Road- and a lower density suburban centre high street- Station Road in Edgware. While far from comprehensive, this small survey should be indicative of the general accuracy of the classification data within London. The city centre street has been chosen as the potential for error is particularly high, due to the high diversity of uses, frequent changes of use and vertical complexity with multi-storey buildings.

The survey included a number of variables to measure various sources of error. These can be divided into two general types – classification errors and geometrical errors. Classification errors result from errors in the attribute data of the Address Layer 2 dataset (discussed in Section 6.1.1), while geometrical errors are related to the 2D generalisation of three dimensional buildings (discussed in Section 6.1.2).

6.1.1: Classification Errors

Looking in detail at the classification errors, the survey measured errors in the identification of residential addresses (relevant to the residential classification); errors in the recorded business name (caused by out-of-date information); incorrect 'Basefunction' attributes (see Section 5.4) where a clear classification error has been made; and finally the number of 'General Commercial' classifications (see Section 5.4.2) where a detailed classification attribute has been omitted.

Table 5 shows results of the survey and as expected, errors are higher in the city centre example though differences with the suburban high street are fairly minimal and there are significant errors within both streets. For residential property identification very few errors were found and this is a positive sign for the application of the data to residential classification. For commercial properties however there are a high number of classification errors in various forms. The proportion of 'General Commercial' addresses is high, there are some temporal errors, and additionally there are a small but significant number of directly false classifications in the 'Basefunction' attribute.

The Address Layer 2 data used in this analysis is from 2008, and so some temporal errors are to be expected compared to the 2009 manual survey. Generally the currency of the Address Layer 2 data appears to be good. The high number of 'General Commercial' and 'Basefunction' errors are more serious problems for the functional classification, as these will directly translate into errors in the final classification. The 'Basefunction' errors appear to be caused by the method of string matching business names against keywords that is used in the creation of Address Layer 2 classifications. For example, a business name including the phrase 'Estate Agent' would be classed as an Estate Agent with this method. Unfortunately this technique is error prone. For instance the key words of 'Press', 'Communications' and 'Workshop' have been used to classify businesses as industrial, but this produces commission errors where small offices, print shops and electronics stores are incorrectly classified as industrial. This was the most common 'Basefunction' error found, particularly in the Tottenham Court Road example. There are similar errors caused by the commercial nature of business names. For example, a

‘Television Entertainment Centre’ is classed as an Entertainment Venue, and the ‘Church of Scientology’ is classed as a Church. This highlights the limitation of string matching without manual validation.

			Classification Errors (Addresses)			
Street	Total Addresses	Missing Addresses	Residential Classification Error	Business Name (Temporal)	Incorrect Basefunction	General Commercial Classification
Tottenham Court Road	351	9	1 (0.3%)	28 (7.7%)	17 (4.7%)	129 (35.8%)
Station Road Edgware	78	0	0 (0%)	4 (5.1%)	2 (2.6%)	16 (20.5%)

Table 5: Functional Classification Validation Survey- Attribute Errors.

A final problem encountered was that some addresses were missing in the Tottenham Court Road survey. This was mainly a temporal error, as some buildings had been redeveloped with subdivided premises creating new addresses. There were however some other significant missing addresses, particularly underground stations. This indicates some incompleteness in the data.

Overall the validation survey points towards the Address Layer 2 data being accurate in distinguishing residential and non-residential properties, but being unreliable as a detailed source of commercial classification data with a high proportion of missing attributes and a small but significant percentage of incorrect classifications. The data is still very useful for residential classifications and more aggregate visualisations of land use, but should not be used in detailed analysis without being aware of the various errors in the commercial classifications.

6.1.2: Geometrical Errors

The manual surveys shown in Table 5 include variables for measuring geometrical generalisation and errors. This is intended to explore to what extent geometrical measures of commercial properties can be estimated from this data, and how 2D generalisation may lead to errors in the algorithm linking address data to building polygon footprints (see Section 5.2).

The survey recorded several measures including the number of Address Groups that combine multiple functions within a single building polygon; the number of times 2D generalisation is present due to vertical changes in the geometry of premises; and finally the number of major 2D generalisation errors found where a clear misrepresentation in the geography of function has occurred. These are all highlighted in Table 6.

		Geometrical Generalisations and Errors (Address Groups)		
Street	**Total Address Groups**	**Multiple Functions**	**2D Geometry Generalisation**	**2D Geometry Major Error**
Tottenham Court Road	85	76 (89.4%)	27 (31.7%)	3 (3.5%)
Station Road Edgware	32	29 (90.6%)	6 (18.8%)	0 (0%)

Table 6: Functional Classification Validation Survey- Geometrical Errors.

The main result that stands out from Table 6 is that the vast majority of buildings from the surveyed streets are mixed-use buildings (around 90%), typically combining retail and office or retail and residential functions. This means essentially that geometrical measures are limited to assessing the footprint of buildings in this model. The vertical distribution of uses in multi-storey buildings is not known from the 2D topographic data, so measures such as floorspace and volume cannot easily be estimated.

The remaining survey measures describe related 2D generalisation issues. For several types of buildings the geometry of property divisions changes in the vertical dimension. A common example is found in large commercial buildings where open-plan offices occupy the first floor and above, while retail premises with smaller property divisions are located on the ground floor. In this case the OS MasterMap data records the larger office outline rather than the ground floor retail divisions. This type of arrangement accounts for over 30% of Address Groups in the city centre survey and over 18% of the suburban centre survey. For these buildings the footprint area cannot be assessed accurately from the topographic data.

Occasionally the 2.5D generalised data will lead to a poor match between the address data and the building geometry. This occurs when the address delivery point is not coincident with the building polygon it relates to. This error was fairly rare (found three times in the surveys) yet it further increases uncertainty in the geometrical measurement of commercial premises.

In summary the vast majority of commercial buildings in the surveyed streets are of mixed-use, often with vertical variation in the geometry of premises, and this greatly limits the ability to make GIS based measures of commercial property size using 2D topographic data. The particular case study of London where densities are high and mixed use buildings more prevalent exacerbates this issue. This has several implications. Firstly for analyses that require geometrical quantification of commercial properties, this cannot be accurately estimated from the 2D methods used here and other quantification data (such as from the VOA surveys) is required. Secondly for the visualisation of urban function, the majority of commercial buildings are mixed-use so methods to sub-categorise these mixed-use buildings (as in Section 5.4.3) are needed if commercial functions are to be visible in map form.

6.2: Residential Classification Validation

The residential classification results can be validated against dwelling type data from the 2001 census. The aim of the validation process is to assess the degree of error in the definition of dwelling types in the Built Environment Model. Classification errors could result from issues such as trivial building links (see Section 5.4). In addition to accuracy issues, contrasts in the methods used to create the datasets are likely to result in some discrepancies. In the 2001 census the householder identifies their residential type, and so there may be differences between definitions commonly used by residents and the geometrical definitions that form the basis of the Built Environment Model. One example for instance could be classifying an end of terrace as a semi-detached. As in all censuses, the 2001 census is affected by under-enumeration. It so happens that London, particularly Inner-London, has by far the highest under-enumeration of any Government Office Region UK with a 10% lower response rate than the average (Office for National Statistics, 2006). The Office of National Statistics have modelled the characteristics of those missing individuals to improve accuracy, but this process will have errors and will likely affect the dwelling type data (Orford and Ratcliffe, 2007).

To make the Built Environment Model data comparable with the 2001 census, it must be spatially aggregated to the census zonation used, in this case, to the Output Area. The fine-scale nature of the model makes this process straightforward. A more significant obstacle to data comparability is temporal variation, as the MasterMap data used to form the Built Environment Model is from 2008, while the census is from 2001. Therefore buildings constructed after 2001 needed to be removed from the Built Environment Model. This was achieved using a combination of the temporal attributes stored in OS MasterMap Topographic Layer, and the postcode-based temporal data stored in the All Fields Postcode Directory (AFPD). This process may not remove all the new buildings as the AFPD only includes those developments large enough to require a new postcode, and for the OS MasterMap attributes it is difficult to assess the temporal accuracy.

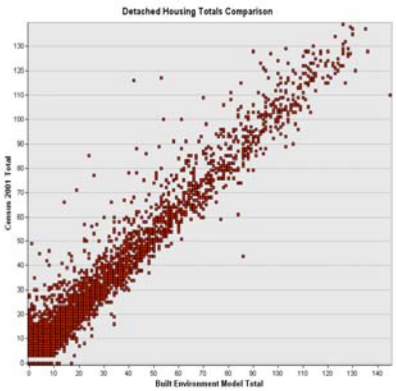
The totals by residential type are shown in Table 7. There is a reasonable correspondence, but two main errors can be seen- an overestimation in the total number of flats and terraced houses, and an underestimation in detached and semi-detached houses. There is a discrepancy of over 250,000 dwellings in the total number of flats. Likely causes of this difference are a failure to remove all new residential developments built after 2001 (as recent new build in London has been overwhelmingly in the form of flats), an over-count of residential functions from the Address Layer 2 data, and errors related to under-enumeration in the 2001 census (which is most prevalent in Inner-London where flats are predominantly located). The validation process carried out here is not detailed enough to gauge the influence of each of these factors and this is therefore an avenue of future work.

Residential Type	Built Environment Model		2001 Census		Difference	
	Totals	%	Totals	%	Absolute	% BEM Total
Detached	141801	4.28	187577	6.15	-45776	-1.38219
Semi-detached	428258	12.93	594906	19.51	-166648	-5.03187
Terraced	1023092	30.89	806301	26.44	216791	6.545922
Flats (inc. maisonettes)	1718697	51.90	1460827	47.90	257870	7.786287
Total	3311848	100.00	3049611	100.00	262237	7.918147

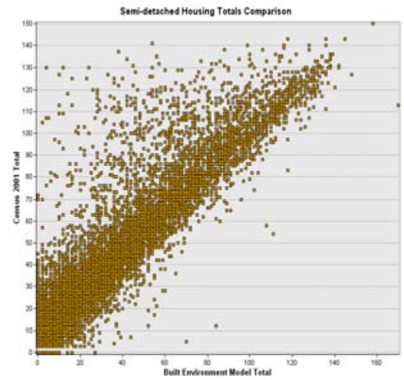
Table 7: Residential Type Totals from Built Environment Model and 2001 Census.

A second trend evident in Table 7 is the under-estimation of detached and semi-detached residences compared to the census, combined with an over-estimation of terraced houses. Figures 10 A-D further highlights this error. This is very likely to result from the trivial building links error discussed in Section 5.4 that will cause detached or semi-detached houses to be misclassified as terraced. In fact the number of over-classified terraced dwellings (216,791) is very close to the number of under-classified detached and semi-detached houses (212,424). Figure 10E sums these dwelling types together and shows a very close correspondence which backs up this conclusion. An improvement in the algorithm to identify these trivial links is therefore a priority for future developments of the model.

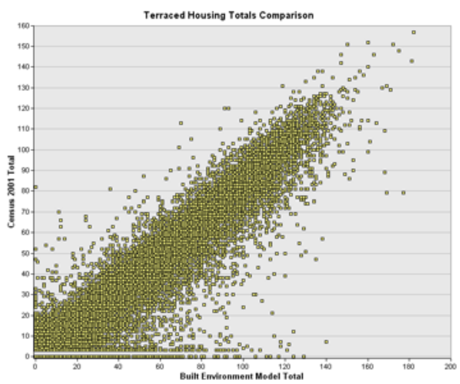
The graphs in Figure 10 highlight the classification errors clearly. Where the Built Environment Model is under-predicting the results are skewed towards the top left with high census results and low Built Environment Model results, as seen in the semi-detached graph Figure 10B. The opposite effect of over-prediction is seen in Figure 10**C** and most strongly in the flats graph Figure 10D**.**



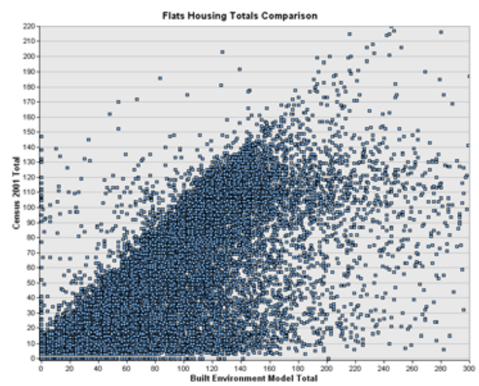
A: Detached



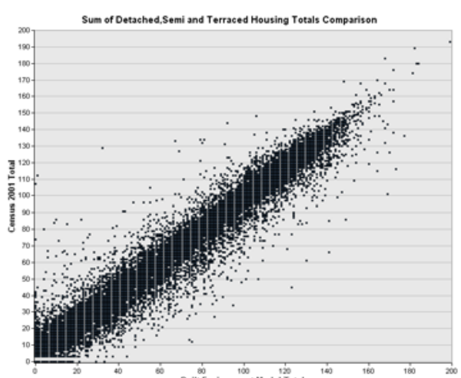
B: Semi-detached



C: Terraced



D: Flats



E: Sum of Detached, Semi-detached and Terraced

Figure 10: Built Environment Model and 2001 Census Residential Type Totals at Output Area Level.

In conclusion the residential classification validation has shown that the Built Environment Model data can be easily spatially aggregated to be compared against demographic data. The validation process highlighted two main errors in the classification. One is the under-prediction of detached and semi-detached houses, misclassified as terraced. This could be reduced using roof line analysis similar to that proposed by Orford (2010), however such a method is computationally expensive for such a large study area. The second issue of a large over-estimation of flats may be due to either temporal comparison issues or errors in the census related to under-enumeration. High quality temporal data is therefore an important feature of the Built Environment Model and postcode-based data and the OS MasterMap attributes may not be sufficient in this regard. Using data sources such as planning permissions, one may be able to augment the temporal attributes of the model to reduce such errors.

7: Analysis of the Built Environment in London

With the data infrastructure for the London Built Environment Model in place, and the accuracy of the classifications assessed, it is now possible to use the model to explore and analyse spatial patterns in urban form and function. Two main examples are provided here to illustrate the capabilities of the model, the first focussing on spatial patterns in urban function, and the second exploring spatial patterns in residential types and density in London.

The urban function and dwelling type data sets are extremely rich and detailed. A comprehensive analysis of the data is not provided here, rather we wish to illustrate the potential applications of such data sets that the methodology has produced. We use visual exploratory analysis techniques to gain an impression of general distributions and patterns. This visual analysis is intended to form the basis of more rigorous statistical analysis in future research. Furthermore, we are interested in exploring general questions regarding how useful the datasets are for enhancing the application of city models to urban planners and researchers. Does the addition of built environment data greatly add to the understanding of urban context and geography, as compared to typical thematic mapping approaches? What scales of visualisation and aggregation are most appropriate for particular applications?

The latter issue of scale is of interest as the model enables multiple scales of output so visualisations and analysis can be tailored to applications. Urban processes generally operate at multiple scales with interactions between local and city-wide levels (and indeed beyond to national and global scales). The Built Environment Model should be well suited to exploring the multi-scale nature of urban phenomena.

7.1: Urban Function Analysis

The functional classification data created by processing Address Layer 2 attributes (see Section 5.4) produces a rich dataset of urban functions at the building level (although with significant errors for commercial property data - as we discussed in Section 6.1). This data can be explored at a variety of scales, and we begin the analysis at the local urban scales and then move up through larger scales to city-wide analysis.

At micro scales the data is of relevance to local planning and urban design applications concerned with issues such as urban massing, public space and local accessibility. For these applications the context of built form is significant. The addition of topographic mapping data such as MasterMap augments the visualisation in this regard. Building heights data also contributes to urban context, as patterns of building use and form are highlighted. Figure 11 shows an example from the Isle of Dogs in Inner London, with commercial high rise buildings at Canary Wharf, surrounded by lower rise residential and commercial clusters. The lack of sub-building data prevents the vertical classification of buildings, and mixed-use categories are used (see Section 5.4.3). The visualisation of 2.5D building heights is difficult to employ over larger urban areas. This is partly a software problem, as GIS software is designed primarily for 2D visualisation with limited 3D capabilities, and partly a visual clarity problem, as intelligibility decreases as the study area expands.

To explore larger urban extents, we felt that 2D visualisation was more appropriate. As noted in Section 2 larger scales of analysis are more directly relevant to geographical research, as spatial patterns in function correspond to the agglomeration of commercial and service activities. We illustrate such patterns of urban function found in London in Figures 12 through to 15. The majority of the city is suburban in character such as that shown in Figure 12, which is dominated by residential functions, with small scale local retail and service centres. Schools (shown in pink) serve local residential populations and, in contrast to the clustering of commercial activities, display negative spatial autocorrelation, generally locating in quiet suburban areas away from busier centres.

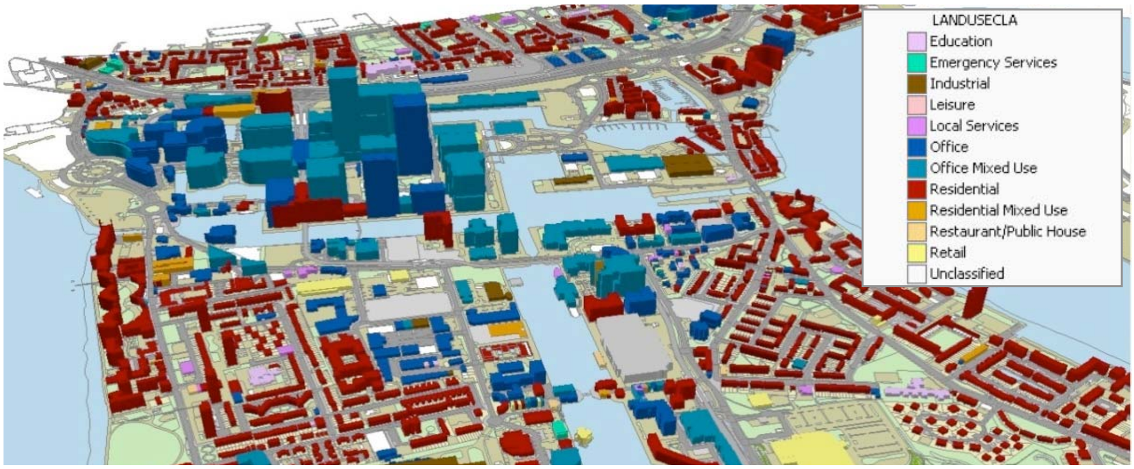


Figure 11: Building Function and Land Use Classification Applied to Tower Hamlets.

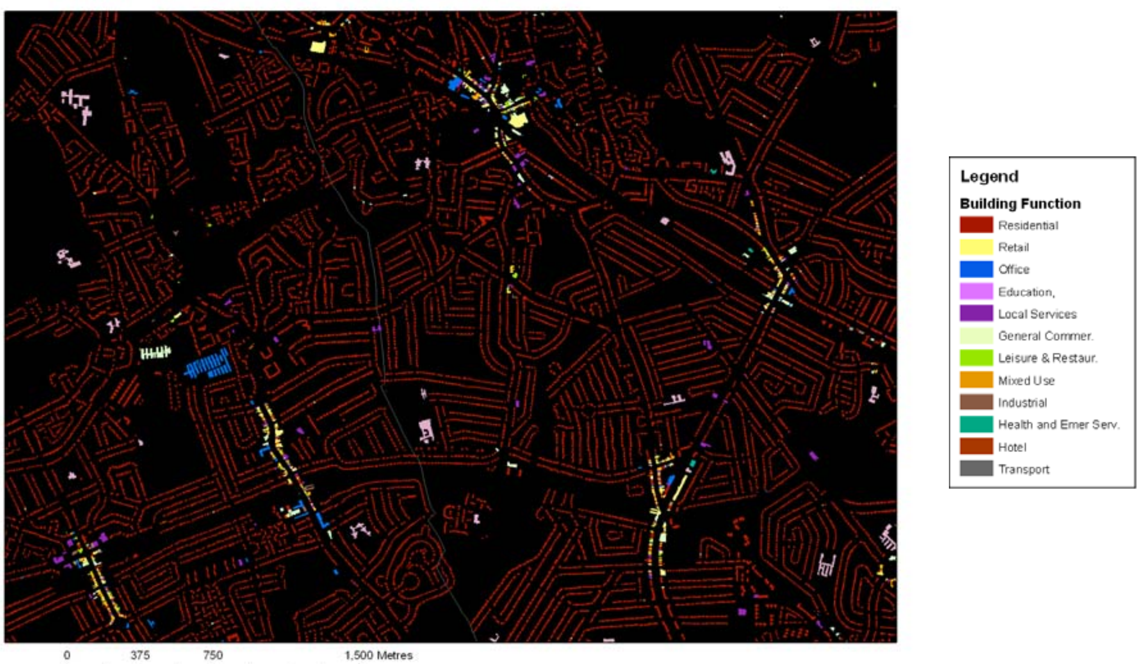


Figure 12: London Suburbs Function Map (North-West London).

Metropolitan centres such as Croydon shown in Figure 13 feature larger scale clustering of commercial and service activities, and offer a wider range of functions including office activities, shopping centres, leisure activities and higher level education and health services. Figure 13 in fact combines two types of agglomeration commonly found in Outer London. The first, in the centre of the map is a metropolitan centre, Croydon, which is fine grained with small scale patterns of functional diversity. Secondly, to the west is an extensive retail and industrial park with large dispersed building footprints. The combination of metropolitan centres with industrial parks is fairly typical in Outer London, and other centres with similar characteristics can be found at Wembley and Stratford. With continued deindustrialisation, these parks are diversifying into retail and office functions, and remain highly car dependent.

As we move to the inner-city as shown in Figure 14, functional diversity increases with inner-city centres linked by linear clusters of activity along major roads (or at least the major roads that function as streets). This is most visible in London along radial routes such as the historic Edgware Road (A5) and Kingsland Road (A10), but occurs at varying densities on a great number of links. The attraction of commercial functions (primarily retail) to high accessibility locations is a core part of theory in urban street network configuration research (see Hillier, 1996). The techniques and data developed here could be a useful evidence base to test such theories, though it is unlikely to support any straightforward relationship between accessibility and urban function. There are numerous complexities in relationships between accessibility and location, as different commercial functions have varying needs in terms of agglomeration, transport mode priorities, labour accessibility, customer accessibility, and ability to meet rental costs.

In the city centre of London the diversity of function and density of urban grain reaches its peak. Figure 15 shows the functional pattern in Central London. The mix of functions is so fine-grained that the map appears as a dense multi-coloured patchwork. Agglomerations can be seen such as the dominance of blue coloured office activities in the City of London (to the east) and greater frequency of retail and leisure functions in the West End[10].

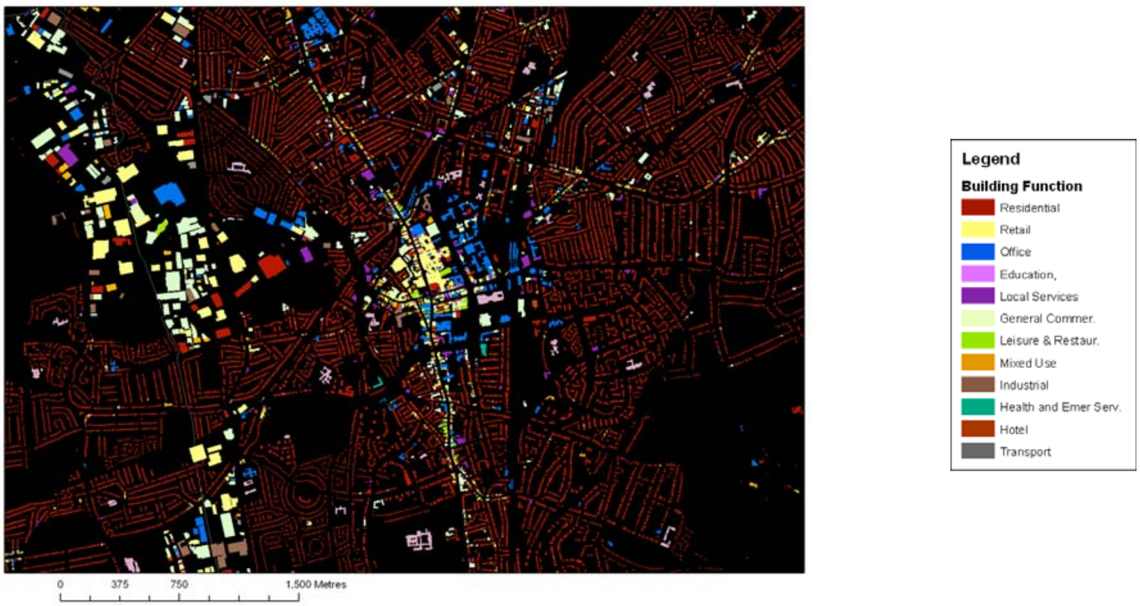


Figure 13: Metropolitan Centre and Business Park Function Map (Croydon).

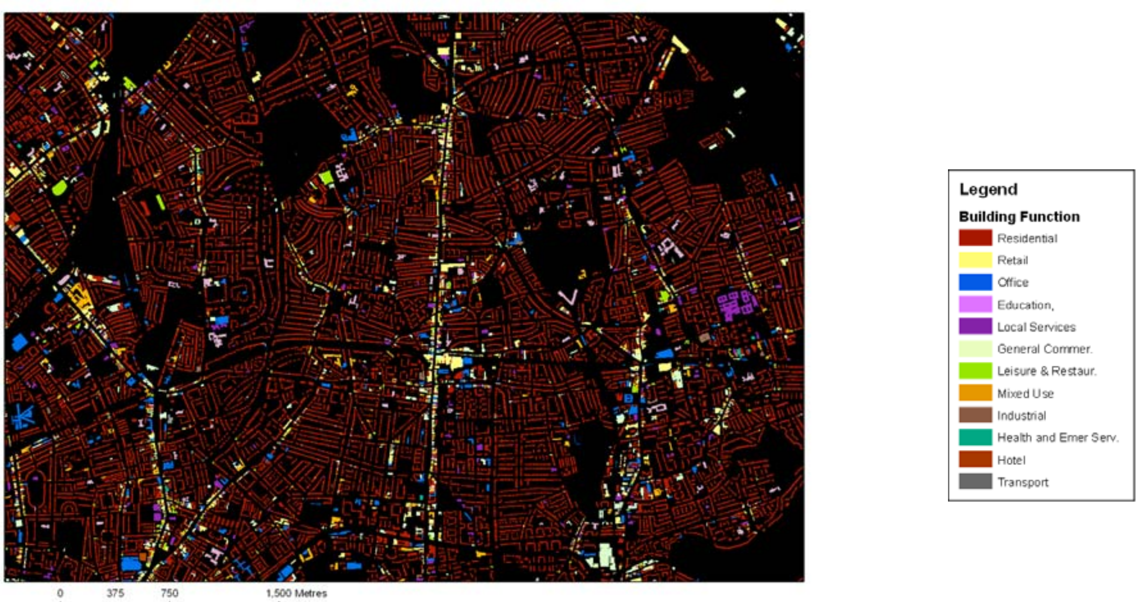


Figure 14: Inner-City Function Map (North London- Kingsland Road).

[10] The exact definition of the West End varies depending on the data source used however; many of the definitions include Oxford Street, Regent Street and Bond Street in the term.

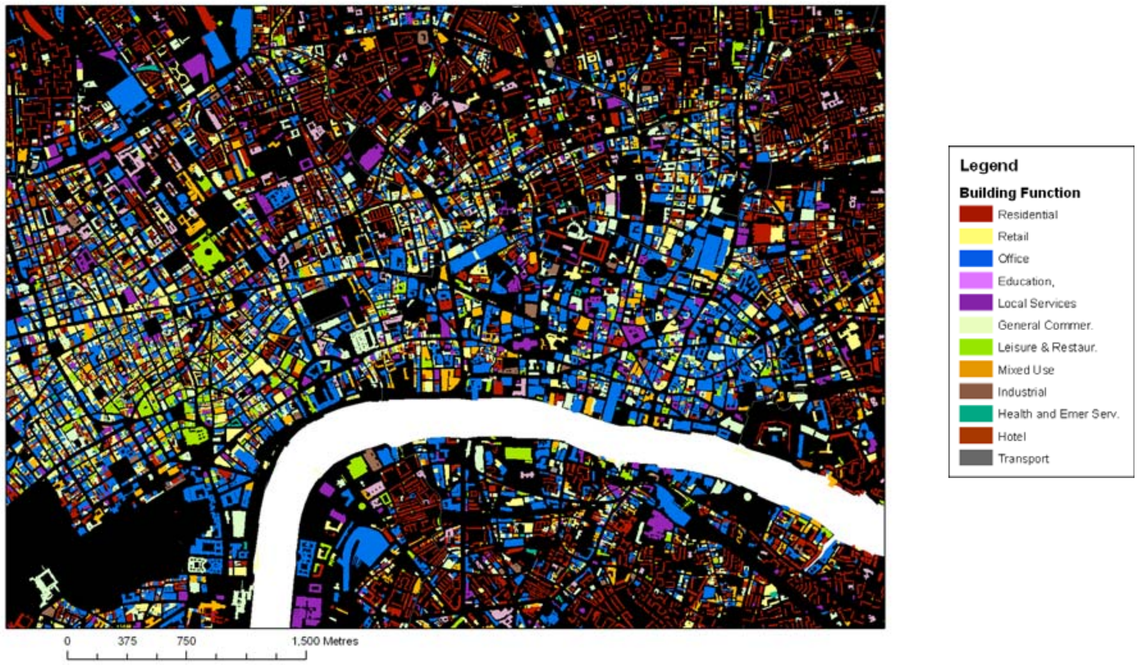


Figure 15: City Centre Function Map (West End and City of London).

We can now move up from meso-scale analysis to visualising the distribution of urban function across the whole of London. The fine grain diversity of function identified in Figures 12 through to 15 is problematic for city-wide visualisations as patterns are too intricate to be intelligible to the viewer. There are several approaches to solving this problem, such as simplifying the classification scheme, aggregation methods and statistical indices. Here we pursue the classification simplification method (aggregation techniques are used with the residential data in Section 7.2).

Figure 16 shows the building level function data for the whole of London, classified into three groups: commercial, local services and residential. This basic classification method greatly improves the legibility of the map, allowing macro scale patterns across the whole of London to be identified. The graphic technique of using a black background also improves clarity. It is interesting that this city-wide visualisation is possible without performing any explicit data aggregation processes. The GIS software (ESRI ArcMap 9.3) is itself is effectively performing some aggregation through its graphics engine, particularly with respect to how the different classes of features are overlaid in the map. Changing the hierarchy of feature classes in the symbology options changes the map output. In Figure 16 the symbol hierarchy used is reflected in the Legend, with commercial and local service classes prioritised over residential.

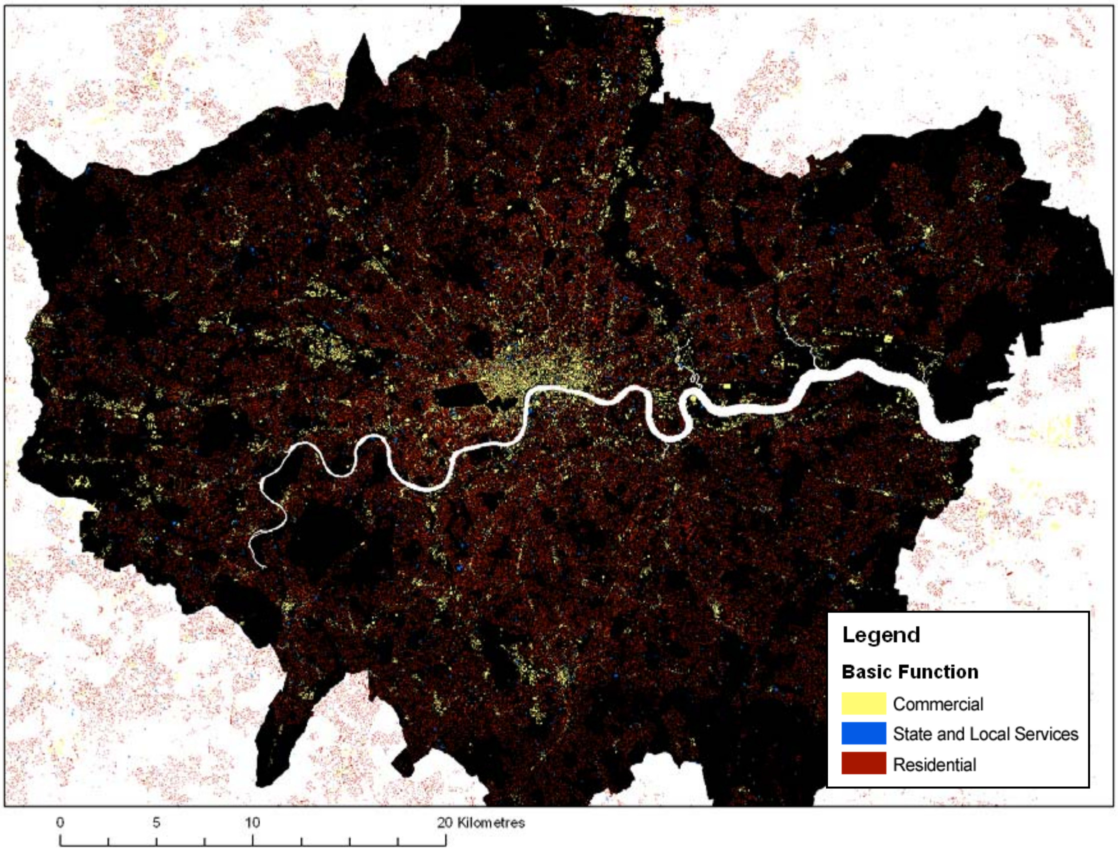


Figure 16: Macro-Scale Basic Functional Map of London.

Macro-scale patterns of commercial activities can be clearly seen in Figure 16, with commercial centres at various scales linked by a network of linear agglomerations along major roads. The main hub of Central London forms by far the largest commercial agglomeration, while local centres, business and industrial parks form smaller outer clusters, linked to the central hub through a network of radial routes. These radials are highlighted by linear commercial activities, producing a 'hub and spoke' pattern.

The disadvantage of grouping all the commercial activities together (and of the incomplete commercial classification in the underlying data) is that we cannot distinguish between types of commercial centre (which was possible to a degree at the meso-scale analysis). In Figure 16 industrial parks tend to have added prominence due to the large building footprints. For example, Park Royal to the west of London which appears as the largest cluster outside of the city centre. Aggregation methods would give a greater degree of control and clarity over the macro-scale analysis, and would be necessary for any aggregate statistical analysis.

In summary, visualisation of the urban function data clearly identified spatial patterns in land-use at a series of scales. Micro-scale visualisations relate to local building configurations and were most practical with the added context of topographic data and building heights. The lack of sub-building data is a shortcoming at this scale. Meso-scale analysis identified very clear patterns in the clustering of commercial functions, and has potential for use in geographical research. The macro scale visualisations showed the potential to visualise function over an entire city, highlighting large scale agglomeration patterns. This required simplification of the functional classification to make the map legible to the viewer.

7.2: Residential Typology and Density Analysis

Spatial patterns in dwelling types are the result of both micro and macro scale urban processes. The clustering and configuration of dwelling types is influenced by the nature of the developer (for example, state or private sector), the era of development, local geography, the wider historical evolution of local centres and transport infrastructure (particularly roads), and conversions through densification and gentrification processes. These local actions result in macro scale patterns as the attraction of major employment and service centres which in turn increases housing demand and increases residential density in accessible locations while suburban locations can offer larger properties and gardens. These processes were originally theorised in Alonso’s (1964) influential Bid Rent model.

The level of detail provided by the Built Environment Model highlights the great diversity and complexity of spatial patterns of housing in London, with small scale clustering leading to a wide range of housing types in close proximity. Some areas include the full selection of dwelling types from detached houses to flats within a few hundred metres as highlighted in Figure 17. At the meso-scale a patchwork of housing clusters is clearly evident dotted with local centres and parks as shown in Figure 18. Higher density flats can be clearly seen following major roads, indicating local accessibility effects.

Figure 17: Local Scale Diversity of Housing Types in Outer London.

While local diversity and clustering of housing types is clearly apparent, this does not outweigh the strong macro scale trend of a density gradient from the city centre to the low density outskirts. Figure 19 visualises the data at a 50m grid scale, with the most frequent residential building type (based on counts of buildings rather than dwellings) within each grid square shown. This is an example of a grid based aggregation technique being used to simplify the fine-scale built environment data (see Section 2.3). As one would expect flats dominate the city centre and are prevalent in the inner city. Terrace housing takes up the largest area, bridging between the inner-city and the suburbs. Major areas of lower density detached and semi-detached housing are found beyond the inner-city. There are small areas that break these general trends, such as outer town centres with local clusters of flats distant from Central London, but generally the monocentric housing density gradient is evident throughout London.

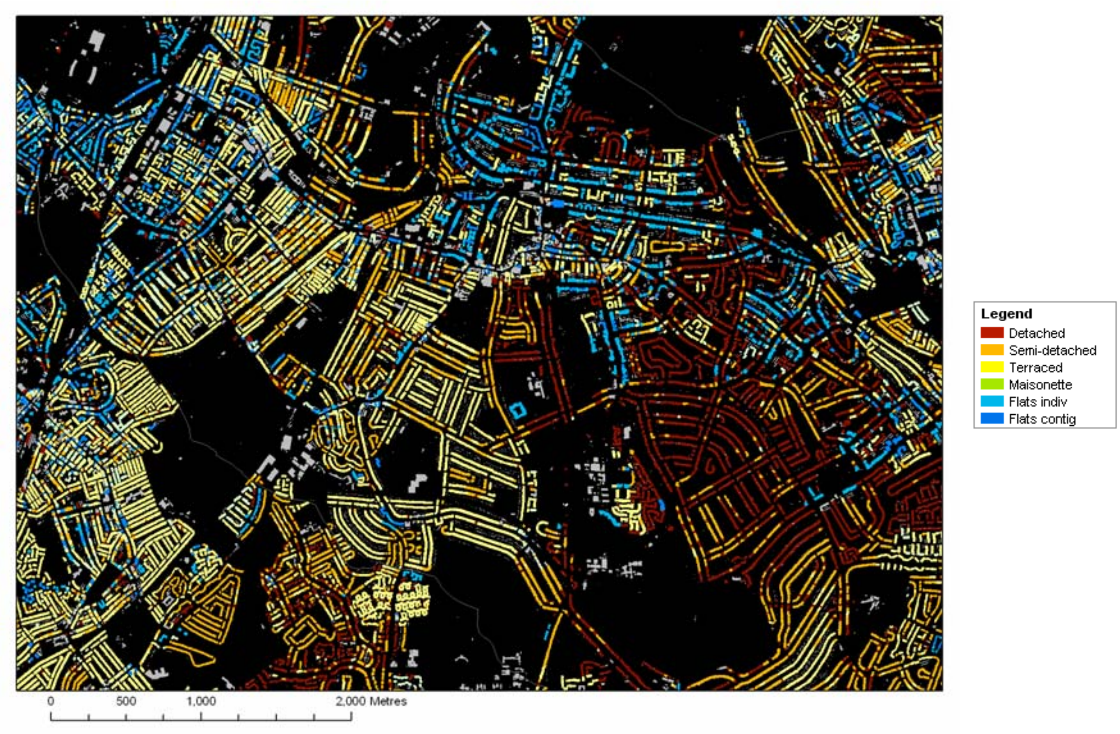


Figure 18: Meso-Scale Neighbourhood Clustering of Housing Types in South London.

This density gradient can be highlighted in graph form. Figure 20 shows mean residential density on the y-axis and distance from the city centre[11] on the x-axis. The graph displays several characteristics in accord with Alonso's (1964) Bid Rent theory. Firstly at the very centre of the city, residential density is low as commercial activities outbid residential functions in the most accessible locations. Outside of the commercial centre residential density peaks at five kilometres from the centre and then declines sharply, tailing off with an inverse distance relationship. The density decline is also reflected in the change of housing types with the prevalence of higher density flats declining with distance, while

[11] There is no definitive central point for measuring distances in London. By historical convention Charing Cross is used. This falls between the central districts of the City of London, Westminster and the West End, and so we have followed convention in measuring from this point.

lower density detached and semi-detached housing takes a larger share in suburban areas. This change in housing type is more clearly highlighted in Figure 21 where the total number of dwellings by distance is shown. Terraced housing is the largest group and is prevalent over a greater range of distances compared to other housing types.

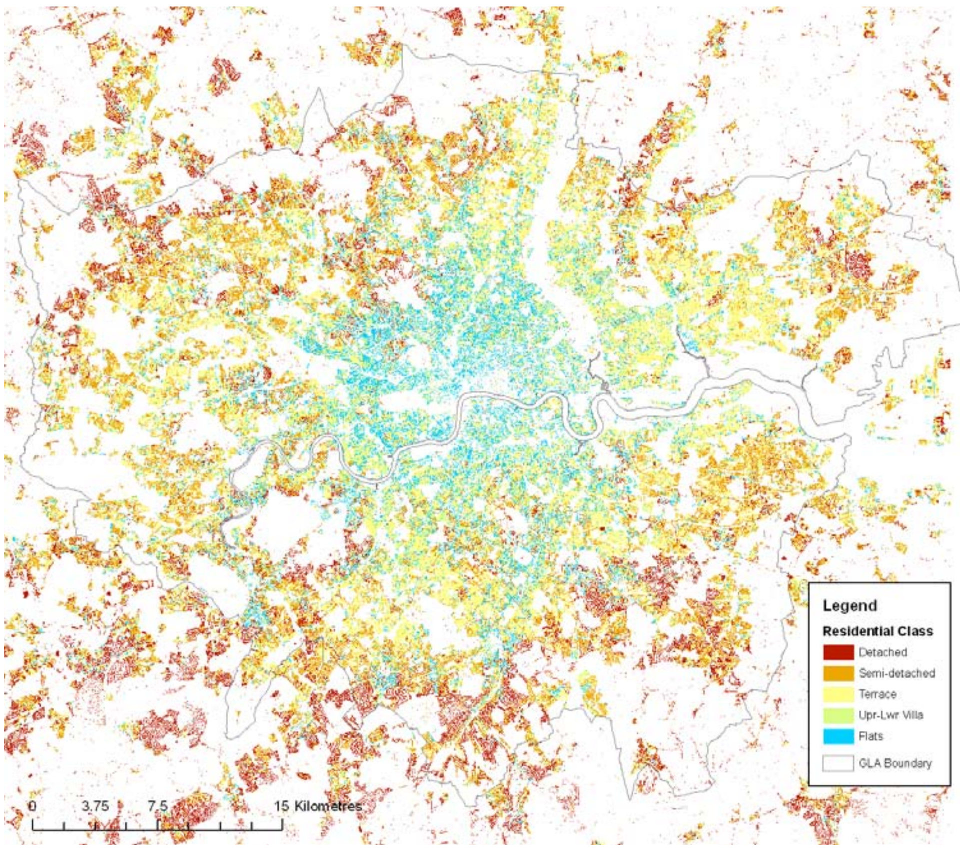


Figure 19: Most Frequent Residential Building Type in London, 50m Grid.

Alonso's (1964) Bid Rent theory provides a useful framework to describe the macro-scale patterns of residential type and density in London, highlighting the continued dominance of the historic monocentric structure. This is however a rather static view and it would be interesting to consider whether more recent urban development is reinforcing the historic structure or if new trends are evolving. There are current debates surrounding the densification of the suburbs in London, which implies that denser housing types are becoming more prevalent in outer London locations. Temporal data linked to the built environment database could enable this kind of analysis, for example, through planning permission data.

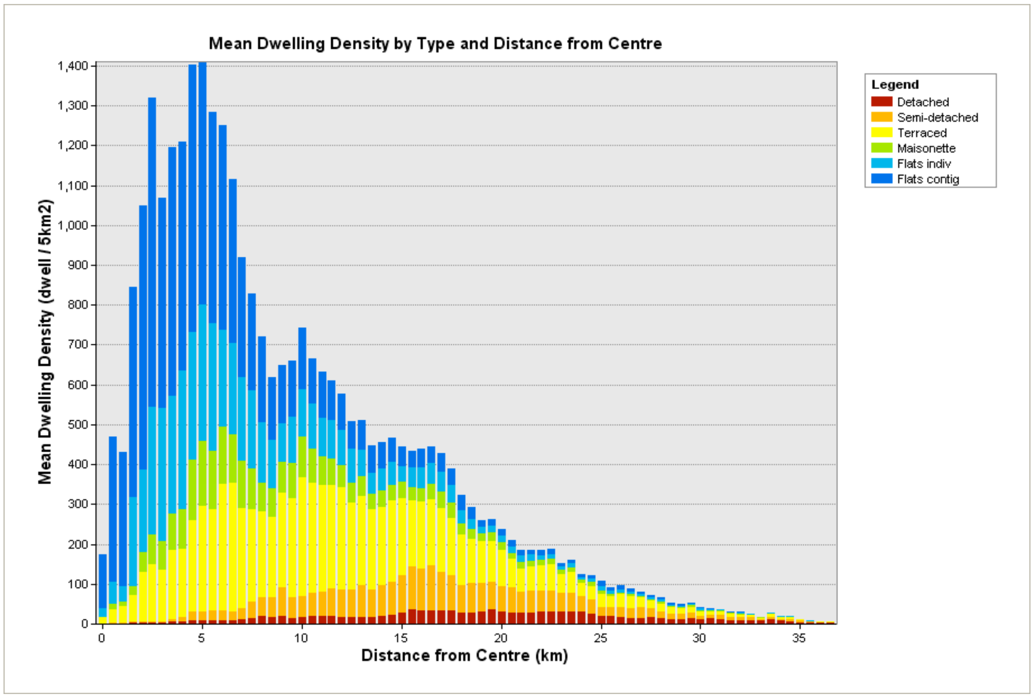


Figure 20: Graph of Mean Dwelling Density by Type and Distance from Centre.

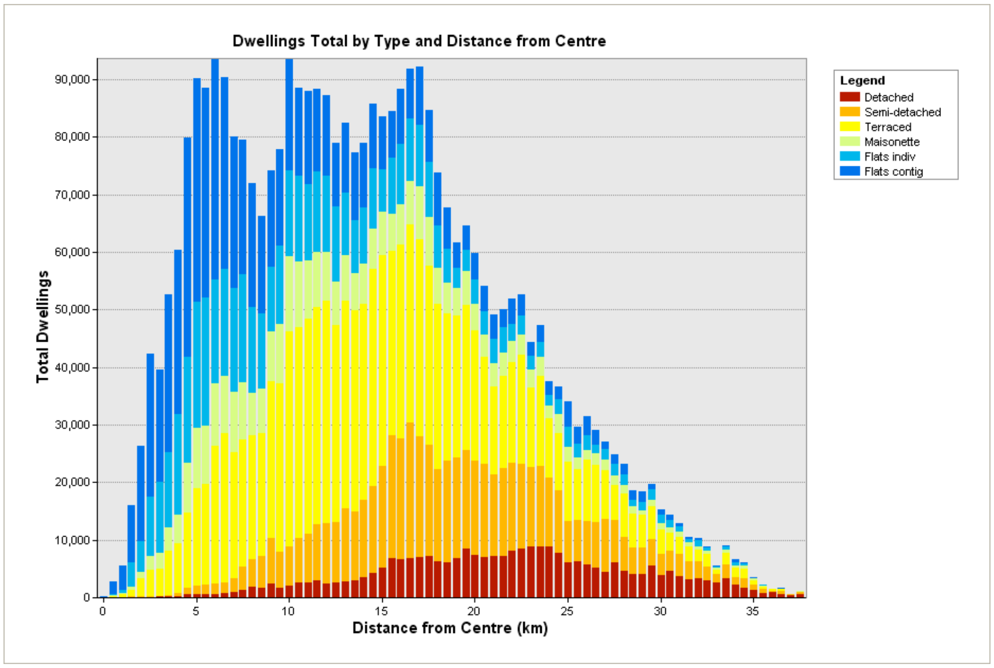


Figure 21: Graph of Total Dwellings by Type and Distance from Centre.

In addition to new build, another significant process in housing dynamics is building conversion. Micro scale patterns in dwelling type highlight a great variety of probable conversions that have occurred, particularly of terraced housing into flats and maisonettes as highlighted in Figure 22 (where yellow is for terraces, blue is for flats). These patterns indicate that densification processes are frequent and widespread. Gentrification can later result in the reversal of these trends. A more comprehensive temporal analysis than is provided here is required to understand these patterns, and this is a promising area for future research.

In summary the residential typology data can be applied in a range of analyses at both micro and macro scales. In the London context there is both highly complex micro scale clustering of housing types, and a clear macro scale monocentric pattern of declining densities from the centre. The visualisation and aggregation methods employed are successful in translating between these scales. The analysis could be enhanced with more detailed temporal data to explore evolution over time. Many data sets, such as house sales and geodemographic data, would be interesting to analyse in combination with the residential typology data and we intend to explore these in future research.

Figure 22: Probable Conversions of Terraced Housing.

8. Conclusions

This research has successfully demonstrated the viability of developing a Built Environment Model of a large city with micro-scale building level data. The flexibility and power of this approach are highlighted in the ability to integrate datasets together at an address level, and to perform spatial analysis on these datasets to create the functional and residential classifications described. This flexibility also extends to how the data is output, with visualisation and aggregation at multiple scales illustrated. This approach is a complementary method to the traditional aggregate approach in quantitative geography and overcomes several of the problems associated with scale and the lack of direct built environment representation in geographical analysis.

A host of research opportunities are possible with such a model, particularly around relationships between urban form, function and accessibility; detailed analysis of residential property markets, and analysis of urban change and development. The analysis section of this paper has introduced patterns of form and function in London, highlighting both micro and macro scale clustering of urban functions and residential types.

While such fields have great potential for the application of the Built Environment Model, this paper has also identified challenges to the geography and geometry approach. These issues are principally the limitations of 2D data, data accuracy, algorithm complexity and difficulties with temporal data.

The ability to link geometrical data to other address-based data sources hinges on linking spatial addresses to topographic mapping data. The OS MasterMap address and topographic layers have been key to this process, and this research has shown how the data model for integrating addresses and buildings can be extended. There are however some fundamental limitations with using 2D topographic data as vertical changes in premises geometry are missing; therefore the size of premises in mixed-use buildings (the majority of commercial buildings) is not known. Solutions to this problem can either involve using 3D data (which is not widely available for the UK) or introducing attribute based measures of premise size, such as from the VOA surveys. The integration of the VOA data could also help with the incomplete commercial classifications in the Address Layer 2 data.

The algorithms used in this research to assess address geometry and classify residential types are based on straightforward adjacency relationships. This approach has limitations as shown in the residential classification errors (See Section 6.1). More advanced algorithms including some morphological techniques such as roofline analysis would improve these algorithms and should be a priority in future advances of the London Built Environment Model. These advanced algorithms will have to be computationally feasible for the large extend of the study area.

Finally the comparison of the model with the 2001 census and the analysis of residential form patterns (Section 6.2) have highlighted the importance of temporal data. While the Built Environment Model has great spatial flexibility this is not yet mirrored in its temporal representation. Neither the OS MasterMap temporal attributes (which are incomplete for buildings built earlier than the MasterMap release in 2001) nor the postcode-based AFPD approach are sufficient in this regard. Future research should investigate how richer temporal data can be integrated with the Built Environment Model, possibly through data sources such as planning permission data.

Acknowledgements

The authors would like to thank the Greater London Authority Economics Unit especially Margarethe Theseira, without which this work would not have been possible. We also wish to thank the following data providers for this research: the Ordnance Survey, Infoterra and the Greater London Authority.

References

Aguilera, A. (2005), 'Growth in Commuting Distances in French Polycentric Metropolitan Areas: Paris, Lyon and Marseille', *Urban Studies,* 42(9): 1537-1547.

Alonso, W. (1964), *Location and Land Use: Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, MA.

Atisreal (2005), *Property Value Study - Assessing the Change in Values Attributable to the Jubilee Line Extension*, Transport for London, London, UK, Available at http://www.tfl.gov.uk/assets/downloads/JLE-Final-Report-May-2005.pdf.

Banister, D. (2005), *Unsustainable Transport: City Transport in the New Century*, Routledge, London, UK.

Batty, M. (1976), *Urban Modelling: Algorithms, Calibrations, Predictions*, Cambridge University Press, Cambridge, UK.

Batty, M. (2000), 'The New Urban Geography of the Third Dimension', *Environment and Planning B,* 27(4): 483-484.

Batty, M. (2007), 'Model Cities', *Town Planning Review*, 78(2): 125-178.

Batty, M., Chapman, D., Evans, S., Haklay, M., Kueppers, S., Shiode, N., Hudson-Smith, A. and Torrens, P.M. (2001), 'Visualizing the City: Communicating Urban Design to Planners and Decision-Makers', in Brail, R.K. and Klosterman, R.E. (eds.), *Planning Support Systems: Integrating Geographic Information Systems, Models and Visualisation Tools*, ESRI Press, Redlands, CA, pp. 405-443.

Batty, M. and Hudson-Smith, A. (2002), 'Virtuality and Cities: Definitions, Geographies, Designs', in Fisher, P. and Unwin, D. (eds.), *Virtual Reality in Geography*, Taylor & Francis, London, UK, pp. 270-291.

Batty, M. and Hudson-Smith, A. (2005a), *Imagining the Recursive City: Explorations in Urban Simulacra*, Centre for Advanced Spatial Analysis (University College London): Working Paper 98, London, UK.

Batty, M. and Hudson-Smith, A. (2005b), 'Urban Simulacra', *Architectural Design,* 75(6): 42-47.

Brail, R.K. and Klosterman, R.E. (eds.) **(2001)**, *Planning Support Systems: Integrating Geographic Information Systems, Models and Visualisation Tools,* ESRI Press, Redlands, CA.

Bruhns, H., Steadman, P. and Herring, H. (2000), 'A Database for Modeling Energy Use in the Non-Domestic Building Stock of England and Wales', *Applied Energy,* 66(4): 277-297.

Clarke, M. and Holm, E. (1987), 'Microsimulation Methods in Spatial Analysis and Planning', *Geografiska Annaler. Series B, Human Geography,* 69(2): 145-164.

Davidson, M. and Lees, L. (2005), 'New build 'gentrification' and London's riverside renaissance', *Environment and Planning A,* 37: 1165-1190.

de Smith, M.J., Goodchild, M.F. and Longley, P.A. (2007), *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools (2nd Edition),* The Winchelsea Press, Winchelsea, UK.

Department for Communities and Local Government (2005), *Towards the National Spatial Address Infrastructure - Outline Prospectus*, The Department for Communities and Local Government, London, UK.

Fotheringham, A.S., Brunsdon, C. and Charlton, M. (2002), *Geographically Weighted Regression & Associated Techniques*, John Wiley and Sons, Chichester, UK.

Galster, G. (2001), 'On the Nature of Neighbourhood', *Urban Studies*, 38(12): 2111–2124.

GLA (2002), *London's Skyline, Views and High Buildings*, The Greater London Authority, London, UK, Available at http://www.london.gov.uk/mayor/planning/docs/tr19_high_bdgs_lowres.pdf.

GLA (2004), *Defining and Analysing London's Housing Submarkets*, The Greater London Authority, London, UK, Available at http://www.london.gov.uk/mayor/economic_unit/docs/workingpaper_07.pdf.

GLA (2006a), *DMAG Briefing 2007/32 – GLA 2006 Round Demographic Projections*, The Greater London Authority, London, UK.

GLA (2006b), *Draft Alterations to The London Plan: Spatial Development Strategy for Greater London*, The Greater London Authority, London, UK, Available at http://www.london.gov.uk/mayor/strategies/sds/lon_plan_changes/index.jsp.

Gordon, P. and Richardson, H.W. (1996), 'Beyond Polycentricity: the Dispersed Metropolis, Los Angeles, 1970-1990', *Journal of the American Planning Association*, 62(3): 289-295.

Gröger, G., Kolbe, T.H., Czerwinski, A. and Nagel, C. (2008), *OpenGIS City Geography Markup Language (CityGML) Encoding Standard, Version 1.0.0*

International OGC Standard, Open Geospatial Consortium, Doc. No. 08-007r1 2008, Available at http://www.opengeospatial.org/standards/citygml.

Hillier, B. (1996), 'Cities as Movement Economies', *Urban Design International*, 1(1): 41-60.

Longhorn, R.A. and Blakemore, M. (2007), *Geographic Information: Value, Pricing, Production, and Consumption*, CRC Press, Boca Raton, FL.

Longley, P.A. (2003), 'Geographical Information Systems: Developments in Socio-Economic Data Infrastructures', *Progress in Human Geography*, 27(1): 114-121.

Longley, P.A. and Mesev, T.V. (2000), 'On the Measurement and Generalisation of Urban Form', *Environment and Planning A*, 32(3): 479-488.

Longley, P.A. and Mesev, T.V. (2003), 'On the measurement and generalisation of urban form', *Environment and Planning A*, 32: 473-488.

Lowry, I.S. (1965), 'A Short Course in Model Design', *Journal of the American Institute of Planners*, 31(2): 158-165.

Masser, I. (2005), *GIS Worlds: Creating Spatial Data Infrastructures*, ESRI Press, Redlands, CA.

Moudon, A.V. (2002), 'Thinking About Micro and Macro Urban Morphology', *Urban Morphology*, 6(1): 38–40.

Office for National Statistics (2006), *2001 Census General Report For England and Wales*, Office for National Statistics, Newport, UK, Available at http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14213.

Onsrud, H.J. (ed.) **(2007)**, *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, ESRI Press, Redlands, CA.

Openshaw, S. (1984), *The Modifiable Areal Unit Problem (Concepts and Techniques in Modern Geography: 38)*, Geo-Books, Norwich, UK.

Ordnance Survey (2006), *OS MasterMap Address Layer and Address Layer 2: User Guide (v1.0)*, Ordnance Survey, Southampton, UK, Available at http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/detailedproductinfo/OSMMALandAL2userguide.pdf.

Ordnance Survey (2007a), *OS MasterMap Address Layer 2 Technical Specification (v1.3)*, Ordnance Survey, Southampton, UK, Available at http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/layers/addresslayer2/detailedproductinfo/al2_tech_spec.pdf.

Ordnance Survey (2007b), *OS MasterMap Topography Layer: User Guide and Technical Specification (v1.3)*, Ordnance Survey, Southampton, UK, Available at http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/userguides/docs/OSMMTopoLayerUserGuide.pdf.

Orford, S. (2010), 'Towards a Data-Rich Infrastructure for Housing Market Research: Deriving Floor Area Estimates for Individual Properties from Secondary Data Sources', *Environment and Planning B*, 37(2): 248-264.

Orford, S. and Ratcliffe, J. (2007), 'Modelling UK Residential Dwelling Types using OS Mastermap Data: A Comparison to the 2001 Census', *Computers, Environment and Urban Systems*, 30(2): 206–227.

Pryce, G. and Evans, G. (2007), *Identifying Submarkets at the Sub-Regional Level in England*, Department for Communities and Local Government, London, UK, Available at http://www.communities.gov.uk/publications/planningandbuilding/identifyingsubmarkets.

Rabin, Y. (1987), 'The Roots of Segregation in the Eighties: The Role of Local Government Actions', in Tobin, G.A. (ed.) *Divided Neighbourhoods: Changing Patterns of Racial Segregation*, Sage Publications, London, UK, pp. 208–226.

Skupin, A. (2000), 'From Metaphor to Method: Cartographic Perspectives on Information Visualization', *Proceedings IEEE Symposium on Information Visualization (InfoVis 2000)*, 9th -10th October, Salt Lake City, UT.

Smith, D.A. (2009), 'Polycentricity and Sustainable Development: A Real Estate Approach to Analysing Urban Form and Function in Greater London', *11th International Conference on Computers in Urban Planning and Urban Management*, 16th -18th June, Hong Kong, PLC.

Spooner, B. and Cooper, E. (2006), *Working Paper 20- Employment Projections for London by Sector and Borough,* The Greater London Authority Economics Unit, London, UK, Available at http://www.london.gov.uk/mayor/economic_unit/docs/wp_20_employment.pdf.

Steemers, K. (2003), 'Energy and the City: Density, Buildings and Transport', *Energy and Buildings*, 35(1): 3-14.

Talen, E. (2003), 'Measuring Urbanism: Issues in Smart Growth Research', *Journal of Urban Design*, 8(3): 195-215.

Tuan, Y. (1977), *Space and Place: The Perspective of Experience,* University of Minnesota Press, Minneapolis, MN.

Urban Task Force (1999), *Towards an Urban Renaissance: Final Report of the Urban Task Force,* Routledge, London, UK.

van Oosterom, P., Zlatanova, S., Penninga, F. and Fendel, E. (eds.) **(2008)**, *Advances in 3D Geoinformation Systems,* Springer-Verlag, Berlin, Germany.

Zlatanova, S. and Prosperi, D. (2006), *Large-scale 3D Data Integration: Challenges and Opportunities,* Taylor and Francis, Boca Raton, FL.