B06.1 Mapping the market: Using GIS to visualise the dynamics of property sale

Mark Thurstain-Goodwin, Director, Geofutures

We are a nation obsessed with property. Furtive glances in the estate agent's window to see how our equity has increased; the Sunday morning ritual trip to B&Q; investing the bonus in a buy-to-let. If we're not poring over the broadsheet property pages we're watching Laurence Llewelyn-Bowen telling us that fur is the new lino, or Kevin McCloud tut-tutting at Grand Design self-builders going over budget.

If this isn't enough, the house-buying shows combine every aspect of our obsession: how to choose and purchase your Place in the Sun, your Escape to the Country or the right Location, Location, Location. The last is arguably the best known TV format, offering a snoop around other people's houses mixed with pretty generic buying advice – stay in your budget, buy something that you like, and above all, pick a good location.

It's this last element that is explored here. The themes explored are equally applicable to commercial and residential property, but concentrated on the more generally familiar territory of the housing market.

This paper argues that better understanding of the UK residential property market would have real benefits, and goes on to examine the use of GIS to achieve this, both in theory and in practice. The statistical methods used to create meaningful 'market maps' from available data are discussed, touching on some important limitations of these techniques. While it emerges that geostatisticians may be no more truly objective that estate agents, it is also clear that these are tools with a great deal of potential for helping us to understand our national obsession.

Location, Location, Location – the estate agent's mantra

Ask a property professional the key element of a property value and it is very likely they will conform to the stereotype. Location, location, location really is the estate agent's mantra. The simple but unifying truth underpinning this belief is that property markets vary spatially.

Sometimes considerably. Terraced houses change hands in the pub for £5,000 in the Greater Manchester constituency of MP Andrew Bennet, he recently told the ODPM's Housing, Planning, Local Government and the Regions select committee. In contrast, similar-style housing in the south east is orders of magnitude beyond the reach of key public sector workers.

These imbalances have significant impacts. The state of the housing market is a key consideration in the decision whether and when to join the Euro, and the reason why the Deputy Prime Minister is supporting the construction of 200,000 new homes on the brownfield land of the Thames Gateway.

Residential property value variation is highly relevant to the Valuation Office who have recently started the council tax revaluation, and to those debating property-related funding of public transport infrastructure – of which more later. Building society valuers, meanwhile, live and breathe the spatial variation in price over micro-scale distances, while for us, their customers, a locational mistake eats into the equity of our most important financial assets.

A better understanding of spatial variation in property values would be compellingly useful to all of us.

GIS technology delivering market information

GIS is a location-based technology, which makes it well suited to visualising the way in which residential property values vary spatially.

GI practitioners in the property arena become unusually familiar with Lucas County, Ohio. Here the district auditors office publish detailed information on every residential property in their jurisdiction which is held on their central database, called the Auditor's Real Estate Information System (AREIS).

AREIS provides detailed information on the structure of each house including the year it was built, the number of rooms, number of bedrooms, number of bathrooms, total floor area, garage size, its construction type and size of the land parcel on which the house sits. The purpose of the database is to make available the information that underpins their valuation system, and so the data includes information on the auditors' evaluation of the property and the amount of tax paid for it.

This information is made freely available on the web though a basic GIS interface; a more interactive GIS with additional functionality is provided free of charge on a CD-ROM. A screenshot from the latter, displaying tax information about the property highlighted in green, is shown in Figure 1.

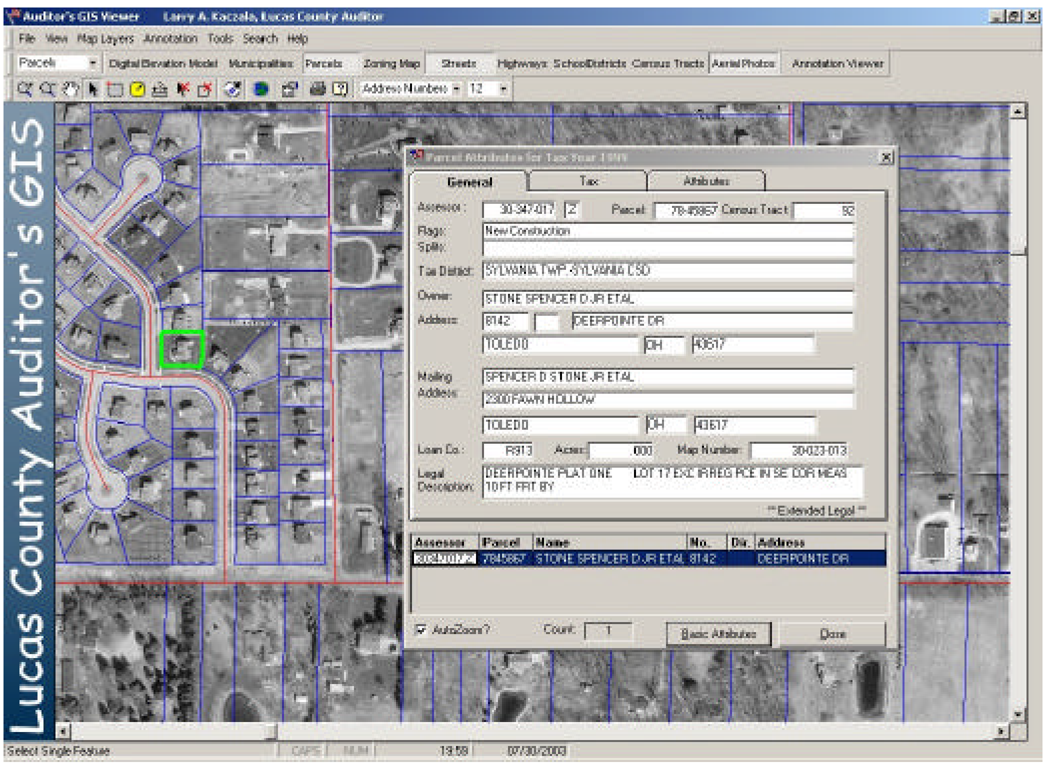

Figure 1 : Tax information screen from the Lucas County Auditor's GIS

As well as recording tax data, information on all property transactions in Lucas County are stored on AREIS. A particularly powerful feature of the CD-ROM GIS is the way it enables the public to gather comparable evidence about property deals that have occurred in a particular area. The wizard enables you to identify all transactions of a particular property type (defined by a series of attributes, as shown below) in the geographical location of the user's choice (Figure 2).

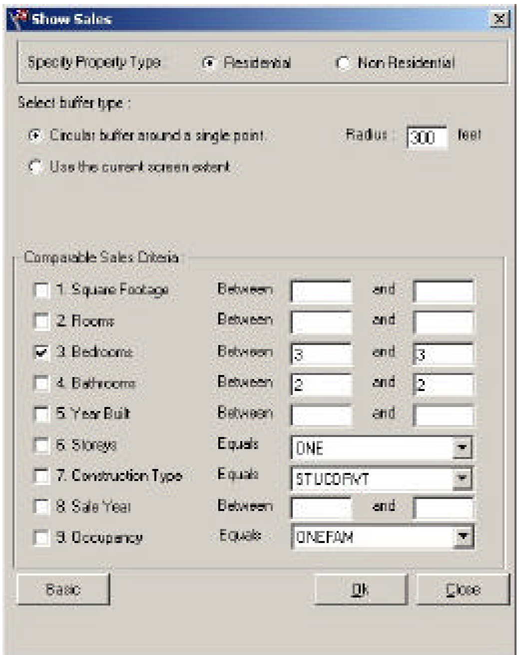

Figure 2: Comparable evidence wizard

This means that the would-be house purchaser can identify all the properties of a certain specification in a locality, as well as getting a view of their current assessed value (which is updated annually). He or she can see the prices of any transactions conducted on those properties going as far back as 1973.

This level of detail and freedom of access to information is unparalleled and would be viewed enviously by housebuyers in the UK. There are a number of benefits. Speaking at the recent AGI event 'GI in Property Tax Assessment', Jerry German, Lucas County's chief assessor, stated that the GIS developed by his office was now integral to the real estate industry in Lucas County, with estate agents and insurance companies regularly consulting the on-line AREIS. This was never more obvious than when the web-server went down and the complaint-line phones lit up.

Visualising market information

While the type of system offered in Lucas County enables the property purchaser to get a sense of the individual property, broader trends are harder to detect with data at this resolution. We can get a sense of the detail of the mapping when looking at figure 3 where those properties with higher valuations are shown with darker shading.

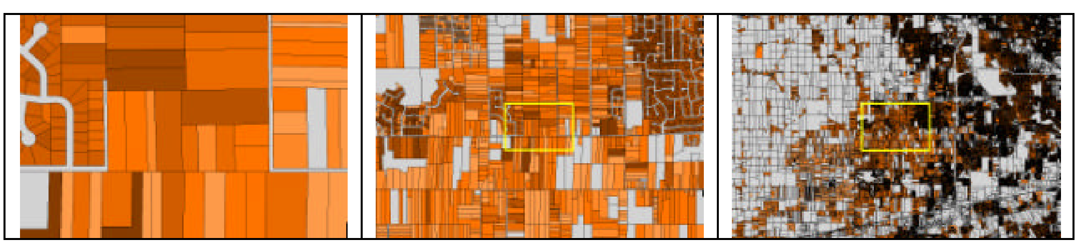

Figure 3: Lucas County property data viewed at different scales

Such a wealth of residential value data, recorded at the individual property level, paradoxically makes it more difficult to see the wood for the trees. This evident as we zoom out from the micro level in Figure 3. The issue of making data collected at one scale meaningful at another is a classic one within GI. One solution is to aggregate the information into zonal systems, but this is fraught with problems such as the well-documented modifiable areal unit problem and the issue of ecological fallacy (Openshaw, 1984).

A more elegant solution to summarising detailed information has been used by the property profession to visualise market data in both the residential and commercial property markets: the data surface. Its key benefit is that it makes it possible to visualise important trends and variations in the underlying geography which can be lost when mapping data using traditional area-based techniques.

An additional benefit is that like a topographic map, the data surface can be visualised as a contour map which enables the spatial variation in the data to be illustrated at a very fine scale. This type of representation has been used to map the property market (Figures 4 and 5).

Figure 4: Anstey’s land value map of the Barbican area, 1969 in £psf (after Howes, 1980)

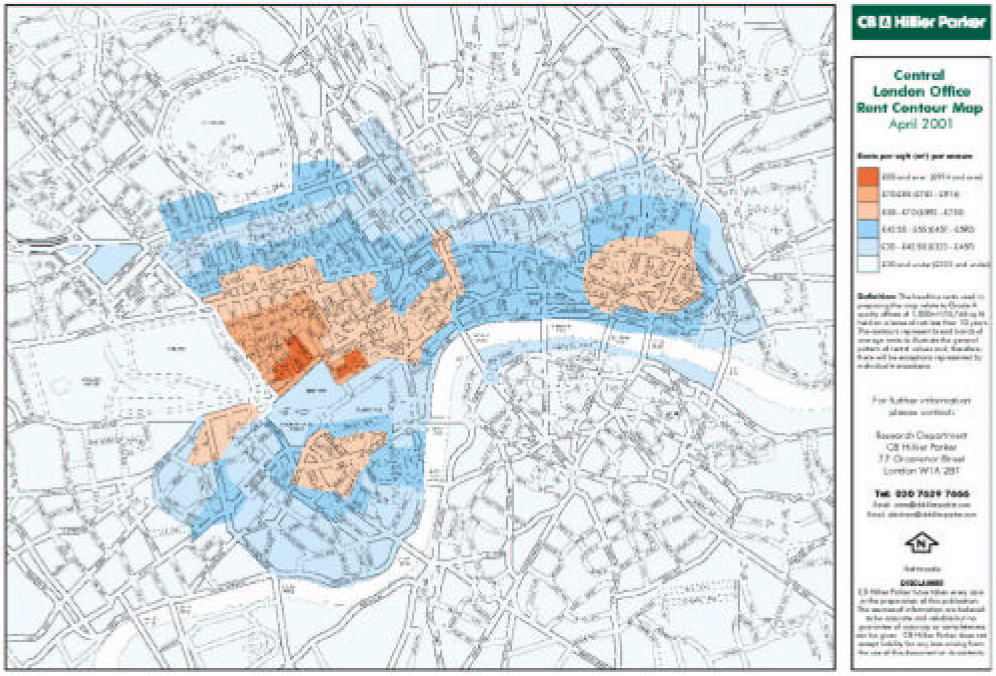

Figure 5: CB Hillier Parker's London Office Rent Contour Map, April 2001

There are two key benefits of visualising property data in this way. Firstly it enables the visualisation of market values at a variety of spatial scales. Secondly, it is highly usable, because it broadly accords with the way in which property professionals have always viewed the market. A topographic-style representation sits happily alongside agents' terms like 'prices dropping off', 'a steep increase on the other side of Acacia Avenue' or 'a peak in the market'.

But how close to the real market are these maps?

Mapping the market – the theory

The property market is by definition an abstraction. Individual prices are agreed by a wide range of people and professions, and the collective sense they have of 'the market' is nebulous. Broad trends are often clear, but detail is hazy – which makes it arguable whether the market can truly be mapped.

Most chartered surveyors would agree that the areas of high office values in central London as shown as red islands in CB Hillier Parker's rental contour maps are correct, but they might disagree on the contour values shown, or their precise location. Anstey's map which shows 1969 land values in the City peaking around the Bank of England is logical enough, but controversy might creep in over the shape of the contours around the Barbican, under construction at the time.

Differences of opinion naturally occur in a profession whose members rarely have access to the same information. Open information sources such as AREIS are very uncommon. And even if a common source of market information existed, there is no guarantee that different professionals would interpret the data the same way.

The maps shown above in Figures 4 and 5 were drawn by hand. Using transaction information, represented as points, it is possible to draw a contour map by inferring overall value from the sample of points. Ultimately, this method of drawing market value maps has to be unsatisfactory: too much depends on who drew the maps and when. Different people using precisely the same sample information will come up with different contours.

An individual's market knowledge will inevitably be limited. They may be a useful source for a small area, but we need to understand the property market's spatial variation over much larger areas.

Fortunately, GIS offers a potential solution to this problem. There is long tradition of creating consistent surface maps from point data using computer algorithms. Many of these techniques were pioneered to develop accurate topographic and geological maps from point samples (such as spot heights or boreholes) but the techniques have been applied to a diverse range of applications, from climate mapping through to property markets.

There are many different techniques which can be used to generate these surface maps, a thorough review of which can be found in O'Sullivan and Unwin's excellent Geographic Information Analysis. All are based on a key underlying principle: that objects which are close together in geographic space are more likely to be similar than those further apart.

Often referred to as Tobler's first law of geography, the degree to which objects vary by distance (be they topographic heights or property values) can be quantified, and is formally known as spatial autocorrelation. In order for data surface to be created, it is essential that spatial autocorrelation is present in the data being modelled.

There are two distinct families of methods which can be used to create data surfaces. The most basic are known as mathematically constrained techniques. The most popular of this techniques, Inverse Distance weighting or IDW, assumes (as its name implies) that the degree of spatial autocorrelation decreases consistently as an inverse of distance. Thus the height of the data surface at an unknown location can be interpolated by calculating the average values of nearby locations where values are known, with those which are closer to the unknown location having a greater influence on the interpolated value. The degree of influence is determined, and constrained, by a pre-set distance decay function.

The second family of techniques take a more statistical approach whereby the distance decay function is actually derived from the sample data themselves, rather than being determined in advance. The most well-known of these approaches is kriging, an approach which was pioneered in the South African mining industry. Since kriging determines spatial autocorrelation from the sample data themselves, it might appear a better way of creating a data surface than the more simplistic IDW. If only things were that easy.

In kriging, spatial autocorrelation is assessed by the semivariogram model. This is a mathematical function which best describes the way in which values vary according to distance. It is calculated by plotting squared difference in value of all possible pairs of locations in the point sample and trying to identify the best fit line which summarises that distribution. In real datasets there may be many hundreds of thousands of points on the graph, and so the distances are banded into lags to make the graph more intelligible.

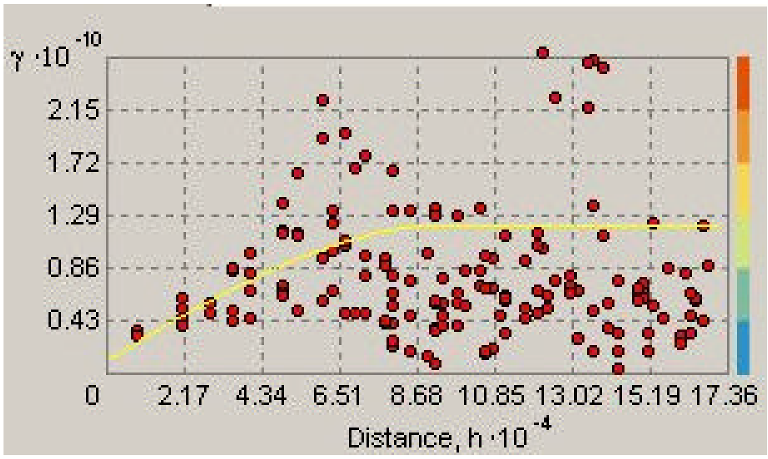

Figure 6 : Fitting the semivariogram model.

Figure 6 shows the semivariogram for residential transaction data in Lucas County in 1999, modelled using ESRI's Geostatistical Analyst, an extension to ArcGIS. We can see that as distance increases (plotted on the x-axis), the difference between the values of pairs of data points also increases (plotted on the y-axis). This is a graphical representation of decreasing spatial autocorrelation. The broad trend is summarised by the yellow line which is then used to interpolate unknown values from the sample data.

Choosing the shape and form of the yellow line is not straightforward. While there are many rules of thumb that can be applied, allowing the process to be automated to a greater or lesser degree, often the selection of the correct function is a down to experience. This is where geostatistics moves from objective science to the more murky area of subjectivity and artistry.

Mapping the market – the practice

Both these techniques have been applied to data from Lucas County and elsewhere. They are a crucial input to the Computer Aided Mass Assessment (CAMA) model used by valuers to derive the property valuations upon which tax is collected.

Surface models using both IDW and kriging are used to create a weighting matrix which accounts for the variation in value in space. It can be viewed as an objective means of quantifying the effect of location, location, location.

So how do they work in the UK?

Geofutures has done some work assessing the impact of the development of the Croydon Tramlink on neighbouring property values.

To do this, we had to create maps showing the spatial variation in the market before and after the Tramlink was constructed. From there, we needed to create a map showing the percentage change in market values across the study area over this time. If the Tramlink had affected values, then we would expect them to have increased more than the average for a nearby comparable area.

(Unfortunately, we are unable to share our results with the AGI conference at the time of writing since the project report has yet to be published, but the methodology used illustrates the practicalities of property market mapping.)

Our first requirement was detailed transaction data that, taken together for a particular point in time, provided a snapshot of the property market. Ideally, we needed these data for before and after the Tramlink's completion, and for broadly the same time of year to reduce possible seasonal effects.

Market value was determined by mapping transaction prices at unit postcode level using data from the HMLR for a particular point in time, providing a snapshot of the property market for the second quarters of 1996 and 2002. It could be argued that in a rapidly-moving market, these three-month snapshots were too large, but in the interests of maintaining a large enough sample, we assumed that the degree of temporal auto-correlation in the dataset was sufficient. Data from these snapshots were then interpolated into a data surface to show transaction value across the whole of the study area. Figure 7 shows the location of transactions for terraced houses in 2002, while Figure 8 shows the market value surfaces generated using the IDW algorithm from these data.

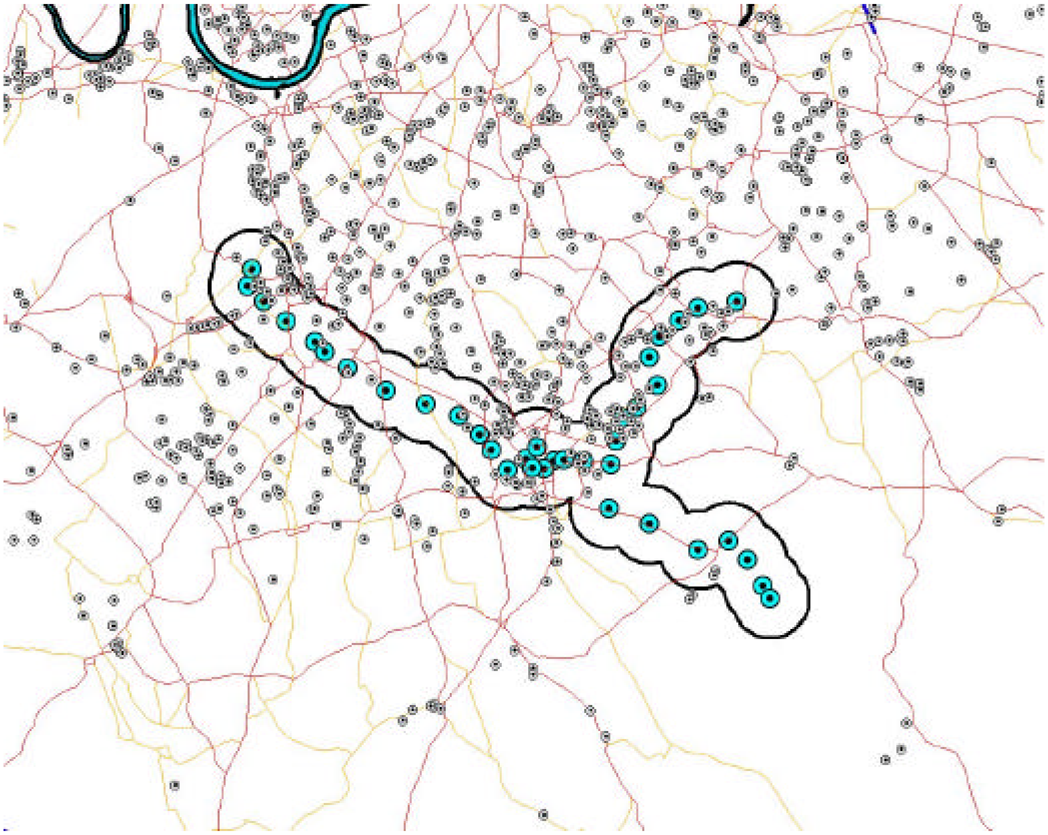

Figure 7: Terraced house transactions in South London, second quarter 2002

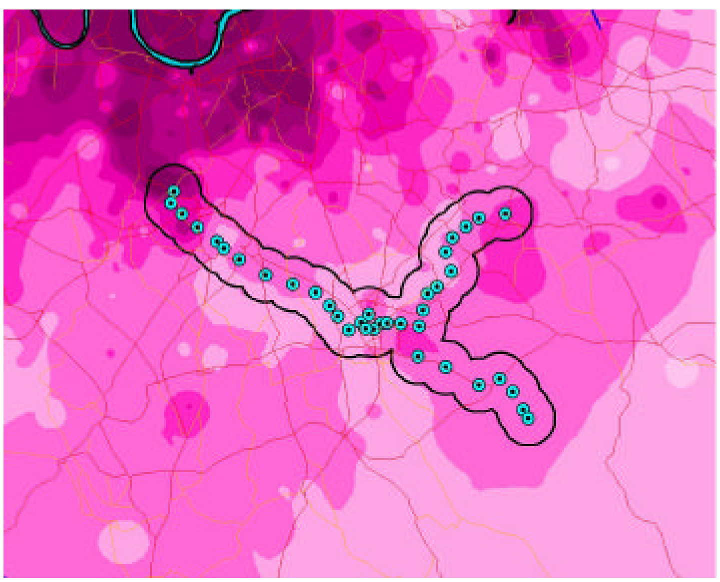

Figure 8: Interpolated value surface for terraced housing in South London, second quarter 2002

It is immediately apparent that the shape and form of the output surfaces are strongly dependent on the spatial distribution of the transaction data used to create them. The calculated values in areas where the spatial sample is relatively sparse (such as in the south west of the map) are therefore less reliable.

Kriging could perhaps resolve this, but as we went through the modelling process it became clear that it would be much more difficult to model property data in the UK than in the US. There was an insufficiently obvious spatial auto-correlation structure, even over relatively short distances, to identify a decent model. Since the presence of an identifiable spatial auto-correlation is a key assumption underpinning the kriging approach, it could not be safely applied.

Indeed, discussion at AGI meeting on GIS and property taxation suggested that it was not possible to automate the selection of the semivariogram model with data from England. If the process cannot be automated, we end up relying on the expertise of a statistician, and are forced to question whether this is more reliable than the hand-drawn contour maps of the chartered surveyor.

Why it works better in the US than the UK

There are a number of reasons why the surface modelling approach proves to be much more tricky in the UK than in the US and in other countries that use this technique as part of CAMA.

First is the lack of detailed data. The HMLR transaction data had very sparse level of attribution associated with it. There is no information on the number of bedrooms or floor area, for example – information which is necessary to normalise the data across a particular sub-market such as semi-detached houses or terraces.

The georeferencing we used is slightly more coarse than in Lucas County. We have no national cadastre, nor any reliable means of georeferencing all properties in the UK. In the CTL study we used the unit postcode centroid (the co-ordinates of the property closest to the spatial average of the 10-15 properties that fall, on average, within a unit postcode). This may not be fine enough, since we know empirically that property values can vary markedly from one side of a street to the other.

While data issues are undoubtedly significant, there may also be deeper structural issues making the use of surface interpolators, especially kriging, more difficult to apply in the British residential market.

Firstly, the housing stock is much more heterogeneous in UK towns and cities than in the US. They are older; it is common to find extremely old buildings located next to newer ones. Lines of post-war houses and flats dissect Victorian terraces following the flight path of Blitz bombers, for example. The pressure for land and the tremendous variety of land uses in our densely packed towns and cities means that Tobler's first law of geography cannot be so robustly applied to property markets in this country.

This spatial heterogeneity in the UK residential market is exacerbated by its lack of transparency. In the absence of any real, publicly available information, valuations inevitably become more volatile and dependent upon estate agents' opinion. This is especially true of the rapidly rising housing market, fed by historically low interest rates, seen in the UK over the past few years. If purchasers and building society valuers could see true evidence of all relevant transactions and property tax valuations, then the smoke and mirrors might be reduced.

We should be aware of how this works in practice, however. Looking at property values in a GIS like AREIS means viewing the results of a modelling process used to calculate the current valuations. A computer model plays out a host of assumptions about the way in which a market behaves. This can create circularity: the model defines the valuation which in turn drives the market, transactions are input into the model which creates a new valuation, and so the wheel turns. Such an effect can be identified by looking at the spatial auto-correlation structures in the data, where evidence of their greater stability and predictability through time would suggest that this process is at work. The self-fulfilling potential of statistical modelling is an important caveat, and should not be discounted lightly.

Conclusion

The idea of market circularity through the use of GIS in mapping property values is controversial. It implies that GIS could have a tangible impact on our property buying, shaping the market rather than merely mapping it.

Let's consider the implications. If we accept that our abstract market can be mapped at all, we also know that currently it is mediated for us by experts in the shape of estate agents. They, if anyone, shape the market now. We could look at the Lucas County example and argue that AREIS works alongside local realtors by smoothing the market, helping to prevent outlandish valuations. Or we could see market mapping as replacing one set of experts with another.

Is this a good or a bad thing? Both the estate agent and the statistician are equally well versed in their own brands of the black arts. Both use data, however they hold it, and as much experience and gut feeling is probably employed when developing a computer model as in valuing a house. And if anything seems more reliable than the shiny-suited agent's patter, equally we should beware the geostatistician hiding behind a veneer of objectivity and scientific rigour, not always justifiably.

In the UK, this question is still some time off. Poor information, the lack of a consistent georeferencing base of individual properties and a very rigid confidentiality environment all stop us using many of the techniques pioneered in Lucas County and elsewhere.

Which makes it timely to open the debate and start exploring the implications thoroughly. Maintaining the long term sustainability of the British economy and society means getting a handle on the housing market. Without more rigorous tools, all we have is the local estate agent and a few TV property shows to help us understand location, location, location.

References

Howes, C.K. (1980) Value maps: aspects of land and property values. Geoabstracts, Norwich. ISBN 0-86094-060-8

Openshaw, S (1984) The Modifiable Areal Unit Problem. CATMOG 38, Geobooks, Norwich

O'Sullivan, D & Unwin, D. (2003) Geographic Information Analysis. Wiley, New Jersey. ISBN 0-471-21176-1

Web addresses

Lucas County Auditor's on-line GIS - http://www.co.lucas.oh.us/AREIS/AREISMain.asp