SpAM

SpAM (Spatial Analysis and Methods) presents short articles on the use of spatial statistical techniques for housing or urban development research. Through this department of Cityscape, *the Office of Policy Development and Research introduces readers to the use of emerging spatial data analysis methods or techniques for measuring geographic relationships in research data. Researchers increasingly use these new techniques to enhance their understanding of urban patterns but often do not have access to short demonstration articles for applied guidance. If you have an idea for an article of no more than 3,000 words presenting an applied spatial data analysis method or technique, please send a one-paragraph abstract to ronald.e.wilson@hud.gov for review.*

Using Dual Kernel Density Estimation To Examine Changes in Voucher Density Over Time

Ron Wilson
U.S. Department of Housing and Urban Development

The views expressed in this article are those of the author and do not represent the official positions or policies of the Office of Policy Development and Research or the U.S. Department of Housing and Urban Development.

Abstract

The measurement of participants in the Housing Choice Voucher Program across time is an important analytical step toward understanding their settlement patterns, particularly whether they concentrate or deconcentrate. Many analyses of voucher-holder settlement patterns employ some areal unit in which counts are divided by unit area to calculate a density. This approach has methodological problems and produces less-than-accurate results because it does not directly measure the locations of voucher holders. In this article, I show how to apply a technique, known as Dual Kernel Density Estimation, to measure directly the concentration of voucher-holder locations to produce more accurate results about where voucher holders have concentrated and deconcentrated over time.

Introduction

Many housing and urban development problems are inextricably tied to place. Because of this link to place, certain questions often arise. Will foreclosures concentrate and spread to neighboring areas through falling house prices? Will tax increases in one county send residents to nearby counties to shop or relocate? Will crime displace to adjacent neighborhoods in the event of a concerted effort to break up a concentration of incidents? Answers to such questions require the measurement of spatial relationships between places that classical statistical techniques are not capable of measuring. In this premier article of SpAM, I demonstrate how to use a spatial smoothing technique to identify changing patterns of voucher-holder concentration between two points in time.

Housing researchers are often concerned about the concentration of voucher holders. A typical approach to measuring voucher-holder density change is by comparing areal densities (events per acre or per square mile) at two different times using already defined political or administrative units (for example, nations, states, counties, or census tracts). Chances are high that measuring change with these units will produce less-than-accurate results, because it does not directly measure the locations of voucher holders. In this article, I show how to apply a more accurate technique, Levine's Dual Kernel Density Estimation (DKDE), using the locations of housing choice voucher holders in the Charlotte-Mecklenburg, NC metropolitan region for purposes of illustration. For a more detailed exposition, see chapter 8 in Levine (2010).

The Housing Choice Voucher Program (HCVP) enables low-income families to relocate to neighborhoods of their choice. In 2010 alone, approximately 2.1 million families received assistance through the HCVP.[1] One common concern about the relocation freedom that HCVP offers is that participants will concentrate in certain neighborhoods. Research has shown that voucher holders often relocate to neighborhoods comparable with those in which they lived before receiving assistance (Freeman and Botien, 2002; Huartung and Henig, 1997; McClure, 2010; Pendall, 2000; Varady, Walker, and Wang, 2001; Wang, Varady, and Wang, 2008).

Moving Beyond Measuring Density Calculations With Areal Units

Many voucher-holder location analyses use census tracts to measure density. In a typical calculation of densities from areal units, a count of observations within the unit is divided by the unit area. This approach has two main problems. First, the aggregation of observations to the areal units forces an incorrect assumption that voucher-holder locations are evenly spread across the unit; the larger the census tract, the more unrealistic the assumption becomes. Second, the variation in census tract shapes will arbitrarily influence the unit within which an observation falls; this method may split up groups of voucher-holder locations.[2]

[1] Public and Indian Housing Tenant-Based Rental Assistance: 2012 Summary Statement and Initiatives. http://portal.hud.gov/hudportal/documents/huddoc?id=Tenant_BR_Assis_2012.pdf.

[2] These problems are symptoms of the Modifiable Areal Unit Problem, which has adverse consequences for data analysis because the unit of geography changes, but the observation data do not. See Openshaw (1994).

With DKDE, single kernel density surfaces are created by interpolating estimates from a geographically distributed set of observations. Estimates are calculated by overlaying a grid system across a geography in which the distance from each cell to every observation within a specified distance is measured and weighted. Measuring and weighting is achieved by using a mathematical function to create a symmetrical distance decay curve (kernel) that decreases from the cell origin. Exhibit 1 depicts the interpolation process.[3]

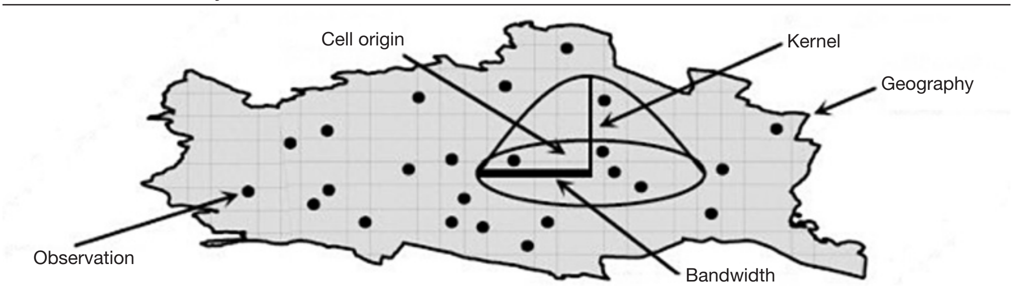
Each observation within the bandwidth is measured and weighted based on how close it is to the cell origin and where it corresponds on the distance decay curve. This process is completed for each cell in the grid to produce an overall density surface of the geographic distribution of locations.

Using DKDE to measure density change has several advantages. First, voucher-holder locations are weighted based on proximity to each other. Second, measurement is standardized using the same mathematical function across the entire geography. Third, an overall geographic density surface with gradients of continuous estimates is produced that summarizes clusters of points for easier interpretation of patterns. Finally, our use of DKDE protects privacy by representing voucher density as an areal estimate rather than as a collection of points.

In CrimeStat 3.3, five mathematical kernel functions weight distance decay differently.[4] A function should be selected based on a theoretical, empirical, or other substantive reason for its use. I selected the negative exponential function for this analysis, because the close-proximity weighting it emphasizes characterizes two substantive aspects of voucher-holder locations. The research cited previously suggested that (1) voucher holders tend to live in places where other voucher holders are in close proximity and (2) voucher holders often do not move far from where they previously lived.[5]

Exhibit 1

The Kernel Density Estimation Process



Source: Adapted from Fischer, Leitner, and Staufer-Steinnocher (2001)

[3] Exhibit 1 depicts the normal quartic function, but the same principles apply to the negative exponential function and the other functions in CrimeStat 3.3.

[4] Those functions are (1) normal, (2) quartic, (3) negative exponential, (4) triangular, and (5) uniform. For a description of each operation, see pages 8.3 through 8.9 in Levine (2010).

[5] The author's unpublished work (Wilson, 2012) on vouchers in the Baltimore area shows that the median moving distance from previous locations is about 3 miles.

The negative exponential function creates a distance decay curve that forms a very narrow peak at the cell origin, which rapidly and monotonically decreases up to a specified distance. Locations closer to the cell are weighted more highly than those farther away; locations outside the specified distance are excluded.[6] The negative exponential function is formally defined as

$$g_i(x_j) = \left\{\sum A \cdot e^{-K \cdot d_{ij}}\right\} \in h. \tag{1}$$

In this function, x_j is a set of voucher-holder locations affecting our estimate of the areal density in cell i; $g_i(x_j)$ is the density estimate; h is a search distance threshold (bandwidth) that encompasses the subset of voucher-holder locations within the bandwidth (that is, value = 0 if distance > h); A is a rescaling constant that ensures the density estimates sum to the number of locations;[7] K is an exponent constant set to 3, producing a steep decrease in the decay curve; and d is a distance between the center of cell i and a location j within h. The estimate $g_i(x_j)$ will be larger when voucher-holder locations j are near cell i and smaller when voucher-holder locations j are farther away from cell i.

One important factor in calculating the density estimate is the bandwidth size of the search distance for identifying locations to measure. Two types of bandwidths—adaptive or fixed—can be employed. An adaptive bandwidth, which is a changing search distance used to identify a specified number of locations within its radius, is used when a need arises to account for changes in the distribution of locations due to variation in the underlying geography; for example, population density. A fixed bandwidth, which uses a constant search distance from each cell origin and identifies a varying number of observations within its radius, is used to detect concentrations of observations within a specified distance.

The type of bandwidth chosen affects the calculation of density estimates and is more important than the type of distance decay (kernel) function selected, because bandwidth size will determine how much smoothing will occur. With a larger bandwidth, estimates will generally be weighted and distributed more evenly. Large bandwidths produce generalized density surfaces that identify subregional clusters. Smaller bandwidths consolidate estimates and produce a more discretely weighted and abruptly changing density surface that creates multiple noncontiguous local clusters. As with the kernel function, the bandwidth size should be determined using a theoretical, empirical, or other substantive reason for its specification. In this analysis, I used a fixed bandwidth to identify condensed groupings of voucher holders in close proximity to one another. The specified distance was 2.5 miles, which was based on a distance analysis.[8]

The DKDE technique creates two separate density grids for each distribution to represent the individual geographic distribution of observations before the grids are combined. The first distribution is voucher-holder locations in 2010 and the second distribution is voucher-holder locations in

[6] This exclusion of observations is true for all the mathematical functions in CrimeStat 3.3, except for the normal function. The normal function includes all observations across the geography.

[7] A is initially set to 1 and is iteratively adjusted until the estimates sum to the number of locations. Exhibit 2 shows the totals in this illustration.

[8] This distance was derived using the Ripley's K technique, in which the level of significant clustering dissipates. (Analysis available from the author.)

2000. Estimates between corresponding cells of the two grids are combined with any one of six different mathematical operations.[9] I selected the relative difference operation for this analysis to identify density changes between 2000 and 2010. This operation subtracts the density estimates in 2000 (secondary file) from the estimates in 2010 (primary file), producing divergent values greater and less than 0 to depict areas of concentration and deconcentration of vouchers.

Measuring Concentration Change in the Charlotte-Mecklenburg, NC Metropolitan Region

The number of voucher holders in the Charlotte-Mecklenburg, NC metropolitan region increased significantly over the past decade, with most jurisdictions showing double-digit percentage increases (see exhibit 2).[10]

Exhibit 3 shows the density change output of the DKDE to reveal patterns of group concentration and deconcentration. I used a standard deviation classification to thematically map the patterns to show divergence from cells with little or no change in densities. Values greater than 1 standard deviation from the mean (> 12.26) are depicted in dark gray, indicating increased density (concentration) and that fewer voucher holders appear per square mile in the second period than in the first. Values less than 1 standard deviation from the mean (< -10.99) are depicted in light gray, indicating decreased density (deconcentration) and that more voucher holders appear per square mile in the second period than in the first.

Exhibit 3 reveals a wide dispersal of local and subregional voucher-holder residential patterns over the past decade. Many areas of marked concentration and deconcentration occurred around the larger towns of Concord, Gastonia, Kannapolis, Monroe, Rock Hill, and Salisbury.

Exhibit 2

HCVP Participant Changes by County in the Charlotte-Mecklenburg, NC Metropolitan Region, 2000 to 2010

Jurisdiction	Particpants in 2000		Participants in 2010		Change	
	N	Percent	N	Percent	N	Percent
Cabarrus County, NC	416	7.7	497	5.6	81	19.5
Charlotte-Mecklenburg County, NC	2,380	43.8	4,983	56.1	2,603	109.4
Gaston County, NC	924	17.0	1,215	13.7	291	31.5
Lincoln County, NC	252	4.6	290	3.3	38	15.1
Rowan County, NC	579	10.7	755	8.5	176	30.4
Union County, NC	233	4.3	293	3.3	60	25.8
York County, SC	648	11.9	856	9.6	208	32.1
Total	**5,432**		**8,889**		**3,457**	**63.6**

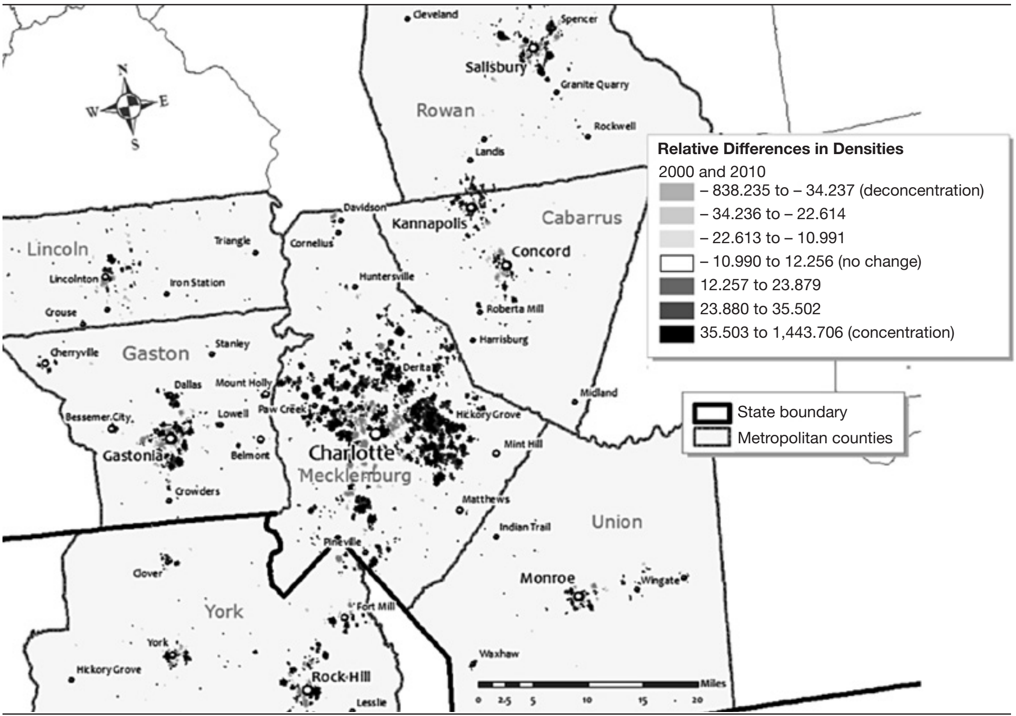
HCVP = Housing Choice Voucher Program.

[9] Those operations are (1) ratio, (2) log ratio, (3) absolute difference, (4) relative difference, (5) absolute sum, and (6) relative sum. For a description of each operation, see pages 8.27 through 8.29 in Levine (2010).

[10] The voucher-holder point location data come from the U.S. Department Housing and Urban Development's 2010 Public and Indian Housing Information Center for the years 2000 and 2010.

Exhibit 3

Differential Density Change in HCVP Participant Densities in the Charlotte-Mecklenburg, NC Metropolitan Region, 2000 to 2010 (standard deviation classification)



HCVP = Housing Choice Voucher Program.

Exhibit 3 also shows the concentration patterns are primarily situated around the town centers of these towns. The patterns around Charlotte show an arc of concentration around the downtown area, from the northwest to the southeast. Those concentrations happen to correspond with areas that had the largest increases in new housing over the past 10 years. Several clusters of deconcentration also emerge in the downtown area of Charlotte, which contains the city's oldest neighborhoods. Similar patterns of deconcentration appear in the central areas of Concord, Gastonia, Salisbury, and Rock Hill, but not in Kannapolis or Monroe. The technique succinctly depicts density changes at the neighborhood level, in both shape and size that would have been lost by using census tracts or other predefined administrative units.

The research cited previously often found that voucher holders tend to concentrate in impoverished neighborhoods. To explore this finding, census tracts with greater than 20 percent poverty (shown in alternating black and white lines) were added to the map[11] and overlaid with the densities to show how geographically specific clusters are in comparison with the areal units. The map in

[11] Poverty data come from the 2005–2009 American Community Survey 5-year estimates.

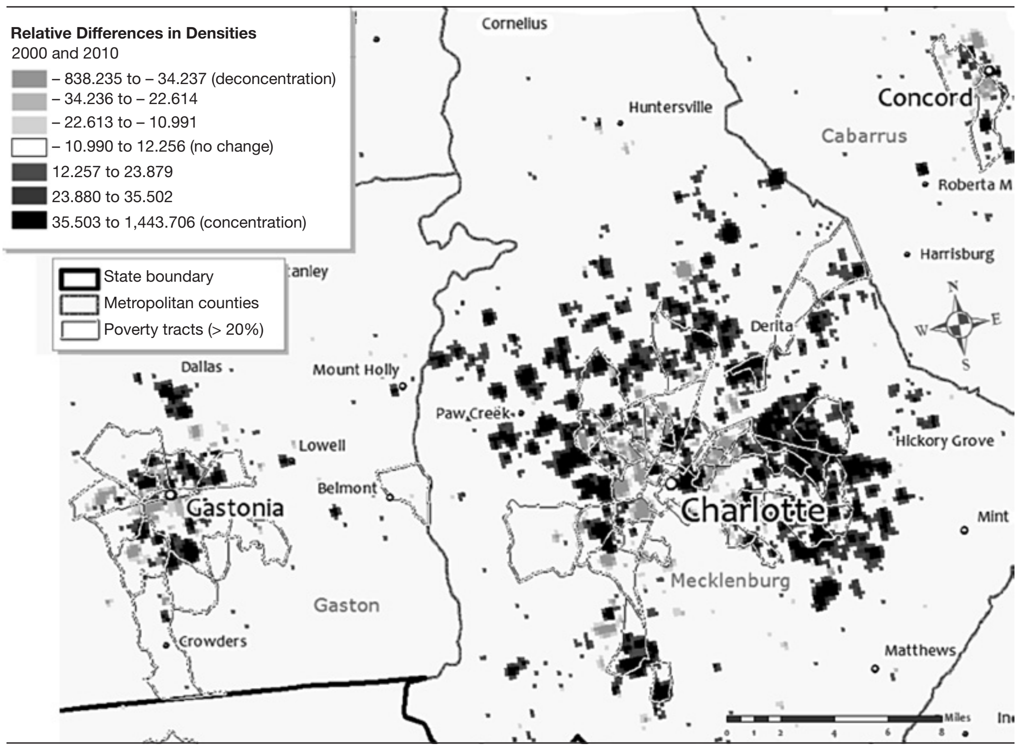
exhibit 4 shows several clusters of voucher-holder concentration and deconcentration that cut across the boundaries of high-poverty census tracts, illustrating the benefit of the DKDE output with respect to identifying specific areas of concentration within census tracts.

Had the densities been calculated using census tracts, the concentration and deconcentration patterns depicted would have been less precise. Using census tract boundaries can also segment groups of voucher-holder locations and reduce the chances of the group being identified as a coherent cluster. For example, several clusters of concentration and deconcentration are split across multiple census tract boundaries in Concord, Gastonia, and central Charlotte. In some instances in which only a small portion of a cluster is within a census tract, any micropatterns of voucher-holder locations would have been dilluted.

These results extend the change rates reported in exhibit 2 by showing exactly where voucher holders have concentrated and deconcentrated beyond the change rates. Although significant growth in voucher holders occurred outside central Charlotte, growth also took place around many of the smaller town centers in the metropolitan region. Considerable levels of voucher-holder concentration remain, however, in areas with higher poverty rates.

Exhibit 4

Differential Density Change in HCVP Participant Densities in Central Charlotte, North Carolina, 2000 to 2010 (standard deviation classification)



HCVP = Housing Choice Voucher Program.

Extensions of the DKDE and Resulting Output

The DKDE output is not limited to visualizing clusters. At the very least, the grid cells could be coded and aggregated to areal units to create ratios between cells that have high estimates and cells that do not to produce a more accurate level of density. This approach provides a marked improvement in the precision of density as opposed to generalized densities with areal units.

More importantly, the density grids can be overlaid with other local-level geographic data to explore microrelationships. For example, clusters from DKDE could be matched with parcel data to examine relevant local factors, such as property type, housing amenities, assessment values, land use, and code violations. Voucher-holder locations could also be analyzed in conjunction with the distribution of businesses to determine how extensively participants concentrate in areas with accessible jobs or good-quality services. Local-level data may even be used with other mathematical functions from DKDE analysis of voucher-holder locations to reveal additional spatial relationships.

Estimates from the DKDE can also be converted into change rates. Visualizing change rates will enhance the understanding of the cluster patterns. To create a change rate for the cells in the surface, the density values for the earlier period can be created with a Single Kernel Density Estimate (SKDE) using the same settings as the DKDE. After the single-density surface is created, it can be joined with the DKDE surface in a Geographic Information System, and rate change can be calculated for each cell by taking the difference estimates from the DKDE analysis, dividing the estimates by the SKDE estimates, and multiplying the quotient by 100 to produce a rate.

Many geographic analyses are still limited in precision because of the continued use of areal units. Researchers with access to point-level data can use a technique like DKDE directly with point locations to more precisely depict the levels and changes of geographic microactivity of concentration and deconcentration.

Acknowledgments

The author thanks Ned Levine from Ned Levine and Associates, Jay Lee from Kent State University, and Fahui Wang from Louisiana State University for providing valuable comments toward improving this article.

Author

Ron Wilson is a social science analyst in the Office of Policy Development and Research at the U.S. Department of Housing and Urban Development and an adjunct faculty member of the Geographic Information Systems program at the University of Maryland, Baltimore County.

References

Fischer, Manfred M., Michael Leitner, and Petra Staufer-Steinnocher. 2001. "Spatial Point Pattern Analysis: Some Useful Tools for Analysing Locational Data." In *Geographischer Jahresbericht aus Österreich*, Vol. 58, edited by H. Wohlschlägl. Vienna, Austria: Vienna University, Institute for Geography and Regional Research: 49–66.

Freeman, Lance, and Hilary Botein. 2002. "Subsidized Housing and Neighborhood Impacts: A Theoretical Discussion and Review of the Evidence," *Journal of Planning Literature* 16 (3): 359–378.

Huartung, John M., and Jeffrey R. Henig. 1997. "Housing Vouchers and Certificates as a Vehicle for Deconcentrating the Poor: Evidence From the Washington DC Metropolitan Area," *Urban Affairs Review* 32 (3): 403–419.

Levine, Ned. 2010. "Chapter 8." In *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations*, v 3.3. Houston: Ned Levine & Associates; Washington, DC: National Institute of Justice.

McClure, Kirk. 2010. "The Prospects for Guiding Housing Choice Voucher Households to High Opportunity Neighborhoods," *Cityscape* 12 (3): 101–122.

Openshaw, Stan. 1994. *The Modifiable Areal Unit Problem*. London, United Kingdom: Geo Books.

Pendall, Rolf. 2000. "Why Voucher and Certificate Users Live in Distressed Neighborhoods," *Housing Policy Debate* 11 (4): 881–910.

Varady, David P., Carole C. Walker, and Xinhao Wang. 2001. "Voucher Recipient Achievement of Improved Housing Conditions in the U.S.: Do Moving Distance and Relocation Services Matter?" *Urban Studies* 38 (8): 1273–1304.

Wang, Xinhao, David Varady, and Yimei Wang. 2008. "Measuring the Deconcentration of Housing Choice Voucher Program Recipients in Eight U.S. Metropolitan Areas Using Hot Spot Analysis," *Cityscape* 10 (1): 65–90.

Wilson, Ron. 2012. Long-Term Mobility Patterns of Housing Choice Voucher Recipients in the Baltimore Metropolitan Region, 2000 to 2010. Unpublished manuscript.

Additional Reading

Miller, Henry J. 2004. "Tobler's First Law and Spatial Analysis," *Annals of the Association of American Geographers* 94 (2): 284–289.

Wang, Fahui. 2006. *Quantitative Methods and Applications in GIS*. Boca Raton, FL: Taylor & Francis.