Development of Output Geographies for Comparative and Temporal Census Research

Mike Poulsen
Macquarie University*

The potentiality exists for a multitude of different geographic objects to be established from the same census dataset. Faced with a variety of aggregation procedures it is critical that we develop research into geographic objects based on a standard methodology which satisfies a set of specific aggregation criteria. If not anarchy will prevail in the form of copious new output geographies. This would be a disaster in that geographic objects are of fundamental importance to the study of both physical and human geographic structures. In census geography the study of geographic objects primarily aims at acquiring vital information on differences between cities, the evolution of those cities and the study of urban management. After initially discussing the issue of which aggregation procedure to use this paper adopts a standard procedure, that of the fragmentation and objectification analysis, which is applied to a race/ethnicity dataset for Los Angeles. As displayed in the Los Angeles application the number of types of major fragments and geographic objects that exist in this city are very limited, but the number of geographic objects are many. Los Angeles beyond the two citadels of the Whites and the barrio of the Hispanics is a very fragmented city.

Keywords: Geographic objects, output geographies, census aggregation, Los Angeles, fragmentation, objectification.

Just as GIS and remote sensing have focused the geographers' attention on the nature and evolution of geographic objects in the physical domain (Couclelis, 1996; Burrough and Frank, 1996), recent research in GIS has generated an increasing interest in the nature of geographic objects in census datasets (Poulsen and Johnston, 2000). In the past the study of urban population structures based on census data has focused on either establishing the major dimensions in the census data (combining different variables in the factorial ecologies of the late 1960s) or establishing the typology structures of urban areas (combining different observations in the grouping/clustering analyses of the 1970s and 1980s). By contrast with those earlier periods of urban research, research into the nature of geographic objects in census datasets is about the identification of the base units of urban structure: fragment types and geographic objects. As Couclelis accurately observed on reflection about the earlier periods of research in geography "who was thinking of mountains, cities and shorelines as

* Macquarie University, Sydney, Australia, 2109. E.mail: Mike.poulsen@mq.edu.au

'objects' before GIS"? (Couclelis, 1996:45). These census-based geographic objects are however not digital representations of physical entities but of human constructs as measured by either economic, or social, or race/ethnic, or population, or housing variables. They exist as structures in the census datasets and not necessarily as real world structures we use to assist us to navigate our way through our spatial environment. Few have been ascribed names or can be identified by individuals in the city, let alone have their crisp or fuzzy boundaries identified. Because of this they have to be extracted from the census datasets by an aggregation methodology plus the use of the censuses digital boundary files.

As we move to analyze the current round of 2000/2001 censuses much has changed in the methodologies used in census geography since the last round of censuses five or ten years ago. Primarily researchers now see the need to use the census data at the enumeration area level as the building blocks for the new output geographies that are independent of administrative structures (Martin, 1998). This shift is primarily due to (i) a greater awareness of the role of scale in terms of the generalization of datasets and (ii) technological advances both in GIS and computing hardware. As Openshaw and Rao (1995) conclude no longer is the analysis and mapping of census datasets based on administration areas acceptable for many applications. Today's researchers maintain that if the study requires that the population data be more accurately located spatially than that afforded by a uniform distribution within each enumeration area then discontinuous areas (built and not built environments) be obtained using population surface modeling (Martin and Bracken, 1991; Harris and Longley, 2000) or dasymetric modeling (Martin, Tate, and Langford, 2000). Researchers also contend that census datasets should be integrated with imagery to provide both better image classifications (Mesev, 1998) and the enhancement of the census data (Hall, Malcolm and Piwowar, 2001).

Importantly at the same time as the development of new output geographies is taking place 'place-to-theory' based census geographies are being replaced by comparative and temporal 'place-to-place' research. In part this is a response to globalization, here represented as the removal of national borders through the availability of digital census information; boundary files and imagery via the internet across multiple nations. Globalization also forces us to distinguish between global and local processes and outcomes. Importantly the distinction needs to be made between global theories and local variants of those theories. We need to know what is unique within a nation or a region and what is common across the world. The new 2000/2001 round of censuses will importantly provide us with our first temporal datasets where digital boundaries cover three censuses. No longer constrained by national boundaries, or administrative aggregations, these developments, along with advances in computer hardware and software, have and will continue to radically change the nature of census geography.

Census-based geographic objects are based on the aggregation of the smallest census areas (or their discontinuous parts from vector based dasymetric modeling) into larger spatial units on the basis of their characteristics. A geographic object by definition is an area which is different from its immediately surrounding neighbors on the basis of at least one value. Therefore, unlike the census areas, which may be part of a geographic object, each geographic object is a unique observation. There are seen to be multiple sets of geographic objects that we can derive from census datasets, each characterized by a specific census variable. Unlike previous research into factorial ecologies or urban typologies the establishment of census-based geographic objects does not include a mix of different variables in the same analysis. Instead we first establish for each census variable a set of geographic objects and only then analyze the nature of those geographic objects intersection. New output geographies based on census datasets however create a dislocation between the structures used by governments for funding (the administrative areas) and those used by academics for research investigation. If spatial policies and grant funding are to be switched by governments from administrative areas to these new output geographies considerable research needs to be undertaken into what are the most appropriate new output geographies.

Processes create census-based geographic objects. The measurements we derive from the analysis of census-based geographic objects not only provide us with important information about those processes but also the outcomes. To obtain this information from new output geographies standardized methodologies need to be developed (Thurstain-Goodwin and Unwin, 2000). As a result the technical details of the method of aggregation we decide to use is of fundamental importance, in that any aggregation method along with digital boundary files can produce a set of census-based geographic objects. To achieve a standard methodology it is necessary that we establish a set of criteria that have to be met by any aggregation procedure we might decide to use. This set, the author contends, contains the following five aggregation criteria:

(i) Be a standard systematic and objective approach which can be easily replicated by others. Otherwise, we will have anarchy in the range of new output geographies produced;

(ii) Be an approach that is extent/frame independent (i.e., the results do not depend on the spatial extent of the study area). Otherwise, the results will not be spatially robust within the study area;

(iii) Be an approach that establishes universal measures of the nature of the geographic objects and their attributes and thereby supports comparative and temporal study. Otherwise, aggregations based on the unique variance that occurs in a study area create structures that are neither directly comparable both geographically or temporally;

(iv) Be an approach that establishes the nature of all geographic objects that exist in the dataset. Otherwise, the geographic objects acquired will be set by *a priori* specifications;

(v) Be an approach that establishes a standard nomenclature independently of the researcher. In other words, the naming of the geographic objects is automatic and not subject to the judgments of the researcher. This enables the measurements of geographic objects to be comparable both geographically and temporally.

We now turn to examine three procedures out of the multitude of aggregation procedures we might use:(1) cluster analysis procedure; (2) automatic zone procedure (AZP); and (3) the fragmentation and objectification procedure. The first two were selected here for review because they are of considerable importance. The third procedure is seen by the author to be the best aggregation procedure relative to the five critical criteria currently available. Each procedure will be examined in terms of the five aggregation criteria stated above, and then we will provide an example of the use of the fragmentation and objectification analysis procedure. The aims of this paper are therefore twofold. First, to display how advances in research in GIS have provided procedures that are rapidly changing the methodologies we use in census geography. Second, to advance the study of census-based geographic objects and in particular the fragmentation and objectification analysis procedures presented here.

ALTERNATIVE AGGREGATION PROCEDURES

Cluster Analysis Procedures

One of two standard approaches to the aggregation of pixels in raster imagery is to use cluster analysis in the form of the standard ISODATA program. Know as an 'unsupervised classification' (in that the person undertaking the classification does not select the geographic entities to be classified) the resulting groups of pixels are those that are most similar in terms of reflective values. The researcher must specify the number of groups that the observations are suppose to form. These groupings are however dependent on (i) the number of groups sought; (ii) the extent (frame) of the study area; and (iii) the unique mix of observations in terms of variance within the dataset. We can obtain the boundaries of these geographic objects by vectorizing that raster-based classification and dissolving the boundaries between contiguous polygons that are members of the same group. The naming of each type of geographic object by the researcher provides us with their thematic attribution.

The same type of analysis can also be applied to census datasets when we have digital boundary files. By subjecting the attribute data (variables) to a cluster analysis, then dissolving the boundaries between contiguous polygons where the observations (census areas) are members of the same group, a set of census-based geographic objects are obtained. However, this direct transfer of the 'unsupervised classification' approach from remote sensing to census geography is problematic (just as it is in remote sensing) in that it fails to meet four of the five specified criteria. First, the approach fails to determine the nature of *all* geographic objects in the data because initially the researcher must specify the number of groups they want classified. Second, the group structures are dependent on the extent/frame of the study area. Expand or subset the study area and many observations will be grouped with different observations. Third, the group structures are dependent on the unique nature of the variance in the datasets, hence classifications undertaken in a different study area or the same study area at a different time are not directly comparable. Fourth, the naming of the groups is up to the researcher and is not standardized. Only the first of the five aggregation criteria is met by this approach in that the procedure is both systematic and objective.

The alternative classification approach to an unsupervised classification in remote sensing is a 'supervised classification'. Here the researchers classifying the image initially select a set of 'areas of interest' that represent the geographic entities under study.[1] These areas of interest form the signatures and the pixels throughout the image and are classified according to the signature (each of which represents a geographic entity) they are most like (by the maximum likelihood classifier). But just as with an unsupervised classification this approach also fails to meet four of our five aggregation criteria. First, with census datasets the basic criteria for a supervised classification, namely an *a priori* list of census-based geographic entities, cannot be met. In census geography we do not know the nature of the set of geographic entities that exist in each dataset, because these are not physical entities to which we ascribe everyday identities. Instead they are either economic, or social, or population, or housing structures that exist in the census dataset (Poulsen and Johnston, 2000). As a result we cannot select a series of areas of interest that form the signatures of each census-based geographic object. Second, the group structures are dependent on the extent/frame of the study area. Third, these group structures are also dependent on the unique nature of the variance in the dataset. Fourth, the naming of the different types of geographic objects established is by the researcher, hence no standard nomenclature is used. Again, as with the use of this methodology in remote sensing, this aggregation procedure would only satisfy the first of our five aggregation criteria.

Taken together the remote sensing techniques of supervised and unsupervised classifications clearly fail to meet all five aggregation criteria established as essential for developing a standardized approach to establishing census-based geographic objects. However, these failures are positive outcomes in that they focus our attention on developing other approaches. Importantly the reasons why these remote sensing techniques fail us sets up the criteria required for a successful aggregation procedure.

Openshaw's Automatic Zoning Procedure (AZP)

Openshaw and Rao, reporting on the development of the automatic zoning procedure (AZP) in the late 1970s, viewed the design of their procedure as an unconstrained zone-design method. The aim was 'to start with data at one scale and then re-aggregate them to create a new set of regions designed to be more suitable for a specific purpose than the original building blocks' (Openshaw and Rao, 1995:427). Its task is to establish an *optimal set* of new regions, where any objective function can be used; for example the sum of the squared deviations from the average population in the regions. The original automatic zoning procedure (AZP) of Openshaw, which was an outcome of the mathematical modeling period in geography of the 1970s, is a heuristic procedure that evolved by trial and error from an iterative reallocation method commonly used in numerical taxonomy. Seven steps are involved and are listed by Openshaw and Rao as:

Step 1. Start by generating a random zoning system of N small zones into M regions;

Step 2. Make a list of the M regions;

Step 3. Select and remove any region K at random from this list;

Step 4. Identify a set of zones bordering on members of region K that could be moved into region K without destroying the internal contiguity of the donor region(s);

Step 5. Randomly select zones from this list until either there is a local improvement in the current value of the objective function or a move is equivalently as good as he current best. Then make the move, update the list of candidate zones, and return to step 4 or else repeat step 5 until the list is exhausted;

Step 6. When the list for region K is exhausted return to step 3, select another region, and repeat steps 4–6;

Step 7. Repeat steps 2–6 until no further improving moves are made (Openshaw and Rao, 1995, 429).

The principal problem, however, with AZP identified by Openshaw is that the procedure can become trapped by a local suboptimum solution. Developments reported by Openshaw and Rao (1995) to deal with this problem include: (a) using a simulated annealing variant of AZP; or (b) a tabu search heuristic.

But while this procedure successfully achieves the aggregation of the building blocks (enumeration areas) and produces a new output geography that is independent of administrative boundaries, it again fails to meet all but one of our five aggregation criteria. First, it aims at the optimization of the zones using some specified function as opposed to establishing the nature of all geographic objects in the study area. This reflects the origins of this approach developed in the era of mathematical modeling in geography. Second, the approach is extent/frame-dependent; hence the results change with the expansion or sub-setting of the study area. Third, the approach is dependent on the variance in the dataset and as such does not meet the specified universality of measurements requirement for comparative and temporal studies. Fourth, AZP does not put forward a standard nomenclature for attribution. But as with the cluster analysis procedures discussed above, AZP does meet the first aggregation criteria in that it provides a systematic and objective methodology.

Poulsen and Johnston's Fragmentation and Objectification Procedure

Poulsen and Johnston's (2000) procedure for aggregation is a fragmentation and objectification analysis procedure initially applied to continuous census data (contiguous census areas at the enumeration area level). It can also be applied to discontinuous census data (the result of dasymetric modeling or population surface modeling) where the non-residential areas are removed from within the enumeration areas. In contrast to earlier census-based factorial ecology research of the 1960s and typology research of the 1970s and 1980s, fragmentation and objectification analysis is undertaken on individual variables. The input data is population counts of individual sets of census variables, such as ethnicity, or employment, or income. Only when sets of geographic objects based on individual census variables have been established do we examine the spatial intersection between them. The procedure is technical and deliberately prescriptive in that it is essential that each researcher follows the same set of standardized steps. When initially published the procedure involved four steps:

Step 1. Based on population count data, a series of universal measures are obtained for each observation (census area) covering majority and dominance status and concentration levels (thresholds) relative to the total population in that census area. Each observation has its census population counts classified in a purpose written computer program by those measures and specified as a set of discrete values (0,1), where the value of 1 indicates that it conforms to that measure. Two discrete datasets are established for each observation so as to distinguish between the root structure of concentrations and the nature of the structures of higher levels of concentration. The first is termed the 'base set' and it has only three discrete values for each variable (majority, dominance, concentration level of 20 percent and above). Second, the 'full set' has nine discrete values for each variable (majority, dominance, and concentration levels of 20, 30, 40, 50, 60, 70, and 80 percent). This is the concentration structure of the geographic objects;

Step 2. For each individual attribute (population variable) the geographic objects are attained using the digital boundary dataset by dissolving boundaries between contiguous census areas that are members of the same type of geographic object. This procedure was undertaken in ArcInfo and produces a set of single attributed geographic objects (SAOs);

Step 3. A union of these single attributed geographic objects was undertaken in ArcInfo to produce a set of multiattributed geographic objects (MAOs); and

Step 4. A union of the multiattributes geographic objects and the original census dataset was undertaken in ArcInfo, followed by dissolving the boundaries between census areas that are members of the same geographic object type. The attribute for geographic object type was obtained in ArcInfo using the frequency command. The census dataset was aggregated (summed) during this process.

Subsequent research has established an alternative and simpler sequence of steps, which achieves the same results and replaces the use of ArcInfo. This approach, first reported here in this paper, is simpler in that single attributed geographic objects are not obtained and better in that it distinguishes between the fragmentation and objectification analyses. Again four steps are involved:

Step 1. As for step 1 above;

Step 2. These discrete dataset profiles are sorted in Microsoft Access (across all discrete fields), then a purpose written computer program is run to establish the number of different fragment *types* that exist (the fragmentation analysis). A fragment type is defined as a set of census areas *irrespective of their location* that have the same characteristics in terms of majority and dominance status and concentration levels (they have the same discrete profile). The purpose written computer program checks to see if the profile changes in the sorted file and ascribes a fragment type number to each observation. This is the classification of fragment types. So that the fragment types are universally attributed in the same manner a standard method of naming is also used. This method specifies the capitalized letter code of the population group followed by 'm' for majority status, 'd' for dominance status, or its percentage concentration level (i.e., 20, 30…). For example WmWpW70H20 indicates that this fragment type is one in which Whites are in the majority, Whites are the dominant group, Whites have a level of concentration of 70 percent or more and they share that space with Hispanics that have a concentration of 20 percent level or more;

Step 3. The classification number of each fragment type for each observation (census area) and its attribution (its standardized name) are joined to the original census dataset;

Step 4. The multiattributed geographic objects are obtained using the digital boundary file by dissolving boundaries in ArcView between contiguous census areas that are members of the same type of geographic fragment. A geographic object is here defined as an observation that is different from all its surrounding neighbors on the basis of at least one discrete variable. Its attribution is the same as its fragment type. During this dissolving process the census dataset for each geographic object is aggregated (summed).

With regard to our five aggregation criteria necessary for acquiring a standard methodology this procedure meets all five. It is systematic and objective. It establishes the nature of *all* geographic objects in the census dataset. It is extent/frame independent because it is based on a set of universal measures. It uses a set of universal measures (majority, dominance and concentration threshold levels) to facilitate comparative and temporal census studies. And finally a standard nomenclature is used; hence attribution is not an issue. If there is a generic weakness it lies in establishing separate geographic objects for each census variable, however this depending on your perspective is also a strength.

In addition there are four important features of this aggregation procedure that assist us with this research. First, there is a mathematical minimum and maximum number of fragment *types* (and therefore geographic object types) that can exist in any city. The relationship between the observed number of fragment types and the maximum number possible provides us with a valuable comparative and temporal measure. The minimum number of fragment types is 1, whereas the maximum is given by the equation:

$$\text{MaxTF} = \text{Ngroups} \times (\text{Ngroups} \times \text{Nlconc} + 1) \times 2$$

Where:

MaxTF: maximum number of fragment types possible;
Ngroups: number of different populations in the analysis;
Nlconc: number of levels of concentration in the dataset (3 for the base set, 9 for the full set).[2]

Second, by contrast with the number of fragment types, there can be as many census-based geographic objects as there are census enumeration areas, however the number of census-based geographic object *types* and their character (attribution) is the same as for the number of fragment *types*.

Third, in applying this fragmentation and objectification analysis methodology it is advantageous to make the distinction between major and minor fragment types. Major fragment types are defined by either their share of a city's area or total population they contain. We define major fragment types as those that account for one percent or more of the city's area or total population.

Fourth, the distinction between the 'base set' fragment types and 'full set' fragment types allows us to view fragmentation as a four level branching tree diagram, where the 'base set' form the third level of branches of the tree and the 'full set' the fourth level of higher branches. A tree exists for each different census variable set (i.e., ethnicity or unemployment or income...), with the second level of branches being the variables within each dataset (i.e., the different ethnic groups). The 'base set' of discrete data provides us with an output geography that is based on who shares which areas with whom. The 'full set' provides us with an output geography on levels of concentration and the nature of those concentration levels within those shared structures. As such the two output geographies depict different aspects of the nature of census-based geographic objects.

LOS ANGELES

We now turn to an example of the application of the fragmentation and objectification analysis procedure. Los Angeles CSMA (Consolidated Statistical Metropolitan Area) at the April 2000 census had a population of 16.374 million within 3,361 census tracts; an average population of 4,872 per census tract (Table 1). In this study the aim is to establish a set of robust measures on the nature of the racial/ ethnic fragment types and geographic objects within Los Angeles. The dataset used is population counts at the tract level of aggregation for a single set of variables in accord with the fragmentation and objectification analysis design discussed above. The four major populations examined are: Whites, Hispanics, African-Americans and Asians, where the Hispanics as an ethnic category are excluded from the other three race groups. Given that the 2000 census allowed individuals for the first time to identify up to six combinations of racial mix, only those who identified with one group (97 percent) are employed in these analyses. A number of general and detailed measurements are obtained and two output geographies for the study of ethnicity are produced. These output geographies are the census-based geographic objects that display (i) who shares the areas with whom (the 'base set') and (ii) the detailed structures of concentration (the 'full set'). In presenting these results the measurements acquired are discussed only in a general manner, whereas the true strength of the methodology lies in the study of individual populations at a very detailed level (e.g., a study of African Americans in West Los Angeles).

The Fragmentation Analysis

For the 'base set', 42 out of a possible maximum of 64 different types of fragments were established (Table 1). Of these 13 were considered to be major types of fragments (each type accounted for more than 1 percent of the total population in the Los Angeles CSMA). In total these 13 major types of fragments contained 90.1 percent of Los Angeles total population. This indicates that in terms of types of shared spaces Los Angeles has a very simple a-spatial fragmented structure. The 'full set' of data established 149 different types of fragments, but again there were only 16 major types. Again this emphasizes the simple nature of major types of fragmentation in Los Angeles. Yet, these 16 major types only accounted for 69.6 percent of Los Angeles total population. This reduction in the total population, accounted for

Table 1: Fragmentation analysis results.

CSMA	*Los Angeles*	*CSMA*	*Los Angeles*
Population (Millions)	16.374	Number of major fragment types in the Base Set	13
Number of census tracts	3361	Number of major fragment types in the Full Set	16
Average population per census tract	4871.8	Percent of total population in major fragment types—Base Set	90.1
Number of fragment types in the Base Set	42	Percent of total population in major fragment types—Full Set	69.6
Number of fragment types in the Full Set	149		

from 90.1 to 69.6 percent, is a measure of how fragmented the structure becomes when we incorporate the degree of concentration into the shared spaces. Minor types of fragments for the 'full set' therefore contain 30.4 percent of Los Angeles population.

Examined by population group, for the 'base set' the Hispanics and Whites have an equal number of total types of fragments (12 each, Table 2), whereas African Americans and Asian populations have a third less (only 8 and 9 respectively). The sharing of areas by the African Americans and the Asians is therefore much less complex than it is for Hispanics and Whites. For the 'full set' the structures of

Table 2: Total fragment types by population group.

Los Angeles	*Majority and dominance (Base Set)*	*Dominance only (Base Set)*	*Total (Base Set)*	*Majority and dominance (Full Set)*	*Dominance only (Full Set)*	*Total (Full Set)*
Hispanic	5	7	12	22	29	51
White	5	7	12	18	26	44
African American	4	4	8	12	10	22
Asian	4	5	9	13	18	31

concentration are much more varied for the Hispanics (51 types) than for any of the other three populations. Of particular note are the African Americans who have the smallest array of fragment types (22 types), half that of the White population (44 types). Asians have approximately fifty percent more total fragmentation types (31 types) than the African Americans, but a third less than the White populations.

Figure 1 displays as a branching tree diagram the relationship between the 'base set' of types of major fragments and the 'full set', where a major fragment type is defined as one which contains one percent or more of the total population. In that figure the standard method of attributing each type of profile is presented, where the population group is specified by a capital letter code followed in lowercase letters by a code to specify majority or dominance status or the threshold concentration level. Hence WmWdW20 indicates a major fragment type where Whites are in the majority, Whites are dominant, and Whites have a level of concentration that exceeds the 20 percent threshold.

As can be seen in Figure 1 the 'base set' of major fragment types include all four populations, although much more complex structures of fragmentation exist where Whites and Hispanics are in the majority. Importantly where Whites are in the majority they share many of those areas with Hispanics and Asians but not with African Americans. By contrast where Hispanics are in the majority they share their areas with all three other populations. African Americans, however, fail to form a single major fragment type by themselves. Instead, they share their only major fragment type with Hispanics. When we view the major fragment types by variation in concentration levels (the 'full set') it is only White and Hispanic populations that have major fragment types that exceed 1 percent of Los Angeles population. In both cases the largest percentages of the total population occur in areas where their concentration levels exceed 70 percent. These are the areas of extreme segregation; the White citadels and the Hispanic barrios. By contrast areas shared by White and Hispanic populations contain only a small percentage of the total population and are areas with much lower levels of concentration. Finally, the African American and Asian major fragments that exist in the 'base set', become too fragmented for them to form major fragment types when the full range of levels of concentration are included.

In recent research Poulsen, Johnston and Forrest (2001) have developed a rules based typology of communities and enclaves. It consists of six types:

Type 1: Isolated Host Community (Citadels);

Type 2: Non-isolated Host Community;

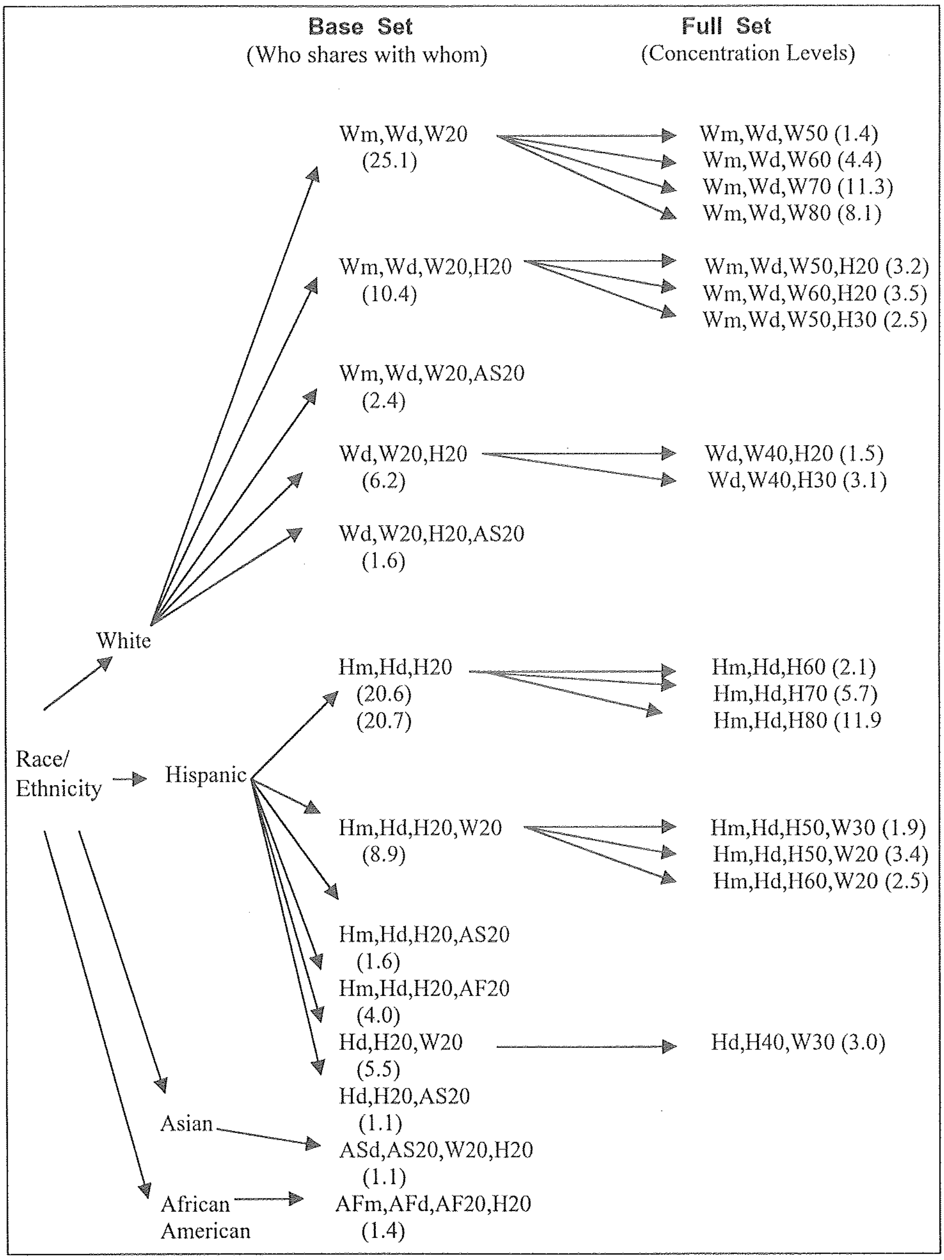
Type 3: Assimilation/pluralism Enclave;

Type 4: Mixed Enclave;

Type 5: Polarized Enclave;

Type 6: Extreme Polarized Enclave (Ghetto).

Figure 1: Branching diagram of the relationship between major types of fragments in the base and full datasets.



W=White; H=Hispanic; AF=African American; AS=Asian; M=majority; d=dominance; 20...80=the concentration levels. The figures in brackets are the percentage of the total population living in each major type of fragment.

As such it establishes a universal homogeneity–heterogeneity–homogeneity continuum using universal measurements along which any community can be classified. If we use this typology to categorize the types of fragments for the 'full set'' a significant finding is achieved. This is that the major fragment types form exclusive subsets within those groups, which reinforces the validity of that rules based system. The isolated host communities of the Whites (the citadels) account for 8 percent of the total population in Los Angeles. This type of area is occupied by fragment types with extreme levels of White concentration (Table 3). At the other extreme, the extreme polarized enclaves contain fragment types of extreme Hispanic concentration and account for a total 22.2 percent of the Los Angeles total population. Heterogeneous areas, which included those classified as assimilation/pluralism enclaves (9.5), mixed enclaves (0.0) and non-isolate host communities (26.3), contain major fragments with shared spaces and account for 35.8 percent of the total population. These results display that Los Angeles is characterized by relatively small proportion of the population living in White citadels, a large percentage living in the Hispanic barrios, with the bulk of the population sharing mixed spaces.

However, a-spatial structures derived from the fragmentation analysis are not spatial structures. We cannot infer that the spatial structure of geographic objects is as simple as the fragmentation structure indicates until we conduct the objectification analysis.

The Objectification Analysis

When we dissolve the boundaries between fragments of the same type, the number of race/ethnic geographic objects established in Los Angeles is 985 for the 'base set' and 1924 for the 'full set'. Given that there are 3361 census tracts in Los Angeles the output geography for the 'base set' has 29.3 percent as many observations as the number of census tracts in Los Angeles and for the 'full set' it is 57.2 percent. This later figure suggests that a large number of geographic objects is a feature of Los Angeles. Of the major types (those which contain one percent or more of Los Angeles total population) 722 (73.3 percent) are part of the 'base set' and 1015 (52.8 percent) part of the 'full set'. Minor types of geographic objects (47.2 percent) therefore are a significant feature of Los Angeles when we take into account the different levels of concentration (the 'full set'). Given the simple major fragmentation structure of Los Angeles measured in the fragmentation analysis we might have also expected a simple structure of geographic objects. Instead, the objectification analysis has established that Los Angeles provides us with a model of an extremely fractured city.

The spatial nature of these output geographies as expected supports Marcuse's (1997) model of the American city (citadels–enclaves–ghettos). However, in Los Angeles the reduced nature of the citadels compared to those that are evident in New York and Chicago, the large area of the Hispanic barrio, the large mix of populations, and the lack of major types of African-American concentration are

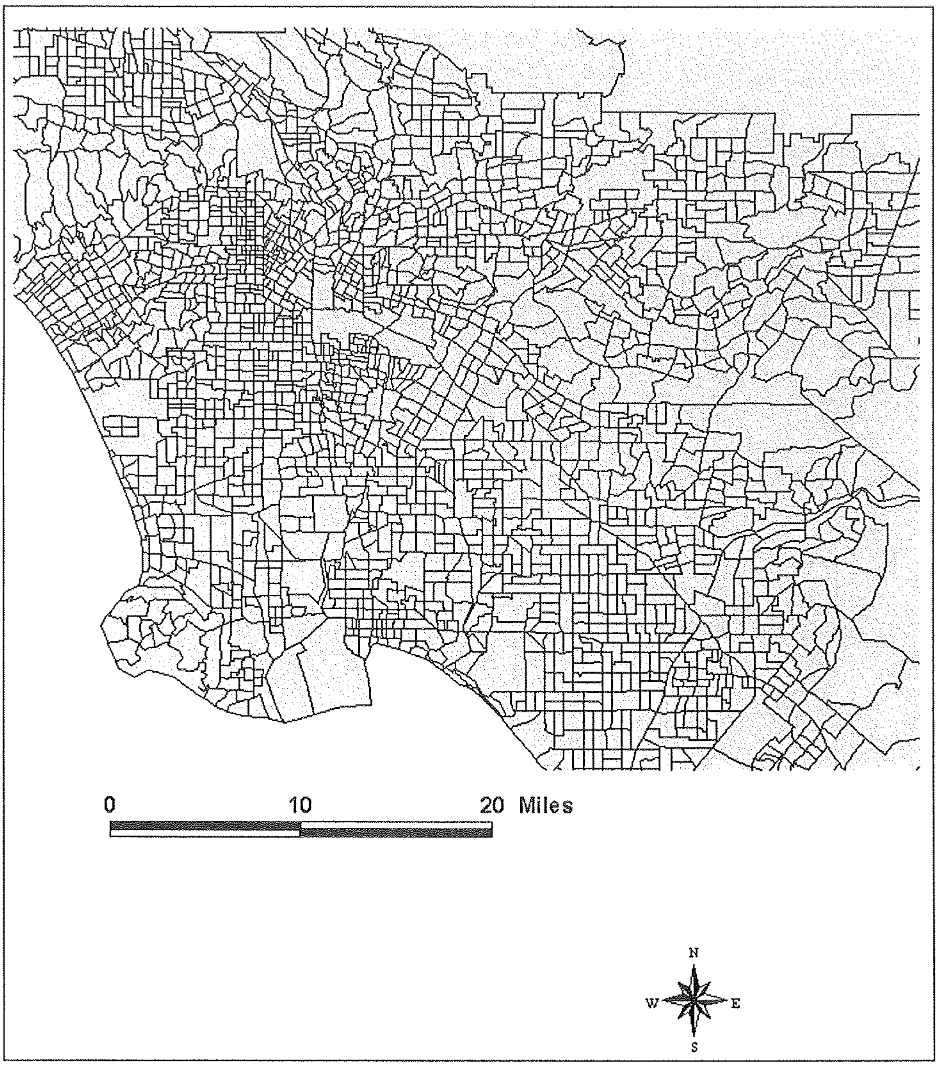
Table 3: Typology of communities and the major fragments of the full dataset.

Typology	*Geographic Object Type*	*Percent of Total Population*
Isolated host society (citadels)		
	Wm,Wd,W80	8.1
Non-isolated host society		
	Wm,Wd,W50	1.4
	Wm,Wd,W60	4.4
	Wm,Wd,W70	11.3
	Wm,Wd,W50,H20	3.2
	Wm,Wd,W60,H20	3.5
	Wm,Wd,W50,H30	2.5
Assimilation/Pluralism enclaves		
	Wd,W40,H20	1.5
	Wd,W40,H30	3.1
	Hd,H40,W30	3.0
	Hm,Hd,H50,W30	1.9
Mixed enclaves		
	None	
Polarized enclaves		
	Hm,Hd,H50,W20	3.4
Extreme polarization (ghettos)		
	Hm,Hd,H60	2.1
	Hm,Hd,H70	5.7
	Hm,Hd,H80	11.9
	Hm,Hd,H60,W20	2.5

W=White; H=Hispanic; AF=African American; AS= Asian.
M=majority; d=dominance; 20...80= the concentration levels.

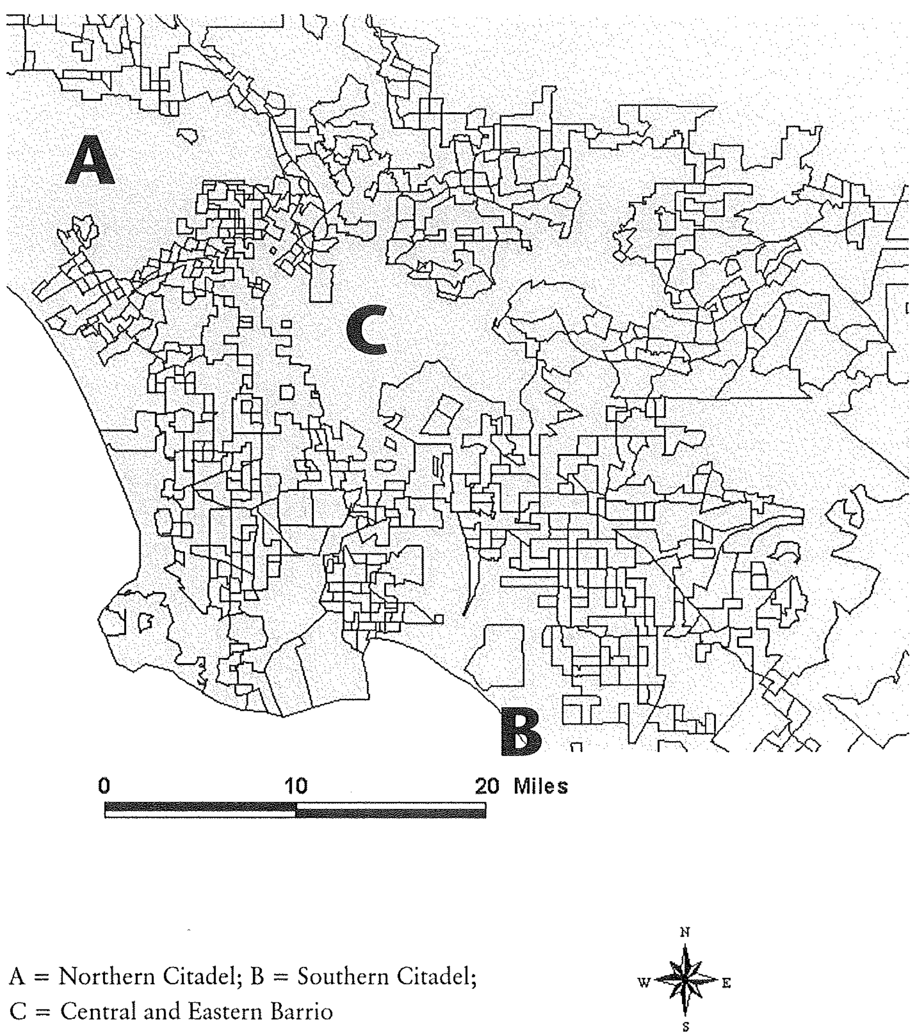
all major and distinctive features of the city. To examine the spatial location and extent of these geographic objects three maps are presented. Figure 2 displays for part of Los Angeles the original census tracts that were subject to the fragmentation and objectification analyses and Figure 3 the new output geography for the 'base set' of geographic objects. By comparing Figures 2 and 3 this objectification analysis reveals the three major homogeneous areas in Los Angeles: the northern White citadel extending westwards from Beverley Hills to Santa Monica and then north; the southern White citadel extending along the coast south from Huntington Beach to San Clemente; and the central and eastern Hispanic barrio extending east from LA Central, through East Los Angeles and north eastward. A few small

Figure 2: Census tracts of central and western Los Angeles.



geographic objects lie west of this barrio within which high levels of polarization by African Americans occurs, but these by definition are minor geographic objects. Between them and abutting the barrio is an array of different minor geographic objects where Hispanics are in the majority, but where there are concentrations of African Americans. By contrast, the areas where the Hispanics are in the majority and they share with Asians are north of the barrio. Importantly the new output geography for the 'full set' of concentration levels (Figure 4) displays that these major geographic objects continue to exist as single units largely undifferentiated

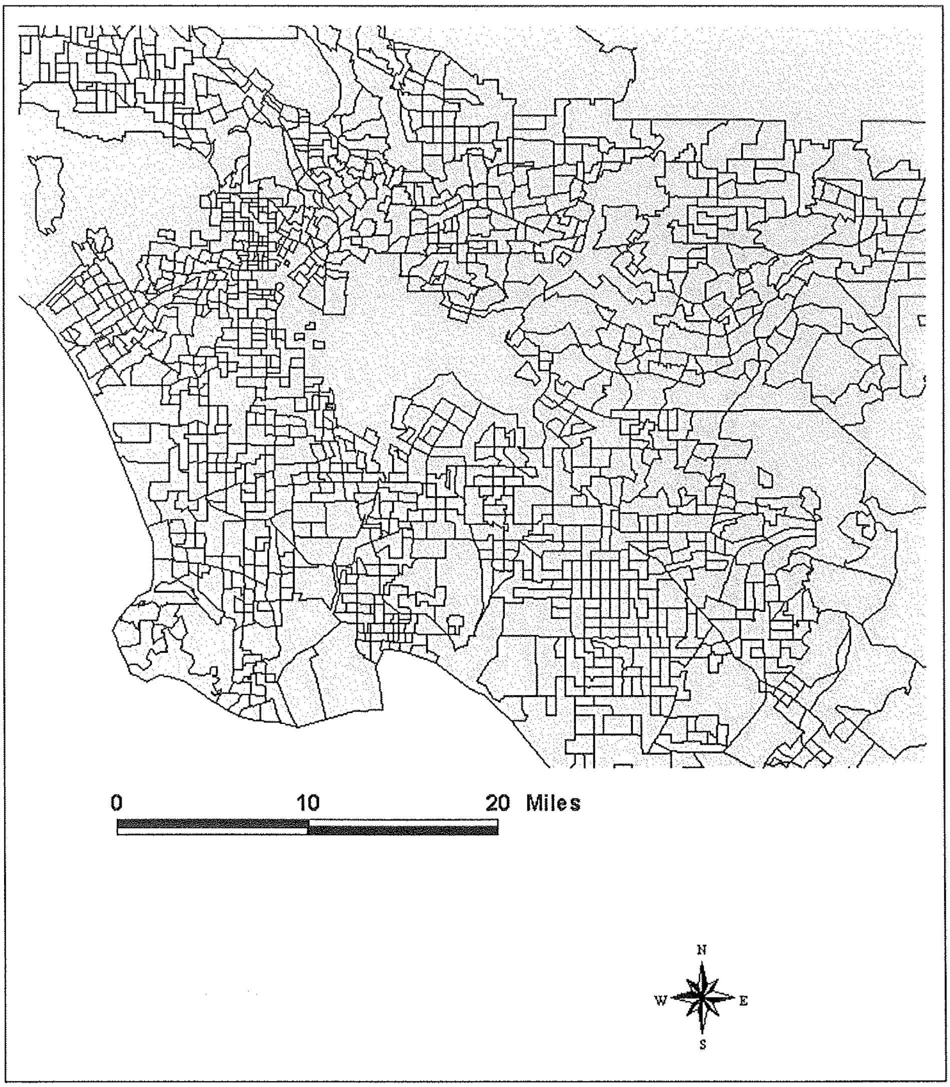
Figure 3: Objectification analysis: 'Base set' output geography of central and western Los Angeles.



A = Northern Citadel; B = Southern Citadel;
C = Central and Eastern Barrio

by different levels of concentration; although the eastern extreme of the barrio does feature some differences in levels of concentration. As a result, we can conclude that the increased levels of fragmentation that we have measured between the 'base set' and the 'full set' in the fragmentation analyses occurs in the spaces between and beyond the White's citadels and the Hispanic's barrio, within which live different mixes of populations.

Figure 4: Objectification analysis: 'Full set' output geography of central and western Los Angeles.



CONCLUSION

GIS has in recent times provided researchers working with census geography with new ways by which we can research that dataset for information. The need to mine that data for information is fundamental to our developing information age, where increasingly greater reliance is being placed on such evidence in the management of our cities. Central to this work has been the development of new census geographies. This includes the integration of census data with imagery for the purpose of improving remote sensing classifications, the integration of imagery with census data to improve census datasets and our use of it, and research into the nature of census-based geographic objects. Research into fragmentation and geographic objects does not have the same objectives as earlier work in the 1960s to 1980s on factorial ecologies and classification typologies. While it is again concerned with urban structure it originates from research in GIS and remote sensing in the physical domain into the importance in studying geographic objects with crisp or fuzzy boundaries. Prior to this focus on fragmentation and geographic objects the use of census data in GIS was largely restricted to choropleth mapping. Today, through the advent of the fragmentation and objectification analyses, we are able to study the nature and evolution of economic, social, population, ethnic geographic and housing objects. Most importantly, through the adoption of a standard methodology and nomenclature for attribution, we are able to acquire local and detailed measurements and information where previously a vacuum had existed.

New output geographies have created a new era in census geography. In terms of comparative and temporal research the a-spatial nature of fragment types and the spatial nature of census-based geographic objects have become a major focus of attention. While it is unlikely that we will see governments switch there funding policies from the census-based administrative units to that of the geographic objects for some while, these structures will become increasingly a major research focus. This is because they are the base elements of urban structure, they evolve over time, and they provide us with measures and in turn information that is comparable between different cities both nationally and internationally. Importantly, in the immediate future they provide a critical source of information about our cities, which we cannot afford to ignore. To become entirely engrossed in discourse about cities and ignore measurements of urban structure—empirical evidence—is to distort urban research. Just as research into geographic objects is a primary focus of research in remote sensing and digital surface modeling using GIS a feature of research within physical geography, research into fragment types and geographic objects derived from census datasets will continue to develop as a major focus of research in human geography. Of concern are a priority list of major research topics that need investigation, including: the evolution of major citadels, enclaves and ghettos (barrios); how concentration levels and fragmentation is evolving; how policy, political, institutional, economic and social settings facilitate fragmentation and the evolution of geographic objects; auditing the outcomes of government and urban management policies through the study of fragment types and geographic objects; projecting the evolution of these geographic objects, and assessing how accurate our projections are. Now that we have a standard methodology by which we can acquire a set of universal measurements from which we derive important information, a major test of future GIS and urban research lies squarely in our ability to use it and develop these new procedures.

NOTES

1. A geographic entity is a real world feature whereas a geographic object is the digital representation of it in a computer (Frank, 1996).
2. We add 1 to the number of levels of concentration to take into account the discrete value for dominance. If we have measures of both majority and dominance in the analysis we include the multiplier of 2 in so that we can have a set of census enumeration areas in which one population is in a majority and thereby dominant, and another set of census enumeration areas where no population is in the majority but one or more are dominant. If we have only a majority or dominance measure we remove the final multiplier. Likewise, if we have neither a measure of majority or dominance we remove the +1 value.

In comparing the number of types of observed fragments against the maximum possible it is necessary to exclude mixed types from the count where two different populations have 50 percent of the population in a census area, or where there is more than one population in a census area that is the dominant group. These occurrences are rare, however we cannot predict them.

REFERENCES

Burrough, P.A. and Frank, A.U. (eds.) (1996) *Geographic Objects with Intermediate Boundaries.* London: Taylor and Francis.

Couclelis, H. (1996) Towards an operational typology of geographic entities with ill-defined boundaries. In Burrough, P.A. and Frank, A.U. (eds.) *Geographic Objects with Intermediate Boundaries.* London: Taylor and Francis, pp. 45–56.

Frank, A.U. (1996) The prevalence of objects with sharp boundaries in GIS. In Burrough, P.A. and Frank, A.U. (eds.) *Geographic Objects with Intermediate Boundaries.* London: Taylor and Francis, pp. 29–40.

Hall, G.B., Malcolm, N.W. and Piwowar, J.M. (2000) Integration of remote sensing and GIS to detect pockets of urban poverty: The case of Rosario, Argentina. *Transactions in GIS,* 4:235–253

Harris, R.J. and Longley, P.A. (2000) New data and approaches for urban analysis: Modeling residential densities. *Transactions in GIS,* 4:217–234.

Marcuse, P. (1997) The enclave, the citadel, and the ghetto: What has changed in the post-Fordist U.S. city. *Urban Affairs Review,* 33:228–264.

Martin, D. (1998) Optimizing census geography: The separation of collection and output geographies. *International Journal of Geographic Information Science,* 12:673–685.

Martin, D. and Bracken, I. (1991) Techniques for modeling population-related raster databases. *Environment and Planning A*, 23:1069–1075.

Martin, D. Tate, N.J. and Langford, M. (2000) Refining population surface models: Experiments with Northern Ireland census data. *Transactions in GIS,* 4: 343–360.

Mesev, V. (1998) The use of census data in urban image classification. *Photogrammetric Engineering and Remote Sensing*, 64:431–438.

Openshaw, S. and Rao, L. (1995) Algorithms for reengineering 1991 Census Geography. *Environment and Planning A,* 27:425–446.

Poulsen, M. and Johnston, R.J. (2000) Threshold analysis approach to the objectification of social space: The structure of household types in Sydney. *Environment and Planning B: Planning and Design,* 27:905–922.

Poulsen, M., Johnston, R.J. and Forrest, J. (2001) Intra-urban ethnic enclaves: Introducing a knowledge-based classification method. *Environment and Planning A*, 33:2071–2083.

Thurstain-Goodwin, M. and Unwin, D. (2000) Defining and delineating the central area of towns for statistical monitoring using continuous surface representations. *Transactions in GIS,* 4:305–317.