2010 Census Block to Subzone09 Assignment Procedures

This document addresses the assignment of the census blocks in the 7-county CMAP region, and provides the general GIS-based approach to CMAP's block-to-Subzone assignments for 2010 Census data. If further information is needed, contact David Clark, dcclark@cmap.illinois.gov

OVERVIEW:

The goal of this project is to assign all 2010 Census blocks in the 21-county CMAP modeling area to subzones (2009 version), for the purposes of creating a new base year socio-economic dataset for transportation modeling and forecasting activities.

The one inviolable rule is this: Census Blocks code to one and only one subzone. In densely-populated areas, this is not an issue because blocks are small and many of them easily fit inside of a subzone. This is helped further by the fact that the 2010 Census geography is greatly cleaned-up in some areas, so blocks that follow PLS section/quarter-section lines don't have nearly the amount of slop that they used to have.

This approach to block assignment is similar to the one used for block-to-quartersection assignments after the 2000 Census, where blocks were assigned through various stages:

1. Block fits entirely (or almost entirely) within zone, so it automatically is assigned to the zone;
2. Block is unpopulated, so assignment is based on location of block's centroid;
3. Remaining blocks are manually assigned based on evidence of which zone the bulk of the population resides in.

Revised approach for 2010 blocks. Assignment tasks were carried out in the order listed:

Method 1: Unpopulated blocks are assigned to whatever subzone the block's centroid falls in.

Method 2: Under-populated blocks (pop < 10) are assigned to whichever subzone has more than 50% of the block's area (if there is no 50% because it is in three or more blocks, it doesn't get assigned in this step).

Method 3: Populated blocks with at least 85% of block area in a single subzone code to that subzone.

Method 3.5: In counties where parcel GIS data includes enough information to generate a housing unit (HU) estimate, parcel centroids with HU estimates are compared to HU estimates for block/subzone intersect areas; where totals are similar, block codes to the subzone that has the highest HU estimate from the parcel data. *(named "3.5" because it was thought up after methods 1-4 were already developed)*

Method 4: The remainder are coded manually in ArcGIS, with the assistance of customized tools.

Method 1: Unpopulated blocks

- Select all census blocks with zero population (P0010001 = 0)
- Create centroid feature class (INSIDE = Y) of selected polygons
- Run spatial join of centroids to Subzones to get Subzone ID
- Join output to Assignments table; calculate Subzone assignment and Method (1)

Method 2: Underpopulated blocks:

- Select blocks where population (P0010001) is greater than 0 but less than 10, AND where the housing unit count is less than 5; create new feature class of these blocks.
- Run Identity with Subzones; calculate the percentage of each block's total area within each block/subzone combination.
- Select from this feature class those features which have more than 50% of the block's area within a single subzone.
- Join selected set to Assignments table; calculate Subzone assignment and Method (2)

Method 3: Blocks entirely or almost completely within a subzone. Minimum 85% of a block has to fall within a subzone to code to it without further work.

- Create feature class of unassigned blocks and run Identity with Subzones
- Calculate the percentage of each block's total area within each block/subzone combination
- Select from this feature class those features which have more than 85% of the block's area within a single subzone.
- Join selected set to Assignments table; calculate Subzone assignment and Method (3)

Method 3.5: Using parcel data

This method was dreamed up while looking at the daunting task of manually assigning the remaining 13,000 blocks in CMAP7_4. This involves using 2010 parcel data, which has been collected for the Land Use Inventory. All counties except for Cook have a parcel-based residential unit estimate generated at CMAP, based on each parcel file's attribute data (procedures varied by county due to intra-county inconsistencies in data maintenance). The theory is to use parcel centroids with residential unit estimates to find which block/subzone combination has the highest number of res units for each block. Procedures are run on a county-by-county basis.

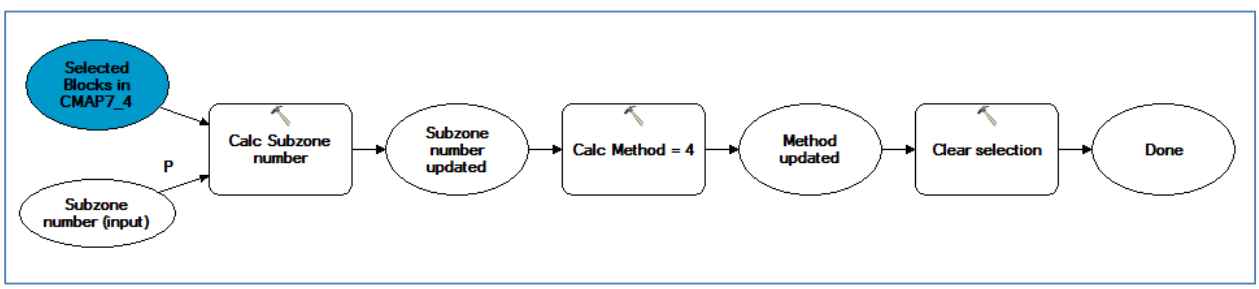
- Select un-assigned census blocks for the county.
- Remove from selection set those blocks that have a group quarters population.

- Run a spatial join of parcel centroids (with HU estimate) with census blocks; summarize parcel HU estimate by census block.
- Select census blocks where the parcel HU estimate is within 10% of the census block's HU count (meaning, the parcel data will serve as a reliable resource for this method)
- Intersect qualifying census blocks with Subzones; run spatial join of parcel centroids through the block/Subzone combinations.
- Summarize parcel HU counts by each block/Subzone combination.
- Assign block to Subzone based on which block/Subzone combo had the highest number of (parcel-derived) housing units.
- Generate list of assigned subzones, join to Assignments table, calculate Subzone assignment and Method (5).

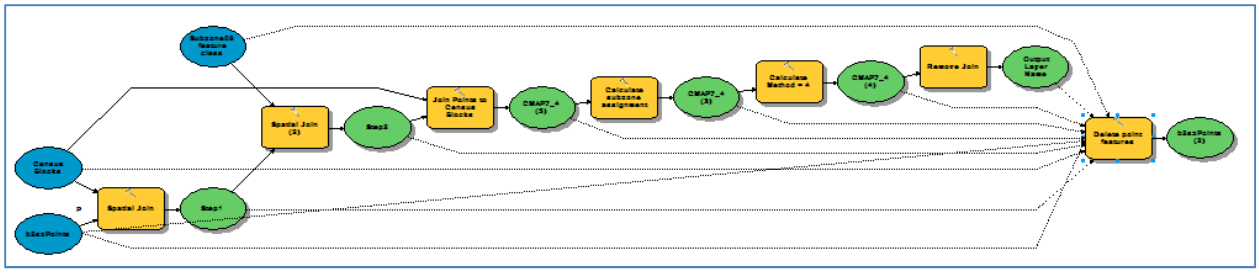
Method 4: The remainder

By this stage, there were only 6,947 blocks remaining (out of 169,489) that required subjective review. We created two ModelBuilder routines which speed up this process. One works better in areas of smaller blocks where there might be many blocks coding to a single subzone ("Polygon Approach"); the other works better in the more rural areas ("Point Approach").

Polygon Approach: Select all census blocks that will code to a specific subzone. Running this model will give you a window where you type in the subzone number. Number is automatically calculated into the block table, Method is updated to "4," and then the selected blocks are un-selected. Requires a screen refresh afterwards (F5).



Point Approach: This is a slightly more complex model, where you work with a point feature class, inserting a point for each block; you place the point so it falls not only within the block, but also in the subzone that the block is being assigned to. Model runs spatial joins to the block and subzone layers to create a block/subzone table, which is then joined back to the blocks layer (on the block ID) to calculate in the subzone number; method is then updated to "4," join is removed, and all point features are deleted so you can start fresh.



Quality Control

1. <u>Is a subzone assigned to every block?</u> No. There were 38 blocks without assignment. They turned out to be lakefront blocks whose centroids fell in the lake, outside of our Subzone polygons. All have a population of zero, so were assigned to the nearest Subzone using the ArcGIS "Near" function.

2. <u>Verify that blocks' state/county designations match</u> with assigned subzone st/co designations. This can happen due to the combination of the messy Census geography along state/county boundaries not lining up well with our Subzone shapefile. This will most likely happen with small sliver blocks along county boundaries, and probably were created during the various automated stages. It is important that any errors here are corrected so that any county-level summaries match between subzone and census data. There were a total of 89 blocks whose county FIPS did not match the county code in the assigned subzone. These were manually inspected and corrected.

3. <u>Check for mis-codings</u>. This should only be necessary for the Method 4 bunch, since all of the others relied entirely on automated procedures. The assumption here is that at least <u>some part</u> of a block intersects with the subzone it's coded to. So we (1) create a block shapefile of Method 4 blocks, with SZ assignment numbers, (2) create a combo field containing the block/subzone number as a single concatenated value, then (3) intersect it with the subzone shapefile and create a new combo field that has every block/subzone combination. (4) If a combination in the first table doesn't have a match in the second combo, that means the block doesn't even touch the subzone that it's assigned to, and could be a mis-coding. One record was found as incorrect through this procedure, and was manually corrected.

Assignment Table Summary

	COUNT	PERCENT
Total blocks	169,489	
Unpopulated Lake Michigan blocks	102	<0.1%
Method 1	45,727	27.0%
Method 2	10,491	6.2%
Method 3	99,987	59.0%
Method 3.5	6,235	3.7%
Method 4	6,947	4.1%

Block to Subzone09 completed, 5 April 2011 dcc