

## Appendix Five

# The Congruence of LandScan USA Boundaries with U.S. Census Block Boundaries

Gerard Rushton and Qiang Cai

The University of Iowa

Iowa Cancer Registry

May, 2004

This research was supported by SEER Special Project #08 “Development of High Resolution Population Distribution Data to Enhance Cancer Prevention and Control Research” (NCI-PC-25014-20). All conclusions are the responsibility of the authors.

---

This paper addresses two problems. The first problem is how LandScan USA population data is estimated when the boundaries of the population grid areas are not congruent with the boundaries of the U.S. Census blocks on which their populations are based. The second problem concerns error propagation when the boundaries of Census Blocks in the TIGER database have geometric error which results in incorrect registration of the population data with data observed from sensor inputs from remotely sensed data. Details of how these problems were addressed were not found in the materials provided to us either at the time of delivery of the population data by the ORNL GIS group nor in their final report (see Appendix One). Methods used by the ORNL group to solve these problems are regarded as proprietary information, we were told. The purposes of the NCI contract cannot be reached, however, without considering how these problems were solved and the likely population estimation errors that inevitably occurred.

When an overlay operation in a GIS is made between a polygon layer and a grid layer, it is inevitable that some cells will cross and straddle the boundaries of the polygons. The larger the cell size in relation to the size of the Census Block, the more geographic area will lie in this area of overlap. The LandScan USA model needs to deal with this problem caused by the fact that the population grid boundaries are not conterminous with the Census Block boundaries. One simple way to deal with this problem would be to use the centroid of a cell to determine which Block the cell should belong to. This approach totally avoids the problem just discussed but introduces considerable error where the percent of the grid areas in the overlap area between Census Blocks is high. It is clear from the information below that LandScan USA uses a different approach:

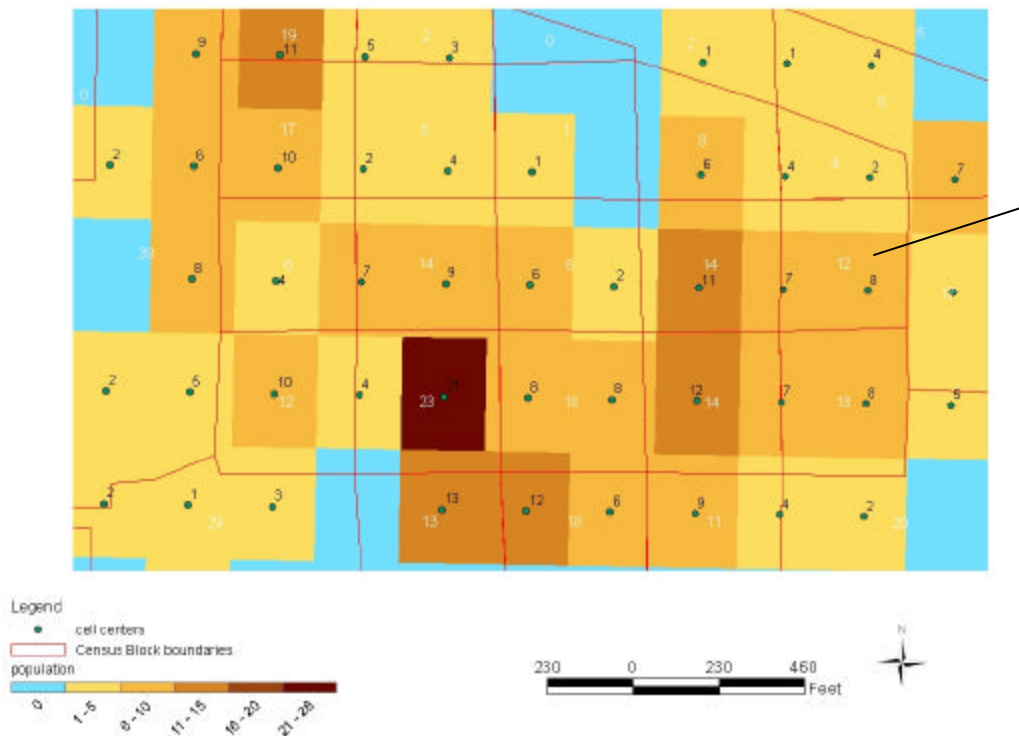


Fig. 1 Illustration of the problem that population cells cross Census Block boundaries and that spatial interpolation is necessary to deal with such spatially misaligned data. The map is of a portion of Carroll County, Iowa.

Some important information can be derived from Fig. 1. In this Figure, green points are centroids of the populated cells. From the relative position of these points and block boundaries; it is easy to see that many cells fall across Block boundaries. The white number at the center of each Census Block is the total population (from census 2000) of that Block. The black numbers near the cell centroids are LandScan U.S.A. populations of the cells. For a certain Block, the Block population usually does not equal the total population of cells with centroids that fall inside the Block boundary. For example, the Census Block that the arrow points to has 12 people according to Census 2000, while the sum of the cell populations is  $7 + 8 = 15$ . It is more likely that those cells that cross Census Block boundaries will be registered to more than one Block, and it appears likely that the membership function was based on the proportion of the cell areas that are in the respective Blocks. It is not clear what spatial interpolation function is used to assign population values to the cells.

The second problem on error propagation of the population estimates due to geometric error in the spatial description of Census Blocks in the TIGER Line file is illustrated in Fig. 2.

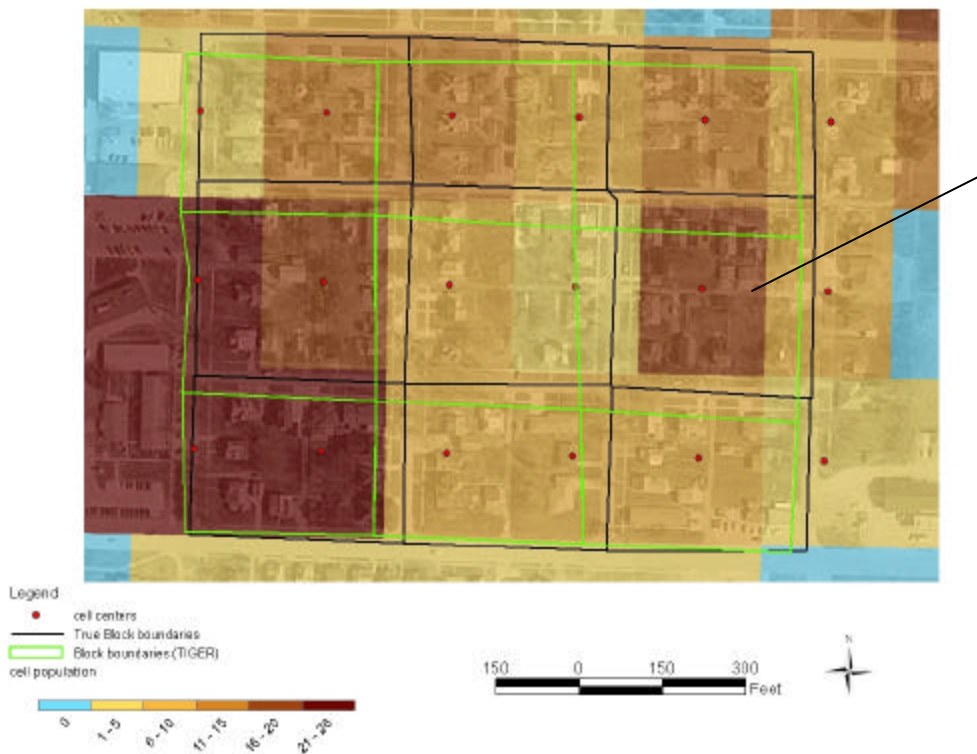


Fig. 2 Illustration of the problem of incorrectly registered cells caused by geometrical errors in the TIGER Line file.

In Fig 2, green polygons are Block boundaries from TIGER, black polygons are digitized Block boundaries based on the overlaid and registered orthophoto map. These are the geometrically more accurate Census Block boundaries in this case. The red points are centroids of populated cells and the semi-transparent color squares are cells in the LandScan USA population grid.

In the cell to which the arrow points in Fig.2, it is obvious that the whole cell falls inside one Block according to the true Block boundaries (Black polygons) and that its population should only be attributed to that particular Block. But when TIGER line files (green polygons) are used to deal with the same cell, it will be attributed to two neighboring Blocks because it crosses their boundaries. The ORNL GIS group report does not explain how LandScan USA

population data were manipulated to solve this problem nor does it estimate probable population estimation errors that may be caused by the problems discussed in this Appendix.