DEVELOPING A GIS TOOL TO ESTIMATE SMALL ZONAL POPULATION FOR LAND-USE AND TRANSPORTATION MODELS

Yang Zhong-Zhen
Chief Researcher
Nagoya Industrial Science Research Institute
Anagahora 2271-130, Shidami, Moriyama, Nagoya 463, Japan
E-mail: yanglian@gld.mmtr.or.jp

Abstract: This study aims to develop a GIS tool, which can estimate small zonal population with a high accuracy and feasibility for land-use and transportation model. With this tool accurate zonal population can be obtained automatically after census database is established. The tool makes a visual view, which is easy for users to interface with the computer. With this tool, zoning for land-use and transportation model becomes much more flexible without inducing further work of population survey. This tool can be further developed into a GUI platform for land-use and transportation integrated model.

1. BACKGROUND

Many integrated land-use and transportation model have been developed. In these models, the first coming sub-model is always the population forecasting model, because forecasted population is deadly needed for calculating zonal generated and attracted trips. Even though population-forecasting sub-model is relatively simple compared to other sub-models, the forecasted population has a strong influence on the accuracy of the whole-integrated model. It means that population data is the basic data for the integrated land-use and transportation model. Without detailed time series zonal population data, an actual-like future population for each zone could not be calculated. However, gaining population data is not an easy work. Especially, when land-use and transportation models are integrated into a system, urban area might be divided into thousand zones. Collecting zonal data for population forecasting model is always an expensive and time-cost work.

Recently, developments in the field of GIS help to simulate urban models. No one has doubted that GIS will ultimately make large-scale urban models thrive again and that land-use and transportation model integration will be intensively improved (Kim, Park & Lee, 1998). In order to provide adequate functions for the integration of land-use and transportation models, GIS technologies will primarily assist in identifying connectivity, contiguity, and proximity on a complicated urban transport network. GIS technologies can deal with more complicated urban structure, and then they can surely help the urban planner to forecast the future urban trend, to create developing plan and to make some decisions with an interface tool. In order to integrate land-use and transportation models into a GIS system, lots of databases are needed to establish. It is obviously that the traditional data surveying, collecting methods could no longer satisfy the new trends. Developing new data surveying methods, especially a GUI (graphic user interface) surveying tool, which is embedded into GIS-oriented land-use and transportation model become one of the key problems.

This study aims to develop a GIS tool, which can estimate small zonal population in land-use and transportation model easily. A platform, which is used for user-interface operation,

is created by "Visual Basic". Maps managed by Mapinfo and databases supported by "MS-Access" are integrated in the platform. According to the zoning design, zonal population can be calculated by using the "menu" or "function button" on the platform. For example, when the zonal layout in the map-window is change by mouse, new zonal population can be calculated immediately. Basic population data sources for this tool are census data.

2. EXISTING DATA AND STUDY METHOD

The tool developed here takes Japan's cities as the study areas. Main data for urban population surveying is from census, which is implemented every 5 yeas in Japan. Normal census process and the available data are stated below. First, an urban area is divided into many census zones, the zonal layout and series numbers will be indicated in 1/2500 urban maps which copies can be obtained from city government afterward. Second, based on the zonal map, staff will carry out the field survey house-by-house to obtain households information such as male and female members, members' names, ages and employment. Finally, above collected information will be aggregated to become several kinds of population data, which will be published as census results.

The results of the above census survey are as follows:
1.1/2500 census map with zonal layout and zonal serial numbers;
2.Building-draft maps with households' position and numbers for each zone;
3.Lists of members of each household number, generally including male and female members (sometimes including employment situation).
4.Statistic census tables published by governments (local or central government), including summed-up and adjusted data.
The data 1, 2 and 3 can not be obtained unless an approval from city governments is gotten. Final table, which show the surveying results of both total urban population and population for each census zone, can be bought without government approvals,

Actually, census is offering plentiful time series and detailed population data for urban planning. However, the problem is that generally census zones are usually either larger or different from the zones in land-use transportation integrated model (hereafter we call them as urban planning zones), therefore, the original census data can not be directly used in the model. In GIS, if you overlay the urban planning zones to census ones, data can be transferred by areal interpolation methods. Figure 1 shows the situation, data in the circle can be estimated from the data in zones 2-5 by the ratios of their areas within the circle (SADAHIRO, 1999). If the data in zones 2-5 distribute randomly, the estimating error may be accepted, however we can not expect the residential houses distribute randomly. Then in the integrated model how to use census data, which is constructed by government regularly, to get the accurate data is an essential problem.

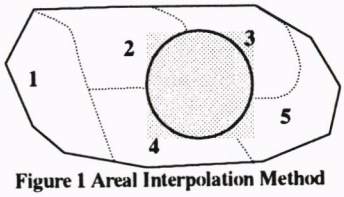

Figure 1 Areal Interpolation Method

Till now, almost all of the existing methods of calculating urban planning zonal population from census zones are treaded by "manual graphic methods". Their procedures can be normally described as follows.
1.Transferring boundaries of urban planning zones to the 1/2500 census maps;

2. Basing on the topographical features, comparing household drafts with census maps to draw urban planning zones' boundaries into household drafts;
3. Counting household numbers and recording the numbers for each urban planning zone in the house draft map;
4. Finding household's members for each household by matching the numbers in both household drafts and household lists;
5. Typing the gotten data to some software such as MS-Excel or Lotus1-2-3 to sum up the population for urban planning zones.

It is obviously that in this process, lots of works are done by hand rather than by computer. The shortages for this method are those, 1.working process can not be repeated easily. In the surveying, the unique check that can find working mistake is to compare the census population with the calculated population in the last step. If there is any error found, almost all of the done works have to be repeated again; 2.This process could not be linked to land use-transport model as a segment automatically.

In fact, in urban land-use and transportation planning, zoning system changes from time to time since zoning is an iterated process, it is impossible for us to decide zonal boundaries once. Therefore, a method that can be repeated easily is necessary in calculating urban planning zone's population. Adopting GIS will alleviate a great deal of data inventory problems and can embed population-surveying process into the integrated models automatically. Analyzing the above available data we can find that there exist two kinds of data. One is topographical data such as census zone, which can be managed by GIS software. Another is text data such as households' list, which can be treated by Database. In GIS, If we link topographical data to the Database, an automatic working process could be established.

3. DIGITAL MAPS AND DATABASE TABLE

3.1 Several Layers in GIS

When the integrated land-use and transportation model is used in an urban area, first the area will be divided into smaller internal zones. Zonal numbers will depend on a compromise between a series of criteria (J.de D. Ortuzar *et,al* 1990). For example, the analysis of traffic management schemes will generally call for smaller zones, often representing even car parks or major generates of trips. Strategic studies, on the other hand, will often be carried out on the basis of much larger zones. According to J.de D. Ortuzar, the following is a list of zoning criteria, which has been compiled from experience in several practical studies.

1. Zoning size must be such that the aggregation error caused by the assumption that all activities are concentrated at the centroid is not too large.
2. Zoning system must be compatible with other administrative divisions, particularly with census zones.
3. Zones should be as homogeneous as possible in their land use and/or population composition.
4. Zone boundaries must be compatible with cordons and screen lines and with those of previous zoning system.
5. The shape of the zones should allow an easy determination of their centroid connectors.
6. Zones do not have to be of equal size.

According to both the above regulations and some facts of Japan's urban area, we create a digital map in which urban planning zones are depicted. The zoning principle in this map is privileged as follows. 1. City's Administrative Boundary, 2. City's Urban Planning Boundary, 3. City's Urbanization and Urbanization Adjusting Boundary, 4. Primary School Ward Boundary, 5. Streets Boundary, 6. Land-Use Category Boundary.

Figure 2 shows the layer for a Japan's urban planning zone, at now information about its zones can include urban plan code, land-use code etc. If we overlap census layer upon urban planning layer, We can get a zonal map that can satisfy land use-transportation model. Secondly, we digitize 1/2500 census map with digitizer to get another layer as illustrated in Figure 3. This layer just includes some geographical information such as zone's ID and centroid. In this stage no social information like household or population exists. If we want it to become a useful GIS layer we have to add the needed social information onto it.

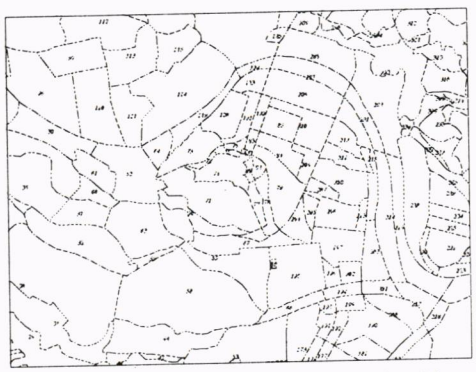

Figure 2 Layer of Urban Planning Zone

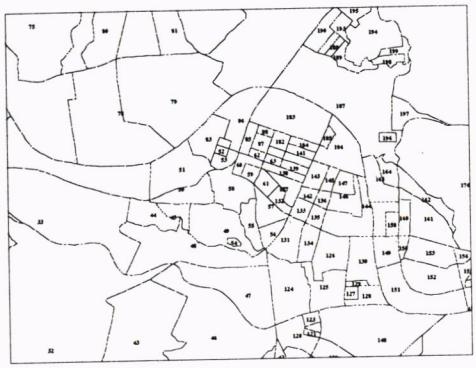

Figure 3 Layer of Census Zones

Another census geographical information is household draft, in which household positions are indicated in the context of census zonal serial numbers. These numbers are digitized to become the third layer named household layer as shown in Figure 4, in which households are treated as point objects with geographical position and serial numbers. The last geographic layer is residential building layer. Generally, in Japan the digital map can be obtained rather than created since an existing building layer should exist. It will be used for error check to find mistake in the working process.

Figure 4 Household Points Layer

Figure 5 Residential Buildings Layer

Till now we have obtained four separated geographic layers. Among them census zones, household points and residential buildings layers are existing data, which will not change during the working process, their digital-type data can be bought in Japan from some agencies. This can ease the work of inputting large amount of data through keyboard and save lot of working time. On the contrary, decision of urban planning zones is a feedback process, government, research institutes or some technical companies might involve in this process. Therefore, without an automatic tool for treating the afterward works, changing of urban planning zones means waste of money and time. We develop a GIS tool to automate the small zonal population surveying based on the mentioned four layers.

4. TOOL FOR TREATING GEOGRAPHICAL AND SOCIAL DATA

We use "Visual Basic" to create a plateform, in which a picture-box is designed to load and display the operation of the geographic layers. In addtion, some manu iterms are desinded to edit or overlap the geographic layers. We also create some event buttons onto the plateform, which will deat with the "Access Table" to distribute or summming up poupulation data. This plateform is shown in Figure 6.

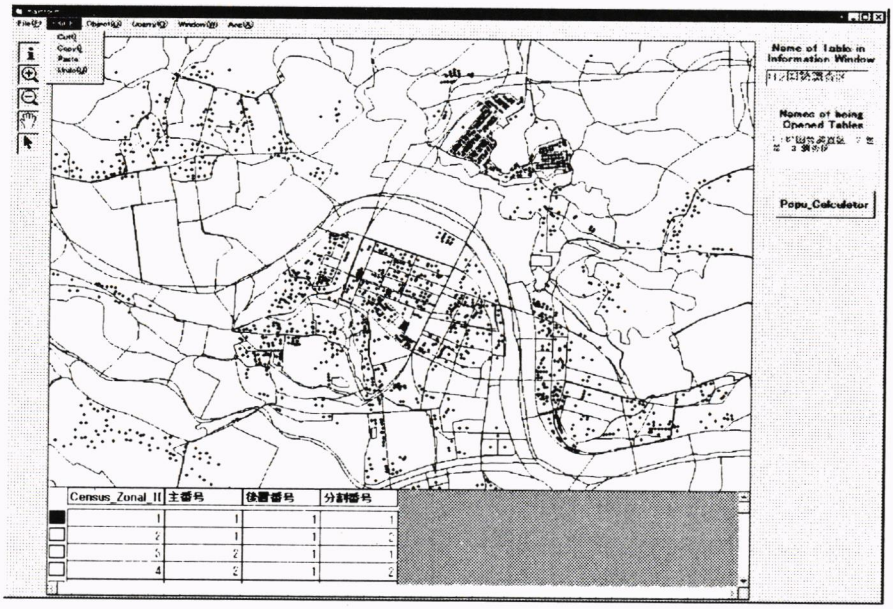

Figure 6 Platform of the GIS Tool

4.1 Overlapping Census Layer onto Urban Planning Layer

Generally, census zones are much different from urban planning zones, calculating urban planning zonal population with census zones data needs overlapping the two layers to each other. A procedure is created to deal with this overlapping automatically. Figure 7 explains the two overlapped original layers, and Figure 9 is the result of the overlapping - a new layer in GIS. Its geographic data are smallest polygons in the new layer, and the text information, which can also be calculated in the procedure, and input to a table such as

Table 1, is the relationships among the polygons in the tress layers.

Table 1 Relation among Three Layers

Polygon ID	Planning Zone ID	Census Zone ID
1	I	J
2	I	J
....	...	...
N	M	L

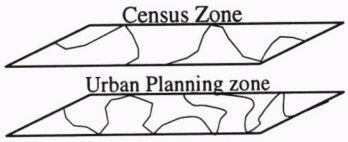

Figure 7 Image of the Overlapped Layers

4.2 Overlapping Residential-Building Layer onto Smallest Polygon Layer

In order to save time and cost, we will not project all households as points into the geographical layer. We will explain this in detail in next section. It is rational to say that the zones without residential buildings should not have households. Therefore, if we previously know that among several smallest polygons, which make up a census zone, only one of them has residential buildings, then zonal households will totally concentrate in this polygon to leave other relevant polygons without households. Then we have to overlap the building layer onto the smallest polygons to find each polygon's residential building amount. This process is illustrated in Figure 8 and Table 2 is gotten.

4.3 Overlapping Household Points Layer onto Smallest Polygon Layer

Till now, except the information in Tables 1 and 2, there is not other information for the smallest polygons. Then the third procedure will be able to overlap household points onto the smallest polygon layer and to find point's household-number in each polygon. Overlapping image is the same as shown in Figure 8, and Figure 10 is the result of the overlapping - another new layer in GIS.

Table 2 Building Amount in Polygons

Polygon ID	Planning Zone ID	Census Zone ID	Building Amount
1	I	J	N1
2	I	J	N2
....	...	...	...
N	M	L	N3

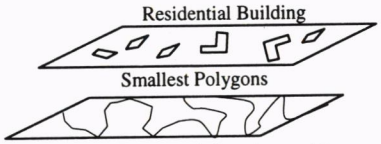

Figure 8 Image of Overlapped Layers

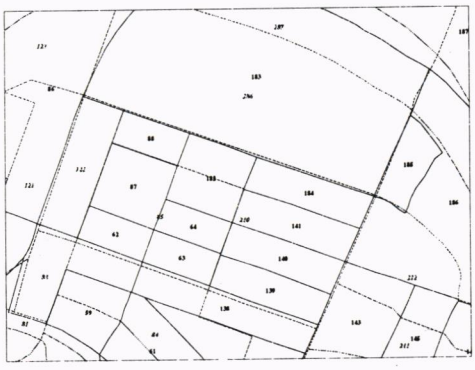
Figure 9 Layer of the Smallest Polygon

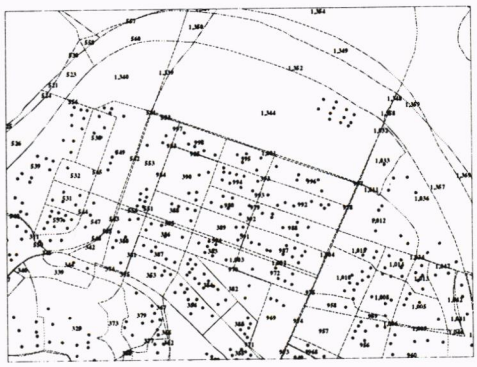
Figure 10 Smallest Polygon with Households Points

Its geographic data are smallest polygons with their household points, and the text information, which can also be calculated and input to a table such as Table 3, is the relationship between the smallest polygons and the households numbers.

Table 3 Households Number in Smallest Polygons

Polygon ID	Household No.	Planning Zone ID	Census Zone ID	Building Amount
1	1	I	J	N11
1	2	I	J	N21
....	...	...	...	...
N1	K1	M1	L1	N31

In previous Chapter we talked about establishment of the households-point layer, here we add some explanation of the technique of projecting the points into the layer. In fact, not all of the households are projected to the layer as point's object. If a census zone is not divided into several segments, it means that the census zone is same or is included in a smallest polygon. In this census zone, households in the zone will be totally included in a smallest polygon, just pass the households and population to the relevant polygon is enough. Therefore, the projected points are such points that their census zones consist of several smallest polygons. Trough treating Table 1, we can tell which census zone is divided into several segments, the method is to count the column of Census Zone ID. If a Census Zone ID appears just one time, it means the zone is not divided by urban planning zone, otherwise the zones are divided as many times as its ID appears.

4.4 Calculating Small Zonal Households/Population from Text Data

Now besides the text data bought from some agent such as census zonal population, household's list, we got some other text data in Tables 1, 2 and 3 by overlapping some layer in GIS. Since just the households in the divided census zones are projected to a layer as point objects, we have to type a table to indicate the corresponding households data for the questioned households as Table 4 based on the household's list for later use.

Table 4 Household's List of the divided Census Zones

Census Zone ID	Households No.	Male Member	Female Member	Remark
1	1	2	2	
1	2	1	3	
...	...	...	...	...
N	1	3	2	
N	2	2	4	

Our next work is to design a procedure to calculate the households and population in the smallest polygons. It is divided into four steps, each step links two tables to each other by a relevant keys. In order to make a good visual image, a button named "Population Calculator" is installed onto the "Main Form" as shown in 6. When the button is clicked, the module will work to finish calculating each smallest polygon's male and female population and households. The button function supported by the "SQL (Standard Query Language)" tool is described below.

First, using Census Zone ID as the linking key, we can add households and population in Census Zones to Table 1 as parent's households and population columns to get Table 5. Meaning time, building's amount in Table 2 is also input into Table 4.

Table 5 Calculated Results for Smallest Zone's Population

Polygon ID	Land-use Zone ID	Census Zone ID	Buildings	Parent's Households	Parent's Population	Households	Population	Error
1	I	J						
2	I	J						
....	...	...						
N	M	L						

Second, using both Census Zone ID and Households Number as linking keys, we add male/female members in Table 4 to Table 3 to get Table 6. Then counting Households-No. and summing up Male/Female Members in the table by grouping Polygon ID, we can get relevant households and population for each smallest polygon (note: the smallest polygons here are those related to the divided census zones)

Table 6 Male/Female Members for each Household with the Smallest Polygon ID

Polygon ID	Households No.	Male Member	Female Member	Remark
1	1	2	2	
1	2	1	3	
....	...	...	...	...
N1	3	3	2	
N1	5	2	4	

Third, using Smallest Polygon ID as the linking key, we will input the population and households gotten in step 2 into table 5. Till now, households and populations in the polygons that are relevant to the divided census zones are calculated. Then we will treat the left smallest polygons, it is very simple, what we have to do is just transferring the data in parent's households and population columns in each record to the households and population columns.

Forth, we will do several checks to find the possible errors in the original data or working process. This error information will be input to "Error Columns" in Table 5. One check is to see if there are households in a polygon but there is not residential building, or there are some residential buildings but there is no household. Another check is to see if the average parent households/population equals to sum up households/population or not when grouped by Census ID. These error-messages can help us to adjust the data to the most updated one.

5. SOME DISCUSSIONS

As we mentioned, census table can always be bought. However, the household's draft or list is not available for some situations. If then, we have to find some substituted data. For example, one of the best substitutions is the residential house map published by some surveying companies in Japan. In this case, households can be numbered, but the household's members are not available. As the result, we can just calculate each smallest polygon's households rather than households and population. In order to get the population, we suggest to distribute population in each census zone to the polygons based on the rates between the households in the polygons and census zone.

This developed tool can used not only for land-use and transportation integrated model, but

also for some general urban planning such as master planning. Using this tool, basic urban zoning is freely, planner can zoning arbitrarily according to their needs. Responding to each census, using this tool the households/population in smallest polygons can be calculated easily. The picture box in the "Main Form" can be used for not only loading layers here mentioned, but also loading any layers such as road-network, land-use pattern etc. Then it is very easy for us to use the platform as a monitor to get suitable visual imagines of the integrated land-use and transportation model.

6. CONCLOSIONS

Integrating MapInfo (a software of GIS) with MS-Access (a database software) through MS-VB, we developed a tool that can calculate small zonal population for land-use and transportation planning. Because form in VB is used as the working platform, this tool has both good visual view and excellent function. On the other side, using "Access Table" can design or output very pretty report for some other purposes. Just with MapInfo this kind of reports can not be accomplished.

At present, in order to run this tool, some original data such as census map, census tables, household draft and households list are needed. Because most of them can be bought with the digital-type data, it is very easy for us to change them into database table for afterward use. On the other hand, creating geographic layers cost lot of time, however, these layers can be also used after population calculation if GIS tends to be applied to an urban model.

The picture box in the platform can also be used for load some other maps such as urban road network, land using pattern and so on. Because we set almost all of the MapInfo functions onto the platform's manus-items or icons, we can demonstrate transportation, land use and environmental items map very easily.

REFERENCE

Yukio SADAHIRO. (1999) Errors in Areal Weighting Interpolation between Incompatible Zonal Systems. **Theory and Application of GIS Vol. 7 No. 1**, 1-9.

Tschangho John Kim, Changho Park and Seung-Kwan Lee (1998). AN INTEGRATED LAND USE-TRANSPORTATION MODEL WITH GIS : APPLICATION TO THE CITY OF SEOUL, **8TH WCTR ANTWARP JULY.**

J.De D. Ortuzar and L.G. Willumsen (1990). **MODELLING TRANSPORTATION**, JOHN WILEY & SONS

Microsoft Corporation(1995). **MICROSOFT VISUAL BASIC, VERSION 4.0 PROGRAMMING SYSTEM FOR WINWINDOWS**

MapInfo Corporation (1995). **MAPBASIC USER'S GUIDE.**

Japan's National Statistic Bureau (1996), **REPORT OF POPULATION CENSUS FOR 1995**.