Online Micro-spatial Analysis Based on GIS Estimated Building Population: A Case of Tsukuba City

January 2010

Ko Ko LWIN

Online Micro-spatial Analysis Based on GIS Estimated Building Population: A Case of Tsukuba City

A Dissertation Submitted to
the Graduate School of Life and Environmental Sciences,
the University of Tsukuba
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Science
(Doctoral Program in Geoenvironmental Sciences)

Ko Ko LWIN

Abstract

Population data used in GIS analyses are generally assumed to be homogeneous and planar (i.e. census tracts, townships, or prefectures) due to the public unavailability of building population data. Moreover, under the GIS domain, spatial analysis functions performed within the census tract do not acquire any significant changes in population. However, information on building population is required for micro-spatial analysis for improved disaster management and emergency preparedness, public facility management for urban planning, consumer and retail market analysis, environment and public health programs, and other demographic studies.

Several methods have been developed to generate smaller geographical units of population distribution based on aggregated values with ancillary datasets, commonly known as "Dasymetric Mapping". From a cartographic point of view, common cartographic forms of population mapping are the choropleth map and the dasymetric map. Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area. However, they have limited utility for detailed spatial analysis of population data, especially where the population is concentrated in a relatively small number of villages, towns and cities. Moreover, choropleth maps cannot express statistical variation within the administrative areal units, such as changing population density.

The main objective of this study is to develop online micro-spatial analytical functions based on GIS estimated building population to improve the population data analysis at micro-scale level, as a case of Tsukuba City central area. This research develops the method and application for estimation of population at building level using modern Remote Sensing data acquisition system and GIS technologies such as LIDAR

(Light Detection and Ranging) data and building footprints to improve the accuracy in micro-spatial analysis as a case of Tsukuba City central area. In this research, LIDAR derived Digital Volume Model (DVM) and building footprints data were used to estimate the building population and iTownpage from Nippon Telegraph & Telephone Corp. (NTT) data was utilized for separation of residential and non-residential buildings.

Due to spatial data availability at finer scale and increasing demands on GIS analyses at micro-scale level, this research also develops and demonstrates the online interactive micro-spatial analysis (Web GIS) based on estimated building population with other datasets such as facility and road network by providing common map tools (i.e. Line, Circle, Polygon, etc.) to demonstrate the population data analysis at micro-scale level for local residents, business owners, urban planners and other spatial information users in timely manners. Moreover, this research proposed several indices and measurement parameters such as Weighted Population Mean Center (W-PMC), Weighted Facility Mean Center (W-FMC), Connectivity Mean Center (CMC), Population Index (PIX), Facility Index (FIX) and Connectivity Index (CIX) to provide abstract idea to make spatial decision process in urban area. These measurement parameters and indices can be used for urban planners, in the case of sustainable urban development planning and for potential home buyers or local residents to evaluate their neighborhood conditions.

Keywords: Online micro-spatial analysis, Digital Volume Model (DVM), LIDAR, Building Population, Dasymetric Mapping, Tsukuba City central area

Contents

Chapter 5 Conclusion (108-113)

List of Figures

List of Tables

List of Abbreviations

AGI	Association for Geographic Information
AJAX	Asynchronous JavaScript and XML
ASP	Active Server Pages
CEDS	Cadastral-based Expert Dasymetric System
CMC	Connectivity Mean Center
CIX	Connectivity Index
CSV	Comma Separated Value
DHM	Digital Height Model
DMSP	Defense Meteorological Satellites Program
DSM	Digital Surface Model
DTM	Digital Terrain Model
DVM	Digital Volume Model
ESRI	Environmental Systems Research Institute
FM	Frequency Modulation
FIX	Facility Index
FTP	File Transfer Protocol
GI Science	Geographical Information Science
GIS	Geographical Information System
GPS	Global Positioning System
GSI	Geographical Survey Institute (Japan)
GUI	Graphical User Interface
ICA	International Cartographic Association
IMU	Inertial Measurement Unit

LIDAR	Light Detection and Ranging
MicroSPA	Micro-spatial Analysis
NGDC	National Geospatial Data Clearinghouse (US)
NTT	Nippon Telegraph & Telephone Corp.
OGC	Open Geospatial Consortium
PIX	Population Index
RMSE	Root Mean Square Error
TBP	Total Building Population
TIN	Triangulated Irregular Network
TIN-DSM	Triangulated Irregular Network Digital Surface Model
TIN-DTM	Triangulated Irregular Network Digital Terrain Model
W-FMC	Weighted Facility Mean Center
W-PMC	Weighted Population Mean Center
XML	Extensible Markup Language

Chapter 1

Introduction

Research into micro-spatial analysis has increased due to remote sensing data available at finer spatial resolution with more diverse geo-information sources (IKONOS, QuickBird, LIDAR, etc.) and the availability of fine-scale GIS data with enhanced attribute information (e.g. building footprints with the number of floors, building use type and building name) (Lwin and Murayama, 2009). In view of the advances in modern **geospatial** information technologies, this is a good time for studying the world at a micro-scale level. Moreover, recent years have seen a rapid growth in interest in the addition of a spatial perspective to population research, and in part this growth has been driven by the ready availability of **georeferenced** data, and the tools to analyze and visualize them: geographic information systems (GIS), spatial analysis, and spatial statistics. The term "geographic information science" has emerged as something of an umbrella in this arena, implying both the use of GIS and other spatial tools for scientific research, and the study of the fundamental principles and issues underlying a spatial perspective (Longley *et al.*, 2005).

Population count is a key anchor for much of the official statistical system and the benchmark for many commercial and research surveys and analyses (Cook, 2004). GIS plays a critical role in population studies and analyses by means of mapping the spatial extent and analyzing it along with other GIS datasets. The benefits of geographic data automation in statistics are shared by the users of census and survey data. The data integration functions provided by GIS, which allow linking of information from many different subject areas, have led to much wider use of statistical information. This, in turn, has increased the pressure on statistics agencies to produce high-quality spatially referenced information for small geographic units. The types of applications for such data are numerous. Examples include planning of social and educational services, poverty analysis, utility service planning, labor force analysis, marketing analysis, voting district delineation, emergency planning, epidemiological analysis, floodplain modeling and agricultural planning.

Estimating and mapping the population is not an easy task due to the nature of human activities which change over space and time. Normally, population can be estimated using statistical and spatial (Remote Sensing and GIS) approaches. For example, night-time city lights imagery has been shown to demonstrate a reasonable correlation with population (Sutton, 2001). Developments in computer hardware and mapping software have already encouraged many statistical and census offices to move from traditional cartographic methods to digital mapping and geographic information systems (GIS) (Rhind, 1991; Ben-Moshe, 1997).

On the other hand, the emergence of user friendly Web GIS technologies such as Google Map/Earth and Microsoft Virtual Earth empower and **mashup** the internet users to make interactive spatial decisions in a timely manner. Nowadays, Web GIS is part of our daily lives. For example, finding a restaurant, browsing our neighborhoods, or choosing a place to live are important considerations for potential home buyers and can be achieved through the internet. Retail shop owners are ever seeking populated areas inside the city. Urban planners are always watching for population growth and their spatial distribution patterns in order to maintain and improve the urban system.

1.1 Living with Error

There are several error issues in GIS such as cartographic error, positioning error, attribute error, statistical error, logical error, and so on. Openshaw (1989) identified the following sources of error in GIS: errors in the positioning of objects; errors in the attributes associated with objects; and errors in modeling spatial variation (e.g. by assuming spatial homogeneity between objects). Population data exhibit spatial variation, especially in areas with a mix of high- and low-rise buildings such as urban areas, residential areas patched with unpopulated spaces (paddy fields, parks, playgrounds or government institutions) as in rural areas. Moreover, most census boundaries do not coincide with the boundaries of geographic features such as land use/land cover, soil type, geological units, and floodplain and watershed boundaries; this is known as "spatial incongruity".

On the other hand, spatial analysis functions performed within the census tract do not acquire any significant changes in population (Lwin and Murayama, 2009). This creates errors when trying to establish accurate rates for GIS analyses pertaining to health studies, crime patterns, hazard/risk assessment, land-use planning, or environmental impacts, among others, which rely on a smaller unit of analysis than the original zones. Examples of this are impact buffers that intersect the census enumeration unit, or a different set of zones altogether that do not coincide with the original set (e.g. overlaying data from units with non-coincident boundaries and/or overlapping spatial units such as census tracts and police precincts or health districts)

(Maantay *et al.*, 2007). This research aims to reduce the logical error (i.e. spatial variation) in population data analysis at micro-scale level by utilizing modern spatial information technologies.

1.2 GIS and Population Studies

As a GIS for population and environmental studies aspect, demographic researchers have also explored population and environment linkages (Pebley, 1998). Remote sensing, GIS, and spatial econometrics have already been used effectively to analyze the relationship between human activities and local environmental change, in particular in the area of deforestation and changing patterns of land use (Chomitz and Gray, 1996; Moran and Brondizio, 1998; Nelson and Hellerstein, 1997; Wood and Skole, 1998). Therefore, it is not surprising that public government agencies as well as private companies spend millions of dollars each year obtaining aerial photograph and other forms of remotely sensed data (Jensen and Cowen, 1999). Empirical research on the reciprocal relations between population dynamics and the natural environment at the local level have been quite rare but as shown in the recent edited collection by Fox *et al.* (2003) the research environment is changing fast as population scientists begin to integrate GIS, remote sensing and spatial analysis methods (Liverman, 1998).

As a GIS for population and Social Science studies aspect, in the areas of labor market research and explorations of the spatial mismatch hypothesis (Mouw, 2000), health inequality (Duncan *et al.*, 1993; LeClere *et al.*, 1998; Roberts, 1998; Yen and Kaplan, 1999; Browning *et al.*, 2003) and crime analysis (Rossmo, 1995; Morenoff and Sampson, 1997; Bowers and Hirchfield, 1999; Morenoff *et al.*, 2001; Baller *et al.*, 2001).

The capacity to gather and organize spatial data on demographic and health events on individuals, families, households, neighborhoods, health facilities, routes and networks, as well as a host of environmental phenomena will continue to grow dramatically during the first decade of the twenty-first century. It is inevitable that geographic information systems and related technologies will be increasingly employed to explore possibilities to integrate and analyze such data (Liverman *et al*., 1998; Fox *et al*., 2003).

However, population data used in GIS analyses is generally at the level of census tracts, townships or prefectures, since building population data is not available for public use due to privacy concerns. For spatial information users, population data has generally only been available in township polygons or city point features with aggregated population data. For non-spatial information users, population data (text and tables) can be obtained from the National Census Bureau or local government offices. All of these datasets are suitable for local and regional analysis, but not for micro-spatial analysis and decision-making processes.

1.3 GIS and Population Mapping

Many early cartographic endeavors were concerned predominantly with producing maps intended for navigational and exploration purposes; these required furthering our abilities to observe and measure the physical world with increasing levels of precision (Hall, 1993). Technical advancements in instrument design and geometric theory made these more precise maps possible, and they generally portrayed tangible aspects of the physical world, such as areal sizes of geographic units, topography, temperatures, and sea depths (Dorling and Fairbairn, 1997). Maps depicting social, cultural, or economic aspects of the world are termed thematic maps—those showing a particular “theme,” such as poverty levels, disease rates, or the flow of migration. Thematic maps (also called statistical maps, if depicting a quantitative data theme) are generally of more recent vintage (Dent, 1993).

From a cartographic point of view, common cartographic forms of population mapping are the **choropleth map** and the **dasymetric map**. Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area. However, they have limited utility for detailed spatial analysis of population data, especially where the population is concentrated in a relatively small number of villages, towns and cities. Moreover, choropleth maps cannot express statistical variation within the administrative areal units, such as changing population density. One way to avoid this limitation is by transforming the administrative units into smaller and more relevant map units through a process known as dasymetric mapping (Bielecka, 2005). Dasymetric maps use ancillary information to help delineate new zones that better reflect the changing patterns over space. Recent research suggests that dasymetric mapping can provide small-area population estimates that are more accurate than many areal **interpolation** techniques that do not use ancillary data (Mrozinski and Cromley, 1999). Many methods to map population distribution have been practiced in geographic information systems (GIS) and remote sensing fields. Many cartographers prefer dasymetric mapping to map population because of its ability to more accurately distribute data over geographic space. Similar to "choropleth maps", a dasymetric map utilizes standardized data (for example, census data). However, rather than using arbitrary enumeration zones to symbolize population distribution, a dasymetric approach introduces ancillary information to redistribute the standardized data into zones relative to land use/land cover (LULC), taking into consideration actual changing densities within the boundaries of the enumeration unit (Langford and Unwin, 1994)

1.4 GIS and Online Micro-spatial Analysis

The increasing popularity of the Internet, from online surfing to e-commerce, has made the Internet an integral part of society. The ubiquitous access to the Internet and user-friendliness of the World Wide Web (WWW) have made them a powerful means for people to exchange and process information, and to make transactions. The Internet has revolutionized ways of doing things in journalism, the sciences, publishing, and many other fields (Plewe, 1997). It is also shaping the ways in which traditional GIS function. The Internet is affecting GIS in three major areas: GIS data access; spatial information dissemination; and GIS processing. The Internet provides GIS users easy access to GIS data from different sources of data providers. GIS data warehouses and clearinghouses, and digital libraries are two common forms of Internet data access systems. The US National Geospatial Data Clearinghouse (NGDC) under the Federal Geographic Data Committee in the USA has been working to build a distributed archive of information for universal access (Peng and Nebert, 1997). The aims of the Alexandria Digital Library project is to construct a centralized or indexed repository of spatial information from diverse collections in order to make it available to the public over the Internet (Chen *et al.*, 1997; Frew *et al.*, 1995; Goodchild and Proctor, 1997). The Internet also enables the dissemination of GIS analysis results and spatial information to a much wider audience than is possible with traditional GIS. The general public can now directly access spatial information and see spatial patterns from their web browsers.

They can even conduct search and query operations for spatial objects from their home or workplace.

Furthermore, the Internet is becoming a means to conduct GIS processing. It enhances the **accessibility** and shareability of GIS analysis tools and functionalities over the Internet. GIS users can work with GIS data interactively on the web browser without owning GIS software on their local machines. The easy access to Internet communications software such as Microsoft NetMeeting allows users to emulate a map server by sharing Windows-based GIS applications with others over the Internet in order to conduct collaborative GIS analysis (Matuschak, 1996). The issue of GIS data access and transmission on the Internet has been addressed in the literature (Coleman and McLaughlin, 1997; Peng and Nebert, 1997). When using the Internet to access and transmit GIS data it is assumed that the user will use the data in his or her local machine with stand-alone GIS software installed. This is useful because it allows GIS users to obtain data more efficiently. However, use of the Internet in this manner is very limited. GIS users have to use traditional GIS software to view and analyze the data. Access to and transfer of GIS data over the Internet is the first step toward interoperability in the GIS community. The ability to access GIS analysis functions and to conduct GIS analysis anywhere over the Internet is the next and yet more important step. GIS processing is already being conducted on the Internet and is evolving rapidly, but it has not been adequately addressed in the literature.

Micro-spatial analysis is such kinds of techniques that are provided within modern Geographic Information Systems (GIS) by utilizing finer spatial resolution or smaller map scale. Nowadays, the term "Micro-spatial Analysis" is frequently appear in Social Science in terms of studying about crime, race and retail market analysis. According to Okabe and Okunuki (2001), described that potential demand of analytical tools for micro-spatial analysis in market analysis is increasing because of marketing is concerned with micro-scale geographical factors with a street network. Such marketing is called micro-marketing. GIS provide a powerful tool for micro-marketing, but a problem with that has poor analytical tools. The most traditional analytical tools are based upon the assumption that the market areas are homogeneous plane, and distance is measured in terms of the Euclidean distance. In a small area, however, irregular streets produce a heterogeneous plane and consumers can access to stores through a street network. This suggests that there be great potential demand for analytical tools for micro-spatial analysis on a network where distance is measured in terms of the shortest-path distance.

On the other hand, a wide variety of web-based or web-deployed tools have become available, enabling datasets to be analyzed and mapped, including dynamic interaction and drill-down capabilities, without the need for local GIS software installation. There are several advantages to construct Web GIS such as reduce maintenance cost, accessible to all levels of spatial information users and updatable and reusable.

1.5 Background Studies

Several methods have been developed to generate smaller geographical units of population distribution based on aggregated values with ancillary datasets, commonly known as "Dasymetric Mapping" by using **GI Science** theory and practice. The population data segregation method was firstly utilized and termed dasymetric by the Russian cartographer Tian-Shansky, who developed the multi-sheet population density map of European Russia, scale 1:420 000, published in the 1920s (Preobrazenski, 1954).

As populated territory, Tian-Shansky mapped areas within the equidistant of one verst (1067 m) from built-up terrains. The first cartographer who popularized dasymetric mapping was Wright (1936). He set forth a new method of presenting population density based upon the division of a given administrative unit into smaller areas complying with different types of geographical environments. Wright (1936) derived population density values subjectively, which is considered as the main disadvantage of the dasymetric method.

Although dasymetric mapping has been in use since at least the early 1800s, it has never achieved the ubiquity of other types of thematic mapping, and thus the means of producing dasymetric maps have never been standardized and codified in the way other types of thematic mapping techniques have been (Eicher and Brewer, 2001; Slocum, 1999). Therefore, dasymetric methods remain highly subjective, with inconsistent criteria. The reason for this relative lack of popularity and the paucity of standard methodology surely lies at least partially in the difficulty inherent in constructing dasymetric maps, and until recently, the difficulties in obtaining the necessary data, as well as access to the computer power required to generate them (Maantay *et al*., 2007). Transferring data from one set of geographic zones or districts to another set of non-coincident zones is often necessary in spatial analysis. For instance, we might have data on the number of people living within a certain census tract but need to estimate the number of people in a smaller area within the tract, or an area that includes only part of that tract and part of other tracts. We may be interested in population or other data at a watershed level and only have population data available at the census enumeration units.

Maantay *et al*. (2007) extensively reviews on existing dasymetric methods and techniques. The following are some developed methods and approaches in dasymetric mapping: areal **interpolation**, filtered areal weighting (binary method) (Eicher and Brewer, 2001), filtering with land use/land cover data (Sleeter, 2004), and cadastral-based expert dasymetric system (CEDS) (Maantay *et al*., 2007).

Areal Interpolation: A common method for calculating disaggregated population values is areal interpolation. This is defined as "the transfer of data from one set (source units) to a second set (target units) of overlapping, non-hierarchical, areal units (Langford *et al*., 1991). Areal interpolation is closely related to dasymetric mapping of population densities (Holt *et al*., 2004). The main difference between areal interpolation and dasymetric mapping is that with the later approach, the data are not re-aggregated into a desired enumeration unit as they are with areal interpolation (Eicher and Brewer, 2001). The simplest type of areal interpolation is area weighted interpolation, which requires no information besides the geography of both sets of zone units and the counts to be interpolated from the source to the target zones (Goodchild and Lam, 1980). In area weighted interpolation the two incompatible zone systems describing a given region are superimposed and intersected, creating a set of intersection zones, each of which describes a unique pair of one source and one target zone (Flowerdew and Green, 1992; Reibel and Bufalino, 2005). Each intersection zone is assigned a fraction of its respective source zone's count corresponding to the proportion of the source zone's area occupied by the intersection zone. The intersection zone counts can then be summed across their respective target zones to complete the integration of data to the incompatible zone system.

Filtered Areal Weighting (Binary Method): By excluding uninhabitable areas in land-use or land-cover data and redistributing the population in the remaining areas is simply known as "binary" method (Eicher and Brewer, 2001).

Land Use/Land Cover as Ancillary Data: Although land-use polygon (**vector**) data sets have occasionally been used as the filtering layer (e.g., Bielecka, 2005; Poulsen and Kennedy, 2004), most dasymetric studies have used satellite data (**raster** data format) to determine the locations of uninhabited areas, and/or to classify inhabited areas by population density (Holloway *et al*., 1999; Langford and Unwin, 1994; Mennis, 2003; Sleeter, 2004). In highly urbanized areas, land-cover data derived from satellites may not yield precise enough results to get a true picture of population density, due to limitations in available **pixel resolution** and intra-pixel heterogeneity of urban areas (Forster, 1985).

Cadastral-based Expert Dasymetric System (CEDS): Maantay *et al*. (2007) used cadastral data as the ancillary data appears to be an innovative and progressive approach to dasymetric mapping. Cadastral data are used in recording property boundaries, property ownership, property valuation, and, of course, for the all-important purpose of property tax collection. The type of cadastral-level data used in her CEDS method is commonly available for most urbanized areas in the United States, western Europe, and other more developed areas. The data are usually organized by township, municipality, or county, and, less often, by metropolitan region. However, in many parts of the world, census and cadastral data may not be readily available, current, or accurate.

Baudot (2001) makes the point that for urban areas in less-developed countries, very often there are no census property tax records or city planning data on population to work with, and even when such data are collected, the exponential growth rates of these cities makes the census data obsolete almost immediately. This is why satellite data are most often used for dasymetric mapping—they are available for almost all parts of the world and are very current. In urban areas where census and cadastral data are available, the CEDS method will be an improvement. For instance, municipalities where property tax records are linked to a digital spatial database (e.g., most cities and larger towns in the United States), the cadastral data required by the CEDS method will be available. Although these data may not be available to the general public for free, they still tend to be less expensive for the end user than high **resolution** remotely sensed images for the equivalent spatial extent.

Moreover, the utility of remote sensing for population estimation has been continuously explored since the 1950s. Various types of satellite imagery have been examined to study population distribution, including Landsat Multispectral Scanner (MSS)(Iisaka and Hegedus, 1982), Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM) (Forster, 1985; Li and Weng, 2005), SPOT (Lo, 1995), and DMSP nighttime imagery (Dobson *et al*., 2000; Sutton *et al*., 2001). Due to emergence of high resolution satellite imagery and smaller map scale with better attribute information, GIS research issues on micro-spatial analysis is increasing. We need to develop more theory, dataset and analytical functions for micro-spatial analysis. Moreover, population analysis at micro-scale level is very rare or absent in GIS arena due to unavailability of population data at smaller unit.

There are several Web-GIS systems have been developed in last decades. For example, modern information technologies are present in various literatures and widely used in studies related to population, environmental resource management and planning. Murayama (2000) introduced Web-based GIS for Malaysia population data analysis. Boston and Stockwell (1995) described Australia's World Wide Web (WWW) service called ERIN (Environmental Resources Information Network), which provides key Australian environmental information. They concluded that the WWW is an effective technology for a data center to provide information of various kinds, including attributive, geographic data, and modeling results. Loh *et al*. (1995) described the importance of information management and exchange to sustainable development using WWW information technology. They also developed an information system which provides users with easy access to data. More recently, Chang *et al.* (1998) carried out a preliminary configuration for a solid waste management system framework in Taiwan. They suggested an approach integrating current hardware and software and combining information systems developed for each type of wastes. This study also proposed a geographic data inquiry function and a client-server environment to improve waste management efficiency.

Chang *et al*. (2001) developed a web-based information system for scrap vehicle recycle management. They used Internet as an information sharing mechanism to enhance operational efficiency within the scrap vehicle recycling program, where the participants in the recycling channel did not have easy and clear communications in the past. To reach the system goal, various information technologies were used including web **DBMS**, Web GIS, and web computing. Chang and Chang (2002) designed a web-based decision support system for sustainable river aquatic environment management in urban areas. The web system integrates the joint research effort, using data warehousing and web computing techniques, allows decision makers to perform remotely the complex data query and modeling analysis. A more sustainable management scenario can be devised effectively using such web system. These various information systems bring modern information technology to environmental resource management fields and enhance managerial performance.

1.6 Research Objective

The main objective of this study is to develop online micro-spatial population analysis based on GIS estimated building population to improve the population data analysis at micro-scale level by utilizing Remote Sensing and GIS technologies. This research intended to introduce population estimation at building level by utilizing modern remote sensing data acquisition system such as LIDAR data and building footprints in conjunction with detailed tabular iTownpage information from NTT. The aim of this population estimation is to reduce the logical errors in population data analysis at micro-scale level by introducing online interactive micro-spatial analytical functions as a case of Tsukuba central area. This research also explores the future potential applications of LIDAR data for sustainable urban development planning.

1.7 Research Structure

In order to achieve this goal, this research divided into threefold or three modules (Figure 1-1). First, develop an algorithm for estimation of building population and its accuracy assessment. Second, apply this algorithm to estimate Tsukuba central area building population by utilizing modern spatial data acquisition system (i.e. LIDAR). Finally, develop online micro-spatial analytical functions based on estimated Tsukuba central area building population.

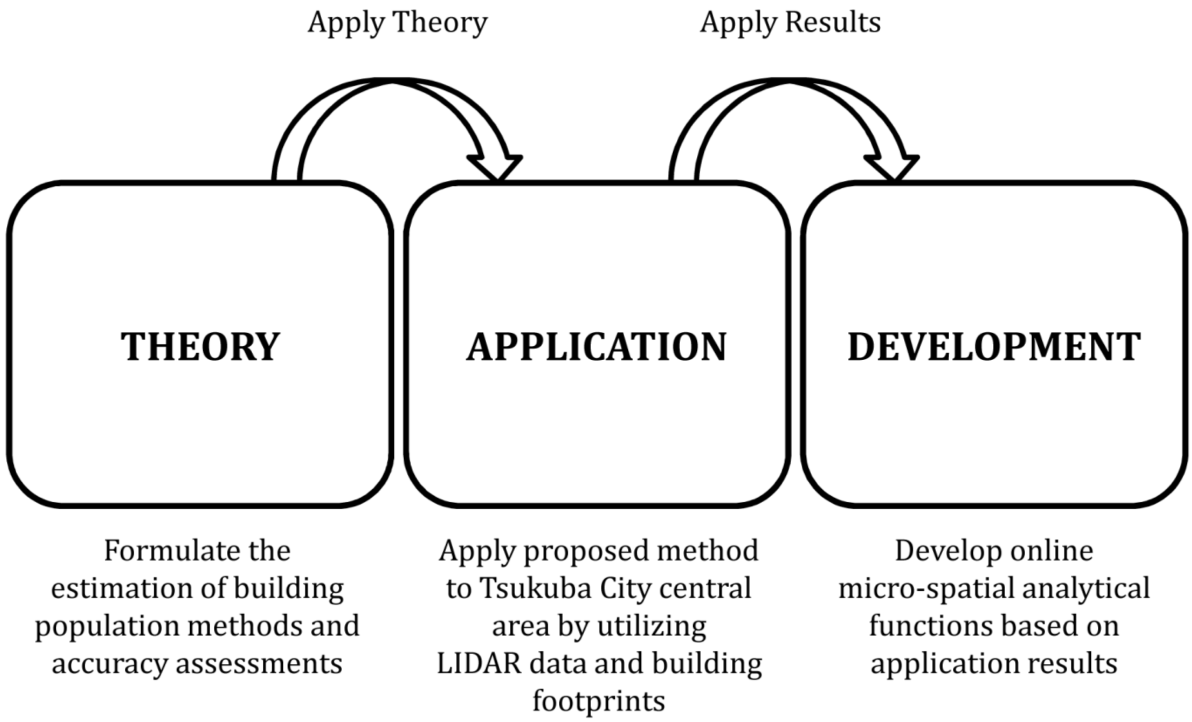


Figure 1-1: Research structure

Chapter 2

Methodology

2.1 Applied Method

This research introduces two estimation methods: (1) Areametric (which does not require information on the number of building floors); and (2) Volumetric (which does require information on the number of floors). For improved the accuracy, this two methods allow filtering by other categories into the computation, such as filtering by minimum footprint area and building use types, e.g. commercial, industrial, educational, and other building use types that are not occupied by residents. The calculation is demonstrated by the following mathematical expressions and graphical illustration in Figure 2-1.

Areametric Method:

$$BP_i = \left(\frac{CP}{\sum_{\mathrm{k}=1}^{\mathrm{n}} BA_k}\right) BA_i \qquad \text{Using building footprint surface area}\ldots\ldots \text{ (2-1)}$$

Volumetric Method

$$BP_i = \left(\frac{CP}{\sum_{\mathrm{k}=1}^{\mathrm{n}} BA_k.BF_k}\right) BA_i.BF_i \quad \text{Using number of floors information}\ldots\ldots \text{ (2-2)}$$

Moreover, advances in remote sensing data acquisition technologies such as LIDAR can be used for extraction of building footprints, building height (Digital Height Model (DHM)) and building volume (Digital Volume Model (DVM)). Equations (2-3) and (2-4) can be used for LIDAR data:

$$BP_i = \left(\frac{CP}{\sum_{k=1}^{n} BA_k . BH_k}\right) BA_i . BH_i \qquad \text{Using average building height} \ldots\ldots \text{(2-3)}$$

$$BP_i = \left(\frac{CP}{\sum_{k=1}^{n} BV_k}\right) BV_i \qquad \text{Using total building volume} \ldots\ldots\ldots \text{(2-4)}$$

Where:

BPi	Population of building i
CP	Census tract population
BAi	Footprint area of building i
BFi	Number of floors of building i
BHi	Average height of building i (from LIDAR data)
BVi	Total volume of building i (from LIDAR data)
i, k	Summation indices
n	Number of buildings that meet user-defined criteria and fall inside the CP polygon

The Areametric method (Equation 2-1) is suitable for low-rise buildings especially in rural areas while the Volumetric method (Equation 2-2 through 2-4) is suitable for high-rise buildings, especially in downtown areas.

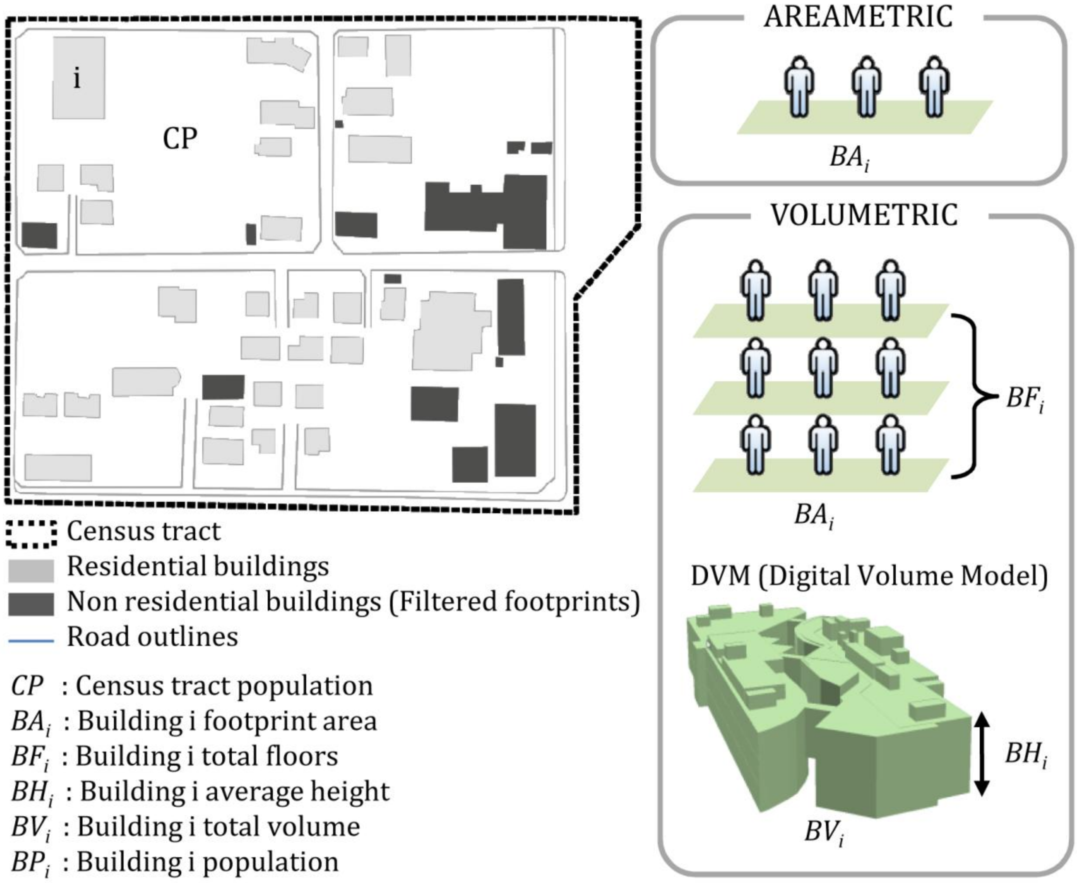


Figure 2-1: Graphical illustration of equations

2.2 Test Data

The two methods were evaluated using actual building population data acquired from the city (X) administration office for study purposes. These data include detailed information about each building such as building age, construction material, building type (detached, non-detached, semi-detached, flat or apartment), building use type (residential, commercial or educational, etc.), number of floors, number of households and total number of people, which is intended for use in disaster management. The test data were converted into the ESRI Shapefile format and after conversion the building footprint polygons totaled 9,913, the census tracts totaled 28 and the total population was 28,000.

2.3 Test Method

In order to find the best results based on data availability, apply both methods (Areametric and Volumetric) with various filtering areas such as 0, 5, 10, 15, 20, 25, 30, 35, and 40 m^2 of residential building use type.

2.4 Implementation of GIS Tool

This research also implemented a standalone GIS tool (Figure 2-3) using the Visual Basic programming language and TatukGIS DK (Development Kit). Users can define the minimum ignored footprint size such as for porticos, garbage boxes and other unpopulated areas. They can also apply filtering by attribute field(s) such as building use type and other attribute information. Three additional approaches are available under the Volumetric method, namely Use Number of Floors, Use Average Building

Height and Use Total Building Volume. Figure 2-2 shows the program flow of PopShape GIS tool.

The operational steps employed (with numbers corresponding to those shown in Figure 2-3) were as follows: (1) Open census tracts file (Shape polygon); (2) open building footprints file (Shape **polygon**); (3) filter by footprint size; (4) filter by building use type(s); (5) select method (Areametric or Volumetric); (6) select approach (use Number of Floors or Building Height or Building Volume); (7) select appropriate field (Floor or Height or Volume attribute field); (8) assign output file name; and (9) start to process (see http://giswin.geo.tsukuba.ac.jp/sis/en/software.html for additional details).

After processed, the estimated building population attribute field, "EST_POP", will be appeared in a new ESRI Shapefile. A map viewer is also provided for viewing the processed results by performing common GIS functions such as add map layer, zoom in, zoom out, get attribute information, label by attribute field and change map layer properties.

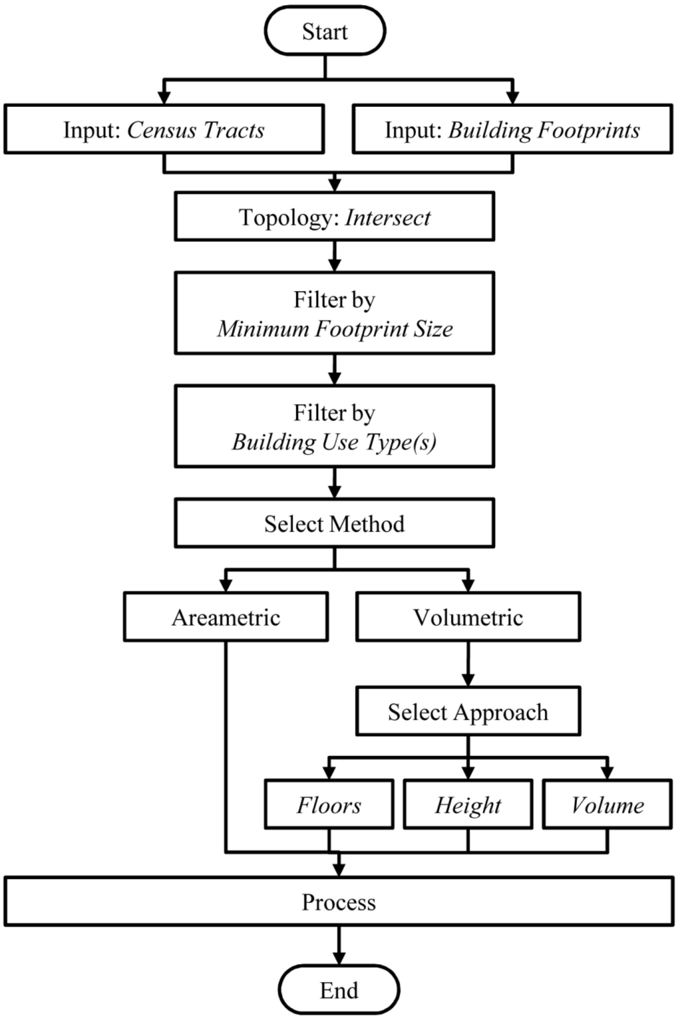


Figure 2-2: PopShapeGIS program flowchart

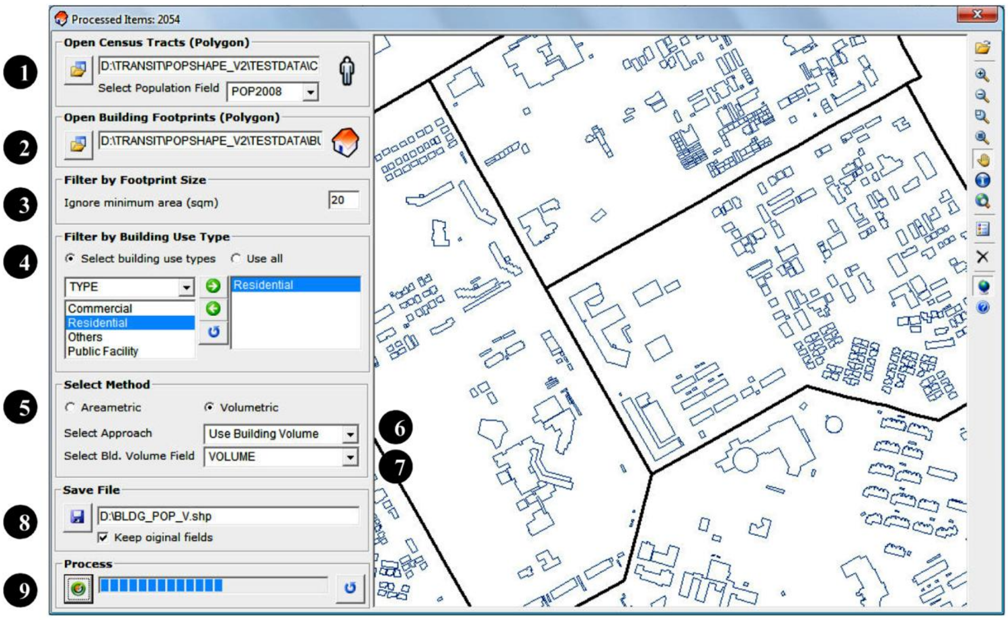


Figure 2-3: PopShapeGIS program Graphical User Interface (GUI)

2.5 Test Results and Accuracy Assessment

All estimated values were evaluated using actual building population data by means of visual, statistical and spatial approaches. The best results were obtained by using the Volumetric method filtered at the 25 m^2 footprint category.

2.5.1 Visual Assessment

Visual assessment is simply compares the two attribute values (i.e. actual building population and estimated building population) by visually (Figure 2-4 and Figure 2-5). Visual assessment is one of the accuracy assessments in Remote Sensing data processing. This visual assessment can be done in under PopShape GIS tool by labeling attribute field, changing color for label text and other graphics.

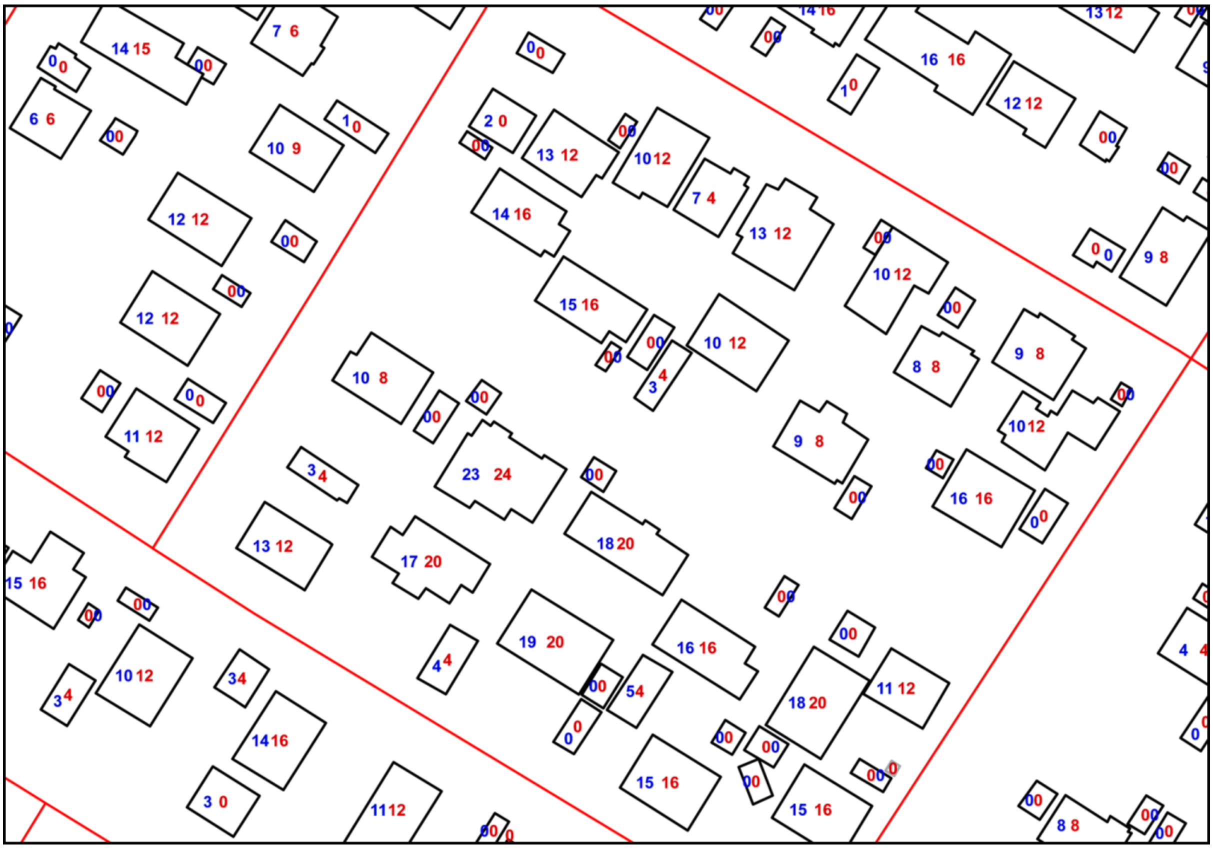


Figure 2-4: Visual assessment for low-rise-building area
Estimated population (left value) vs. actual population (right value)

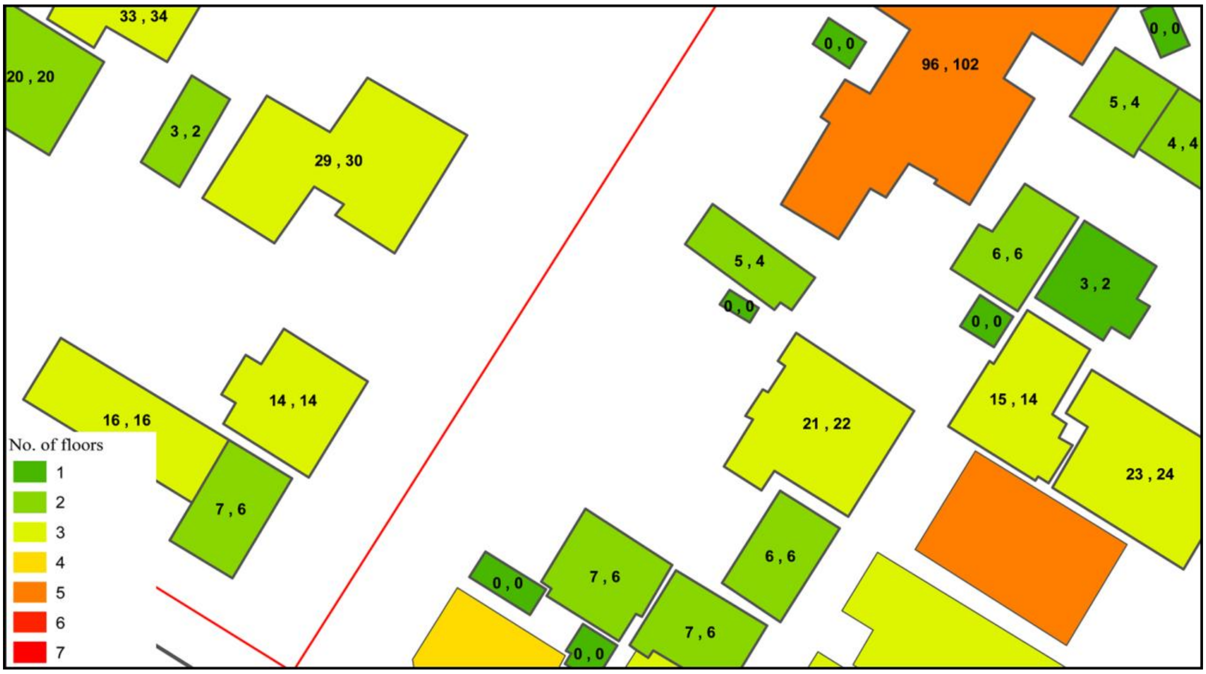


Figure 2-5: Visual assessment for high-rise-building area
Estimated population (left value) vs. actual population (right value)

2.5.2 Statistical Assessment

To examine the **geostatistical** relationship between the two datasets (estimated and actual population), linear regression analysis was applied to determine the correlation coefficient, R^2. The correlation coefficient r is a measure of the linear relationship between two attributes or columns of data. Tables 2-1 shows the results of correlation coefficients for various filtered footprint areas using two estimation methods. Root mean square error (RMSE) (Equation 2-5) was also computed for each category in both estimated methods (Table 2-2). The root mean square error or RMSE is one of many ways to quantify the difference between an estimator and the true value of the quantity being estimated. The smallest RMSE was achieved using the 25 m^2 filtered building footprint applied in the Volumetric method.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\mathrm{Actual} - \mathrm{Estimated})^2}{n}} \qquad \text{Root Mean Square Error} \ldots\ldots\ldots\ (2\text{-}5)$$

With the Areametric method, the estimated values did not agree with the actual values, especially in highly populated buildings (population > 50) (Figure 2-6). This may have occurred due to the presence of high-rise buildings with a large number of floors but a small footprint. The best estimated results were achieved by using Volumetric method where all R^2 values were greater than 0.9. While all R^2 values are acceptable in the Volumetric method, the best value (R^2 = 0.9488) was achieved in the 25 m^2 filtered area category (Figure 2-7).

Table 2-1: Various correlation coefficients for Areametric and Volumetric methods

Filtered Area	Areametric Method R^2 & Y	Volumetric Method R^2 & Y
0 m^2	R^2 = 0.8004 (y = 0.5785x + 1.3214)*	R^2 = 0.9461 (y = 0.8625x + 0.3699)
05 m^2	R^2 = 0.8003 (y = 0.5795x + 1.3241)	R^2 = 0.9461 (y = 0.8628x + 0.3708)
10 m^2	R^2 = 0.7995 (y = 0.5898x + 1.3022)	R^2 = 0.9467 (y = 0.8688x + 0.3765)
15 m^2	R^2 = 0.7995 (y = 0.6160x + 1.2197)	R^2 = 0.9468 (y = 0.8794x + 0.3748)
20 m^2	R^2 = 0.7990 (y = 0.6431x + 1.1348)	R^2 = 0.9479 (y = 0.8934x + 0.3402)
25 m^2	R^2 = 0.8002 (y = 0.6631x + 1.0703)	R^2 = 0.9488 (y = 0.9037x + 0.3088)*
30 m^2	R^2 = 0.7971 (y = 0.6739x + 1.0312)	R^2 = 0.9458 (y = 0.9117x + 0.2773)
35 m^2	R^2 = 0.7954 (y = 0.6823x + 1.0078)	R^2 = 0.9439 (y = 0.9189x + 0.2571)
40 m^2	R^2 = 0.7944 (y = 0.6926x + 0.9751)	R^2 = 0.9425 (y = 0.9275x + 0.2306)

* Best results

Table 2-2: Root mean square error (RMSE) for both Areametric and Volumetric methods

Filtered Area	RMSE (Areametric)	RMSE (Volumetric)
0 m^2	0.0315520	0.0153020
5 m^2	0.0315193	0.0152933
10 m^2	0.0312114	0.0150973
15 m^2	0.0304115	0.0148670
20 m^2	0.0296994	0.0145070
25 m^2	0.0291742	0.0142609*
30 m^2	0.0290846	0.0145403
35 m^2	0.0290064	0.0147061
40 m^2	0.0288724*	0.0148335

* Best results

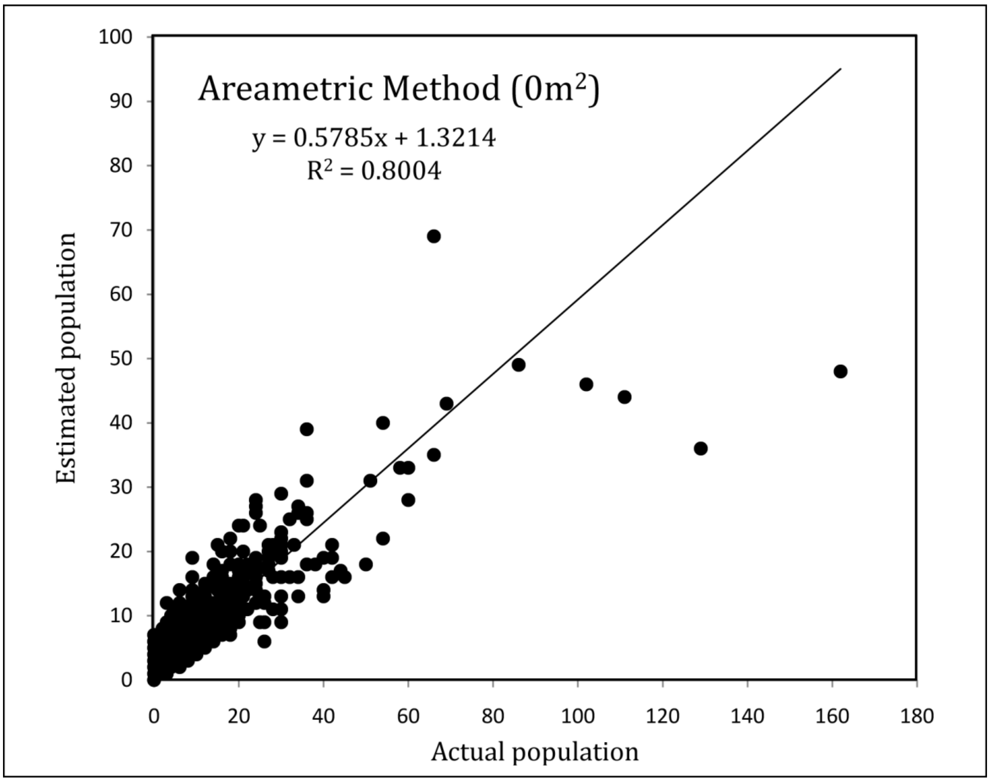


Figure 2-6: Scatter plot for 0 m^2 filtered area in Areametric method (Total sample size = 8854)

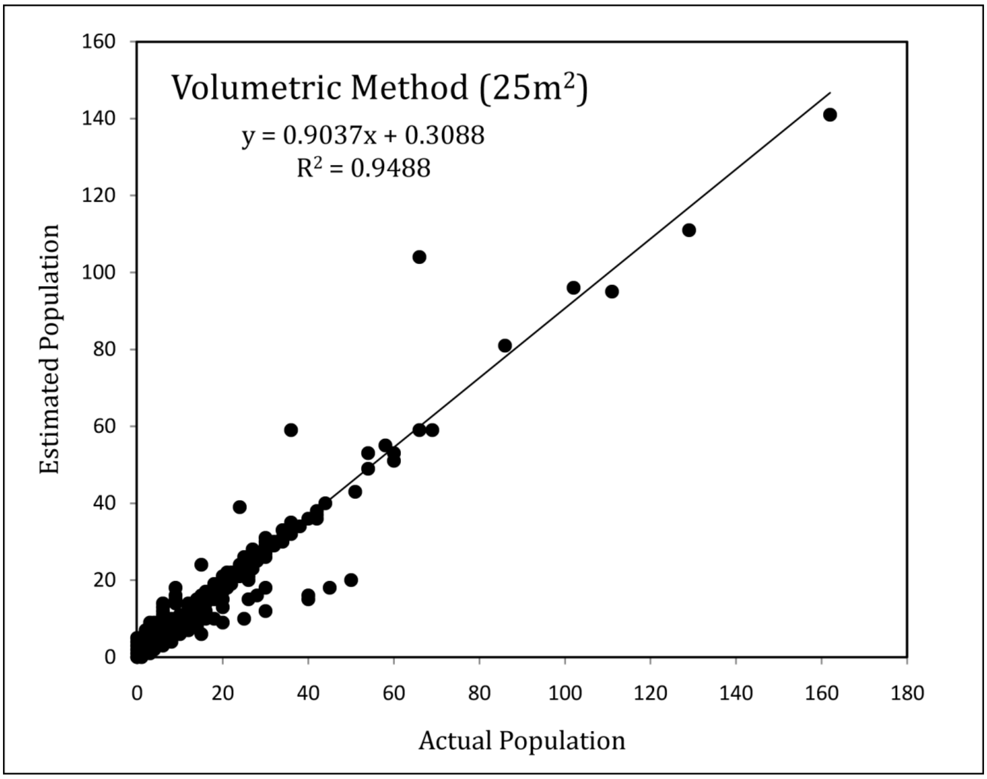


Figure 2-7: Scatter plot for 25 m² filtered area in Volumetric method (Total sample size = 8854)

2.5.3 Spatial Assessment

The purpose of this assessment is to compare spatial distribution patterns between estimated and actual building population. There are several ways to perform **spatial autocorrelation**, the most popular being Moran's I and Geary's C. Like autocorrelation, spatial autocorrelation means that adjacent observations of the same phenomenon are correlated. However, autocorrelation is about proximity in time. Spatial autocorrelation is about proximity in (two-dimensional) space. Spatial autocorrelation is more complex than autocorrelation because the correlation is two-dimensional and bi-directional.

Moran's I was used to compare each value in the pairs to the mean value for all features in the study area, which is also known as global Moran's I. Moran's I was computed for each filtered category in both the Areametric and Volumetric methods based on the estimated population class field using ArcGIS program. This tool measures the spatial autocorrelation (feature similarity or dissimilarity) not only on feature locations or attribute values alone but also on feature locations and feature values simultaneously.

In general, a Moran's I index value near +1.0 indicates clustering while an index value near −1.0 indicates dispersion. A high positive Z score for a feature indicates that the surrounding features have similar values. A low negative Z score indicates that the feature is surrounded by dissimilar values. Moreover, Moran's I for actual population is also computed for comparison. Although Moran's I measures the patterns to determine whether the features are clustered or dispersed, the purpose of using Moran's I in this study was to measure the patterns for each category and then compare them with the feature patterns of the actual building population. The following (Table 2-3) shows Moran's I and Z for actual building population data and Table 2-4 and 2-5 show the

Moran's I indexes for each filtered category for both the Areametric and Volumetric methods. Although both indices intersect at certain filtered areas (Figure 2-8), one of the Z scores of the Volumetric method intersected at a point between the 20 and 25 m^2 footprint areas. This is probably the average single-unit living space in the study area. This Volumetric method with filtered 25m filtered category was applied to compute Tsukuba central area building population.

Table 2-3: Moran's I and Z score for actual building population

Actual Building Population

Filtered Area	Index	Expected	Variance	Z score
None	0.03271	−0.00021	0.00000	62.05492

Table 2-4: Moran's I and Z score for estimated building population in Areametric method

Areametric Method

Filtered Area	Index	Expected	Variance	Z score
None	0.02508	−0.00014	0.00000	65.48493
05 m^2	0.02529	−0.00014	0.00000	66.06920
10 m^2	0.03178*	−0.00014	0.00000	81.18271*
15 m^2	0.04734	−0.00016	0.00000	111.56468
20 m^2	0.06082	−0.00018	0.00000	130.03565
25 m^2	0.06845	−0.00020	0.00000	137.12042
30 m^2	0.07450	−0.00021	0.00000	143.36652
35 m^2	0.07615	−0.00021	0.00000	142.61562
40 m^2	0.07727	−0.00022	0.00000	141.69034

* Best results

Table 2-5: Moran's I and Z score for estimated building population in Volumetric method

Volumetric Method

Filtered Area	Index	Expected	Variance	Z score
None	0.01851	–0.00018	0.00000	39.79721
05 m^2	0.01726	–0.00011	0.00000	54.03807
10 m^2	0.01867	–0.00018	0.00000	40.15002
15 m^2	0.02128	–0.00018	0.00000	45.31663
20 m^2	0.02845	–0.00019	0.00000	58.47552
25 m^2	0.03451*	–0.00020	0.00000	68.89462*
30 m^2	0.03799	–0.00021	0.00000	73.47143
35 m^2	0.03892	–0.00022	0.00000	73.43831
40 m^2	0.04002	–0.00022	0.00000	74.06918

* Best results

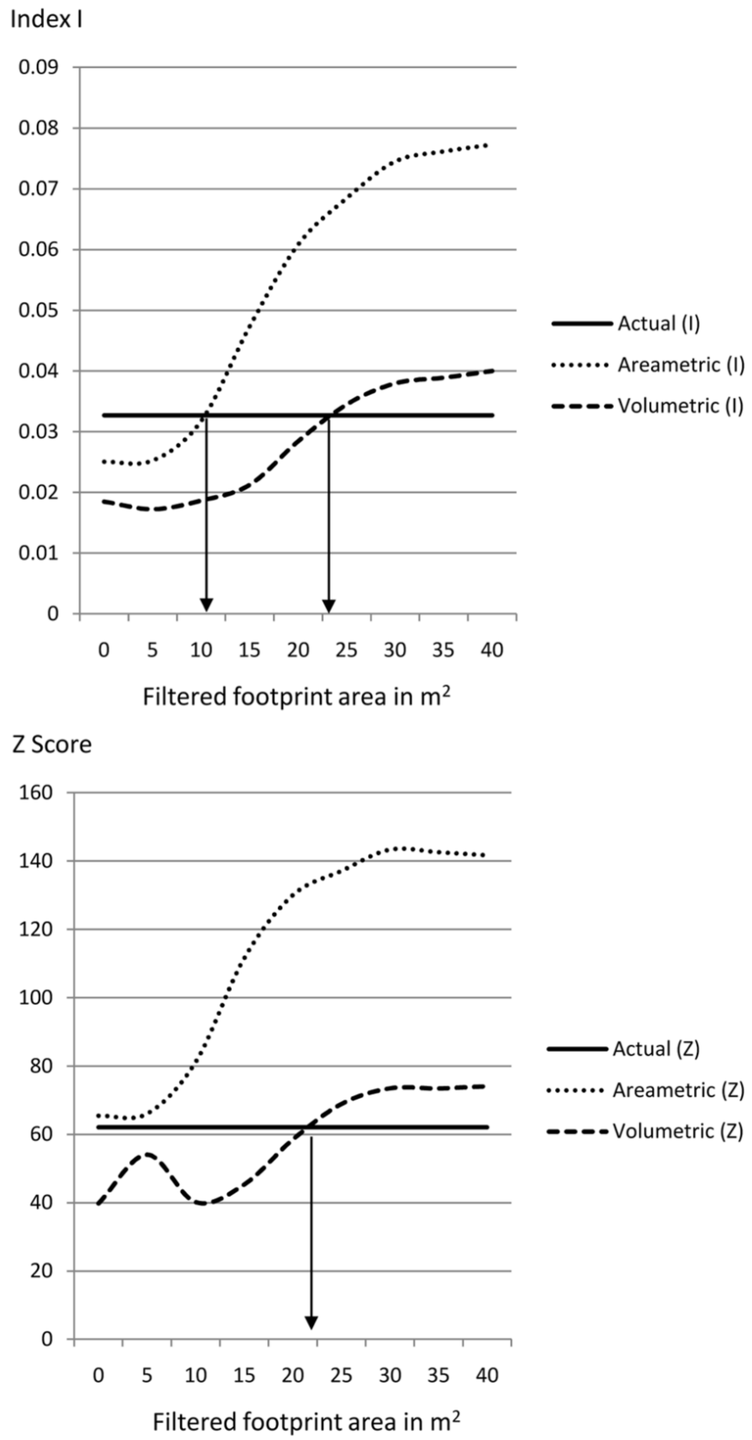


Figure 2-8: Moran's I (top) and Z score (bottom) for actual population vs. estimated population (using Areametric and Volumetric methods for various filtered footprint areas)

Chapter 3

Estimation of Building Population

3.1 Study Area

The study area is Tsukuba central area (Figure 3-1 and 3-2) also known as Tsukuba Science City which is a planned city for research and scientific purposes. Tsukuba is sometimes considered as a part of the Greater Tokyo Area and about 50 km Northeast of Tokyo. In 1963, the new city's construction plan was approved and by 1980 more than 40 research and higher education facilities had been built. Roads, water and sewerage systems, parks and other facilities had also been constructed by then. In Tsukuba there are 88 parks and green areas with a total area of 100 ha. Each of them has been designed individually according to its location and function. These areas serve as places for the residents to rest, do recreational activities, and participate in sports. Many of these parks, as well as other public, commercial, and educational facilities and residential buildings are connected by a path of 48 km (31 miles) for pedestrians and cyclists.

Twenty-two years after the new urban construction project was approved, the International Exposition of Science and Technology (Expo '85) was held in

1985 in order to commemorate the completion of the Tsukuba Science City. Based on the theme of "Humanity, Residence, Environment, and Science and Technology", the exposition established Tsukuba a reputation, both at home and abroad, as Japan's pre-eminent center for science and technology.

Nowadays, the city is home to the country's national testing and research facilities encompassing such research fields as science, industry, agriculture and forestry, environment and space development. It also houses institutions of higher learning including the University of Tsukuba campuses. About 19,000 researchers (40% of the total for the whole country), 5,000 of whom hold doctorate degrees, are conducting cutting-edge research here.

According to emergence of research and development facilities in Tsukuba City, the population is ever increasing and the government requires to establish public facilities such as bus stations, hospitals, glossary stores, schools, shopping centers, etc. Another aspect of selecting Tsukuba City for case study is population integrity. Most of the people are living and working inside the Tsukuba City and home of the University of Tsukuba. Rich of spatial data availability of Tsukuba City (such as high resolution aerial imagery, LIDAR altitude data and other fine scale GIS data) is another favor of study area selection. Although, national census investigation (door-to-door) takes place five year intervals in Japan, most up-to-date resident registration information by wards can be get from city or local government office by monthly intervals in Tsukuba City. The study area is part of the Tsukuba City. In this study, total 94 census tracts and population 84955 were used. Total study area is 5959.13 ha.

Idx	Census Tract Name	Idx	Census Tract Name	Idx	Census Tract Name	Idx	Census Tract Name	Idx	Census Tract Name	Idx	Census Tract Name
1	Kamihara	17	Sengen 1 Chome	33	Yamanaka	49	Matsushiro 4 Chome	65	Takezono 2 Chome	81	Hachimandai
2	Kamizakai	18	Sengen 2 Chome	34	Shima	50	Matsushiro 5 Chome	66	Takezono 3 Chome	82	Kitasato
3	Kamiyokoba	19	Hara	35	Teshirogi	51	Matsuzuka	67	Hanamuro	83	Kitahara
4	Ueno	20	Furuku	36	Arai	52	Syoei	68	Hanabatake 1 Chome	84	Owashi
5	Shimohiratsuka	21	Azuma 1 Chome	37	Kasuga 1 Chome	53	Matsunoki	69	Karima	85	Tennodai 1 Chome
6	Namiki 2 Chome	22	Azuma 2 Chome	38	Kasuga 2 Chome	54	Yagihashi	70	Katsuraginesaki	86	Tennodai 3 Chome
7	Namiki 3 Chome	23	Azuma 3 Chome	39	Kasuga 3 Chome	55	Shibasaki	71	Hasununa	87	Okamurashinden
8	Namiki 4 Chome	24	Azuma 4 Chome	40	Ksuga 4 Chome	56	Sakae	72	Nishiohashi	88	Miyukigaoka
9	Nakauchi	25	Amakubo 1 Chome	41	Higashioka	57	Sakura 1 Chome	73	Nishi Onuma	89	Honzawa
10	Nakane	26	Amakubo 2 Chome	42	Higashihiratsuka	58	Sakura 2 Chome	74	Nishioka	90	Kurihara
11	Ninomiya 1 Chome	27	Amakubo 3 Chome	43	Higashi Arai	59	Sakura 3 Chome	75	Nishihiratsuka	91	Nishihara
12	Ninomiya 2 Chome	28	Amakubo 4 Chome	44	Higashi 1 Chome	60	Umezono 1 Chome	76	Saigo	92	Nishizawa
13	Ninomiya 3 Chome	29	Tennodai 2 Chome	45	Higashi 2 Chome	61	Umezono 2 Chome	77	Kaname	93	Konda
14	Ninomiya 4 Chome	30	Saiki	46	Matsushiro 1 Chome	62	Yokomachi	78	Tateno	94	Nagamine
15	Imaizumi	31	Onozaki	47	Matsushiro 2 Chome	63	Inarimae	79	Uenomuro		
16	Kurakake	32	Onogawa	48	Matsushiro 3 Chome	64	Takezono 1 Chome	80	Namiki 1 Chome		

Figure 3-1: Study area and road network

Figure 3-2: Census tracts (Polygon) overlay on 8 cm orthoimages

3.2 List of Data

In this study, LIDAR data was used in ESRI point format for both a Digital Surface Model (DSM) and a Digital Terrain Model (DTM), provided by PASCO Corporation. Digital Surface Model (DSM) is a topographic model of the earth's surface that includes buildings, vegetation, roads, and natural terrain features. The key benefit of the Digital Surface Model (DSM) is that it provides a geometrically correct base map. Digital Terrain Model (DTM) is a topographic model of the bare earth that has had vegetation, buildings, and other cultural features digitally removed, enabling users to infer terrain characteristics possibly hidden in the Digital Surface Model (DSM).

Each Digital Surface Model (DSM) and Digital Terrain Model (DTM) scene is 800 m wide and 600 m long. There were 164 Digital Surface Model (DSM) and 164 Digital Terrain Model (DTM) scenes used for the whole study area. Digital Surface Model (DSM) points are on average 0.9 m apart and Digital Terrain Model (DTM) points are at a regular 5 m spacing. PASCO also provided 8 cm orthoimages, which were acquired along with the LIDAR surveying. Table 3-1 and Table 3-2 show the list of data used in this study.

Moreover, iTownpage from Nippon Telegraph & Telephone Corp. (NTT) and Tsukuba City monthly registration data were used in this study. Table 3-1 and 3-2 shows the list of data and purpose to be used in this study.

Table 3-1: List of data and purpose to be used (Spatial Data)

Data and Source	Description	Purpose
Digital Surface Model (DSM) (Source: PASCO Corp.)	• Point feature in ESRI shape format • Average point spacing is 0.9 m • One scene (800 m X 600 m) • Acquired in 2006 • A total of 164 scenes were used	To compute Digital Height Model (DHM) where: DHM = **DSM** – DTM
Digital Terrain Model (DTM) (Source: PASCO Corp.)	• Point feature in ESRI shape format • Each point is 5 m regular spacing • One scene (800 m X 600 m) • Acquired in 2006 • A total of 164 scenes were used	To compute Digital Height Model (DHM) where: DHM = DSM – **DTM**
orthoimages (Source: PASCO Corp.)	• GeoTIFF format • 8 cm X 8 cm spatial resolution • RGB True Color • One scene (800 m X 600 m) • Acquired in 2006 (along with LIDAR surveying) • A total of 164 scenes were used	Used as a base map and landscape visualization
Building Footprints (Source: ZMapTOWN-II product by Zenrin Co. Ltd.)	• Polygon feature in ESRI shape format which includes building name, block number, number of floor and other attribute information • Acquired in 2006 • Map Scale 1:2,500	To compute building population
Census Tracts (Source: ZMapTOWN-II product by Zenrin Co. Ltd.)	• Polygon feature in ESRI shape format which includes administration boundary units and names • Acquired in 2006 • Map Scale 1:2,500	To compute building population

Table 3-2: List of data and purpose to be used (Non Spatial Data)

Data and Source	**Description**	**Purpose**
Tsukuba City resident registration (Source: Tsukuba City Office)	• Microsoft Excel Sheet • To combine with census tracts polygons and use as census tracts population • Monthly updated • Used 2006-07-01 resident registration information	To compute building population
iTownpage (Source: Nippon Telegraph & Telephone Corp. (NTT))	• Comma Separated Value (CSV) format • Information of public facilities and business locations • Business name, main category, sub category, business contents, address, phone, URL, etc.	To separate residential and non-residential building To detect mixed building use type

3.3 Data Processing

3.3.1 Digital Height Model (DHM) and Digital Volume Model (DVM) Generation

Traditionally, stereo images matching is a standard photogrammetric technique to generate Digital Surface Model (DSM). However, this technique is good only for open smooth terrain surface. The quality of Digital Surface Model (DSM) in built-up areas is poor due to occlusions and height discontinuities (Haala and Brenner, 1999). The use of Light Detection and Ranging (LIDAR) data for terrain models and topographic mapping has gained wide attention in the recent years as it contains the height data on the earth surface. Global Positioning System (**GPS**) and Inertial Measurement Unit (IMU) acquire the position and the attitude of the airplane, and the laser measurement unit acquires the distance to the earth surface. After that the point coordinate X, Y, Z is obtained by analyzing the position, altitude, distance and projection transformation. Since this data includes many Non Surface Objects such as buildings, trees and cars, the classification is needed in various application fields. This classification is called the filtering process of LIDAR data. LIDAR offers an alternative to in situ field surveying and photogrammetric mapping techniques for the collection of elevation data. LIDAR technology can be used to provide elevation data that is accurate, timely, and increasingly affordable in inhospitable terrain.

Light Detection and Ranging (LIDAR) is a terrain and urban information acquisition technique based on laser technology. It uses a downward-pointing laser, transmitting very short pulses or a modulated signal in the visible or near infrared part of the electromagnetic spectrum (Rees, 1999). Back-scattered radiation is detected and analyzed for time delay, amplitude, or frequency, depending on the application. LIDAR is quite similar to radar, but utilizes shorter wavelengths. Other terms such as LIDAR, laser radar, laser fluoro-sensor, and laser bathy meter are also used for various applications of this technology. LIDAR techniques have been studied and utilized since the early 1960s, but appear to have become more prominent in the past few years. LIDAR has found applications in a wide variety of fields of study, including atmospheric science, bathymetric data collection, law enforcement, telecommunications, and even steel production (Maune *et al*., 2000). Advantages of using LIDAR for terrain and urban applications include the following: LIDAR allows rapid generation of a large-scale DTM (digital terrain model); LIDAR is daylight independent, is relatively weather independent, and is extremely precise (3Di, LLC, 2000). In addition, because LIDAR operates at much shorter wavelengths, it has higher accuracy and resolution than microwave radar (Jelalian, 1992).

Choosing the appropriate method for generating the Digital Surface Model (DSM) and Digital Terrain Model (DTM) is important in LIDAR data processing since surface height information is collected as points. Hug (1997) stated that laser scanners are the best choice for obtaining digital surface models, especially for dense urban areas. Haala and Brenner (1999) reported on similar work using airborne LIDAR data for the generation of 3D city models.

A **TIN** can be constructed by triangulating a set of vertices. The vertices are connected with a series of edges to form a network of triangles. The resulting triangulation satisfies the Delaunay triangle criterion, which ensures that no vertex lies within the interior of any of the circumcircles of the triangles in the network. If the Delaunay criterion is satisfied everywhere on the TIN, the minimum interior angle of all of the triangles is maximized. The result is that long, thin triangles are avoided as much as possible. The edges of TINS form contiguous, non overlapping triangular facets that can be used to capture the position of linear features that play an important role in a surface, such as ridgelines or stream courses. Because nodes can be placed irregularly over a surface, TINs can have a higher resolution in areas where a surface is highly variable or where more detail is desired and a lower resolution in areas that are less variable.

The input features used to create a TIN remain in the same position as the nodes or edges in the TIN. This allows a TIN to preserve all the precision of the input data while simultaneously modeling the values between known points. You can include precisely located features on a surface, such as mountain peaks, roads, and streams by using them as input features to the TIN nodes. TIN models are less widely available than raster surface models and tend to be more expensive to build and process. The cost of obtaining good source data can be high, and processing TINs tends to be less efficient than processing raster data because of the complex data structure. TINs are typically used for high-precision modeling of smaller areas, such as in engineering applications, where they are useful because they allow calculations of planimetric area, surface area, and volume.

In this process, both Digital Surface Model (DSM) and Digital Terrain Model (DTM) point features were converted into a Triangulated Irregular Network (TIN) model (i.e. TIN-DSM and TIN-DTM). Under the ArcGIS program, the TIN process allows users to convert multiple scenes at one time. This reduces the time for mosaicking. Moreover, the TIN process is faster than other interpolation processes such as IDW, SPLINE and Kriging. See details work flow in Figure 3-3.

Each TIN-DSM (Figure 3-4) and TIN-DTM (Figure 3-5) were converted into a raster format setting the spatial resolution to 0.5m. TIN Raster interpolates cell z-values from the input TIN at the specified resolution (i.e. 05 m for this study) or sampling interval to produce the output raster. Because interpolation of the input TIN surface occurs at regular intervals, some loss of information in the output raster should be expected. How well the raster represents the TIN is dependent on the resolution of the raster and the degree and interval of TIN surface variation. Generally, as the resolution is increased, the output raster more closely represents the TIN surface. Because the raster is a cell structure, it cannot maintain the hard and soft breakline edges that may be present in the TIN. Digital Terrain Model (DTM) was subtracted from the Digital Surface Model (DSM) raster layers to achieve the Digital Height Model (DHM). This Digital Height Model (DHM) raster layer was multiplied by the cell surface area (i.e. 0.25 m^2) to convert to a Digital Volume Model (DVM).

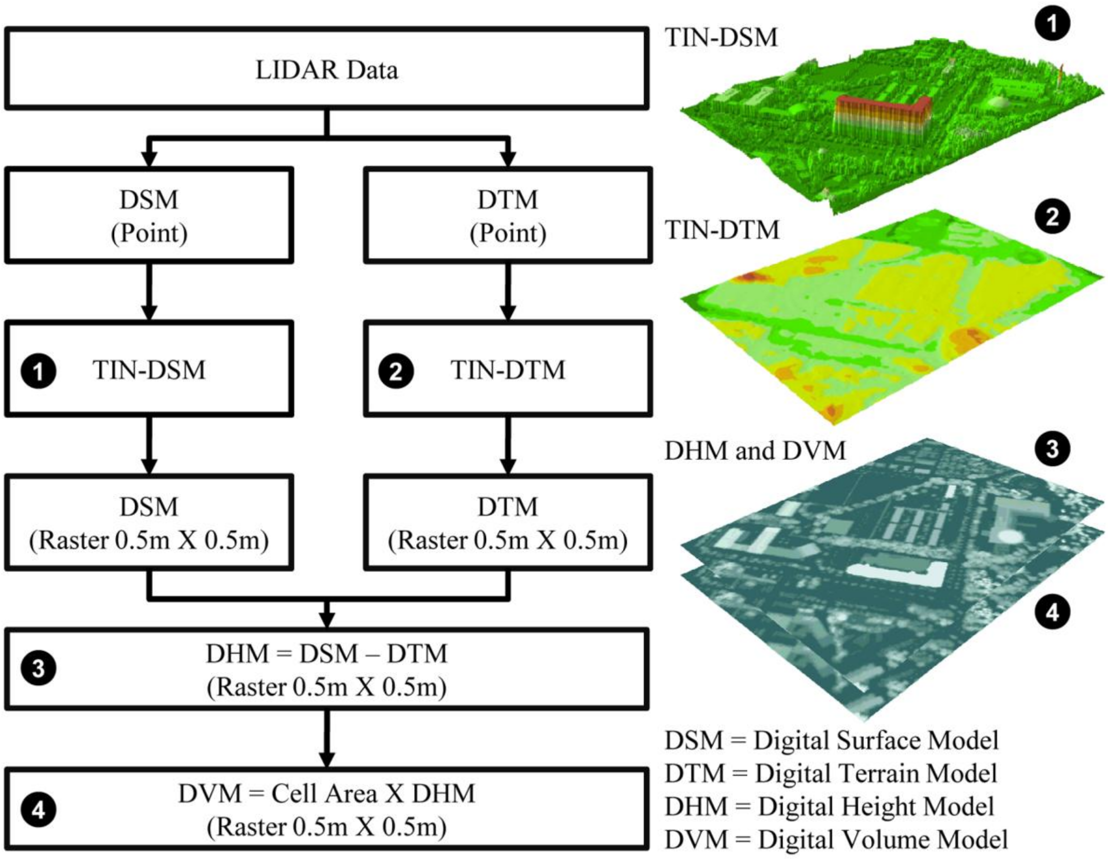


Figure 3-3: Process flow of Digital Height Model and Digital Volume Model generation from LIDAR point data

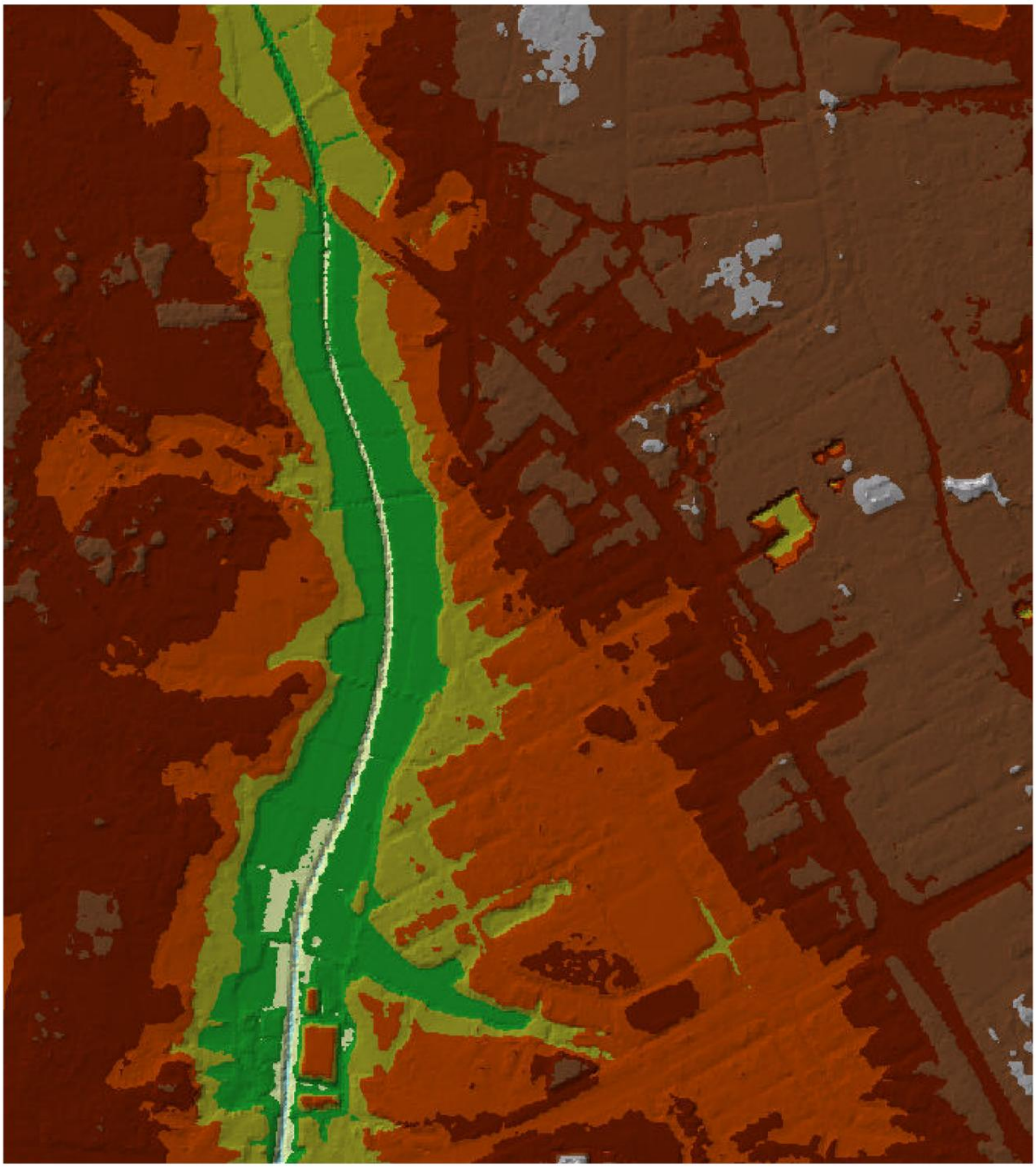


Figure 3-4: Map of Triangulated Irregular Network bare earth elevation

Figure 3-5: Map of Triangulated Irregular Network surface elevation

3.3.2 Spatial Adjustment

In this process, spatial adjustment is required to improve the measurement, due to building footprints data was not generated from LIDAR data. During the past two decades many researchers in Photogrammetry, Remote Sensing and computer vision communities have been trying to study and develop the automatic or semi-automatic approaches for building extraction and reconstruction (Gruen *et al.*, 1997; Mayer, 1999). Several approaches have been presented for building extraction from the laser altimeter data. Maas and Vosselman (1999) extracted buildings from original laser altimeter point data. Sahar and Krupnik (1999) developed a semiautomatic building extraction approach, which buildings were detected interactively and 3D building outlines were extracted using shadow analysis and stereoscopic processing. However, the aerial photographs are typically very complex and contain a large number of objects in the scene. The automatic building extraction from aerial photograph has proven to be quite difficult. Those approaches are far from being useful in practice for images of different characteristics and complex contents (Mayer, 1999). The automatic generation of building footprints from LIDAR data is presently under development and the results cannot be used in this study.

To improve the visualization and to remove the noise from Digital Height Model (DHM) data, a **Kernel** 5X5 low-pass filtering process was applied. And also filtered out pixels with heights < 2 m and then performed hill shading. This removed the cars, small bushes and other objects with heights less than 2 m (Figure 3-6).

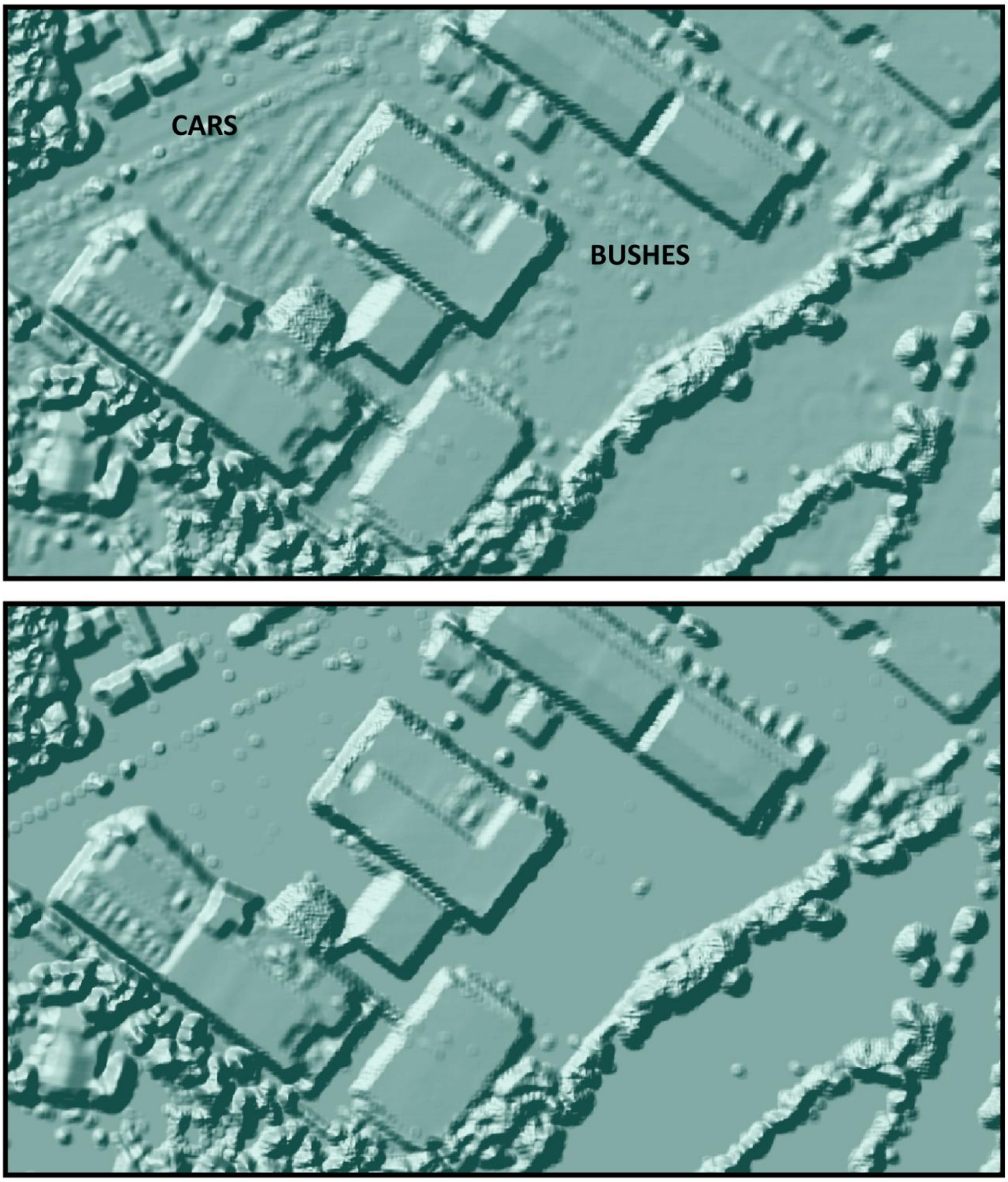


Figure 3-6: Comparison of two hill shaded building heights

To perform the spatial adjustment, building footprints **layer** was overlaid on the shaded Digital Height Model (DHM) and adjust the building footprints edges (Figure 3-7). The advantages of hillshading are well documented in landscape visualization and in its ability to portray terrain details between contours, especially in high mountain cartography. It became possible only after the 1870s when the development of the lithographic process enabled the full range of tones (Raisz, 1935).

As a digital form, **hillshade** is a grid that encodes the reflectance value off an elevation surface given a light source at a certain theoretical position in the 'sky'. It allows for the visualization of the elevation surface that photorealistic, as though you were viewing the terrain from an airplane. The **hillshade** function obtains the hypothetical illumination of a surface by determining illumination values for each cell in a raster. It does this by setting a position for a hypothetical light source and calculating the illumination values of each cell in relation to neighboring cells. It can greatly enhance the visualization of a surface for analysis or graphical display, especially when using transparency.

Using ArcGIS Editor Tool to modify the building footprint edges, vertex points, split and cut the polygons. After spatial adjustment performed, all building footprints areas were calculated again. By applying various **azimuth** and altitude in hill shade process, different building sides can be seen. Hill shading is the quickest way to understand the topographic features in cartography. In this process, altitude 45° and **azimuth** 315° were used to generate hill shade effect.

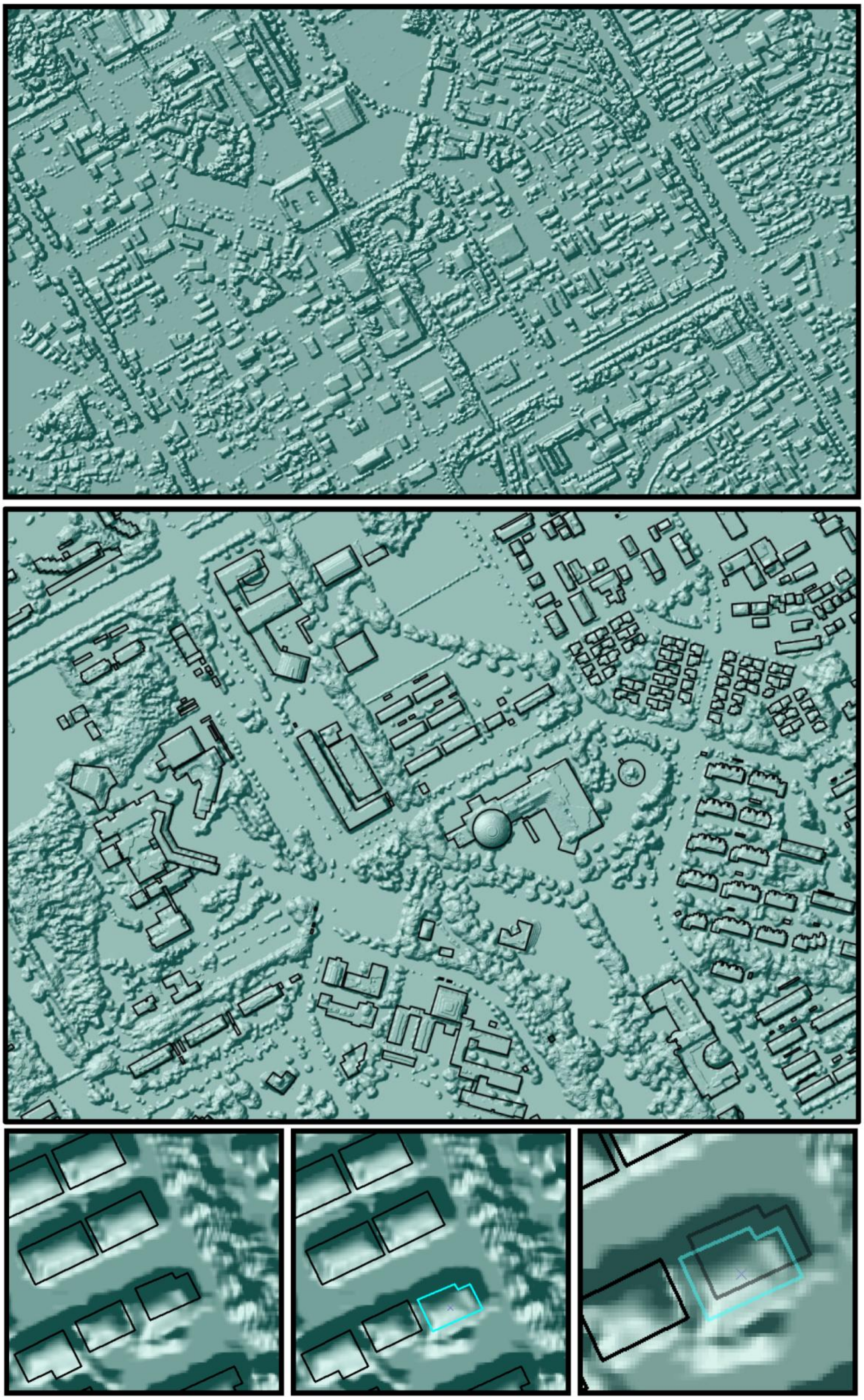


Figure 3-7: Spatial adjustment performing based on 2m above hill shaded Digital Height Model

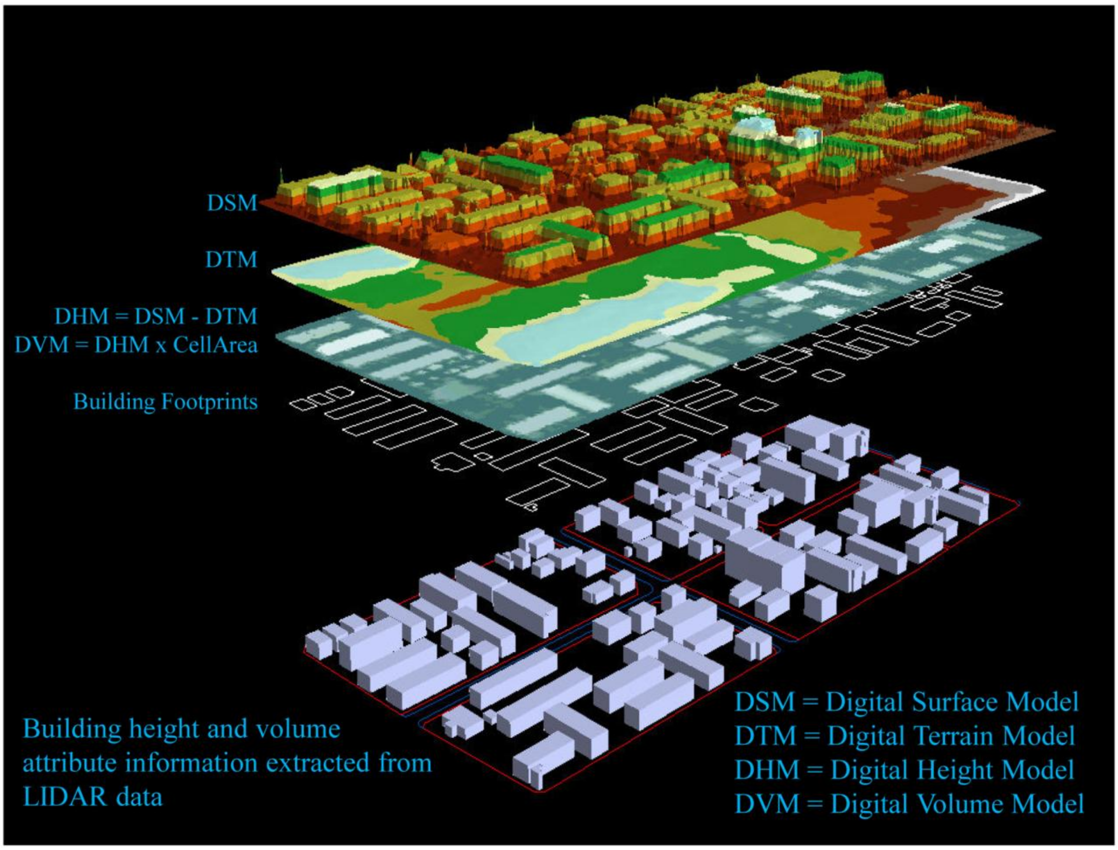
3.3.3 Building Height and Volume Extraction

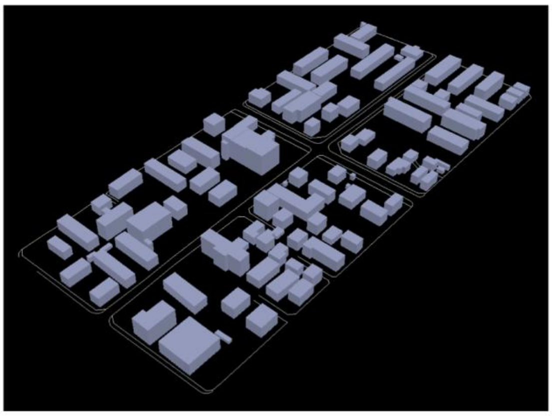
A number of publications have addressed the automatic information extraction from **LIDAR** data in many different fields. Under the automatic building footprints extraction from **LIDAR** data, the orientation and height of a roof face can be estimated accurately, however, the outline of a roof face is more difficult to determine (Vosselman and Dijkman, 2001). Förstner (1999) presented a thorough and informative discussion of the problems encountered in acquiring and establishing the building models. Although fully automatic techniques are improving, manual processing of LIDAR data is still exist.

After spatial adjustment was performed, the building height and volume attribute fields were extracted from Digital Height Model (DHM) and Digital Volume Model (DVM) (Figure 3-8), respectively, by applying **Zonal Statistics** as a Table in ArcGIS (ToolBox → Spatial Analysis Tools → Zonal → Zonal Statistics as Table). This function allows users to summarize the values of a raster (i.e. DHM and DVM raster) within the zones of another dataset (i.e. building footprint polygons) and reports the results to a table. A zone is all the cells in a raster that have the same value, regardless of whether or not they are contiguous. Raster and feature data sets can be used as the "zone data set." So, for example, residential is a zone of a land use raster data set, or a roads feature data set can be the zone for an accident data set. Zonal statistical functions perform operations on a per-zone basis; a single output value is computed for every zone in the input zone data set. Zonal statistics are used to calculate statistics such as the mean, sum, min, max and standard deviation of elevation for each building footprint (Figure 3-8).

Zonal statistical function is useful for finding average slope per watershed, average soil nitrogen content per paddy field, average surface temperature by land use category and so on. Later this table was joined with building footprints polygons based on the same polygon ID. Joining data is typically used to append the fields of one table to those of another through an attribute or field common to both tables. Finally the building footprint polygons with average building height, standard deviation of building height and total building volume attribute fields were obtained. The standard deviation of each building footprint is also useful for identification of building shape (i.e. regular or irregular shape), because the building shape is strongly related to family units.

The use of building height and volume are numerous, for example, urban planning has been identified as a major benefactor of realistic 3D **geovisualization** based on building height. LIDAR has been used for topographic mapping of forested terrain and other areas not suitable for aerial photography (Wever and Lindenberger, 1999). As each data point is **georeferenced**, the LIDAR data can also be easily merged with other data sources (Kletzli and Peterson, 1998).





PID	AREA	H_MEAN	H_SOURCE	H_STD	V_SUM	V_SOURCE
PID_720	174.5978	5.5873	LIDAR	0.9845	972.1980	LIDAR
PID_721	176.9578	6.0804	LIDAR	0.5258	1076.2400	LIDAR
PID_722	188.9076	6.4752	LIDAR	1.5185	1222.2000	LIDAR
PID_724	82.1189	3.9560	LIDAR	1.8953	325.3770	LIDAR
PID_725	43.9445	2.4470	LIDAR	0.5332	107.0560	LIDAR
PID_726	74.8091	2.9327	LIDAR	0.6126	219.2170	LIDAR
PID_727	144.4233	3.2186	LIDAR	1.2545	466.7000	LIDAR
PID_728	17.2848	3.7575	LIDAR	0.8217	64.8162	LIDAR
PID_729	1134.1210	8.4507	LIDAR	4.7853	9585.2002	LIDAR
PID_730	247.4970	5.3646	LIDAR	1.1407	1327.7400	LIDAR
PID_732	144.5287	4.4927	LIDAR	1.8652	650.3230	LIDAR
PID_733	202.2426	6.1844	LIDAR	1.6201	1252.3500	LIDAR
PID_734	125.2634	6.0588	LIDAR	1.5820	761.8950	LIDAR
PID_735	12.0599	2.4802	LIDAR	0.3349	29.7623	LIDAR
PID_736	439.6767	10.4085	LIDAR	3.6050	4579.7500	LIDAR
PID_737	216.5759	5.7869	LIDAR	1.8010	1251.4100	LIDAR

Figure 3-8: Extraction of building average height and total volume

3.3.4 Conversion of iTownpage CSV into Point Feature

In order to separate residential and non-residential buildings, iTownpage data was downloaded from the NTT website, which includes business name, type, category, content, address and other information in Comma Separated Value (CSV) format. The iTownpage website supports the everyday lives and business activities of visitors and expatriates in Japan and people living overseas by enabling users to search for information on stores and businesses via the Internet.

This CSV data was converted into ESRI point features (Figure 3-9) using commercial **Geocoding** software with building level accuracy. There are several table formats that ArcMap can read, in order to bring point data into ArcMap as an Event theme; these include DBase III, DBase IV, PRN, TXT, and CSV. Files in these formats can be created in Microsoft Excel. Geo-coding is the process of finding associated geographic coordinates (often expressed as latitude and longitude) from other geographic data, such as street addresses, or zip codes (postal codes).

In this conversion process, number of shops per point attribute field was also computed by applying the **Dissolve**d Method in ArcGIS based on X, Y coordinate fields. Some points may include multiple shops such as shopping centers or malls, which share the same address. This is important for identification of mixed building use type (i.e. home-based business from shopping-centre based business), where the number of shops per point is less than in a shopping mall.

Point ID	PID_00269
Business Name	ドックキッズ
Main Category	小売業
Sub Category 1	その他の小売業/他に分類されない小売業
Sub Category 2	ペットショップ（犬）
Address	〒305-0068, 中内464－3
Telephone	029-838-1667
URL	http://dog-kids.hp.infoseek.co.jp/index2.html
EMAIL	dog-kids@mail1.accsnet.ne.jp
Comment	あなたと新しい家族との素敵な出会いをお手伝いします！
Business Contents	ペットショップ（犬）
Block Name (J)	中内
Block Name (E)	Nakauchi
Remark	

Figure 3-9: NTT Facility Points

3.3.5 Separation of Residential and Non-Residential Building

This is the heart of the estimation process because the accuracy of the building population result is fully dependent on the building use types, such as residential, non-residential and mixed. There are four steps to separate residential from non-residential buildings.

Step 1: Filter by Zero Census Tracts Building

Remove all building footprints that fall inside the zero census tracts, such as the university and research centre campus.

Step 2: Filter by Building Area and Height

Remove all building footprints with areas of less than 20 m^2 and with a height of less than 2m, thus removing bicycle stand roofs, garbage boxes, porticos and other footprints not occupied by people.

Step 3: Filter by Public Facilities

In this step, facility point layer was used which was generated from NTT iTownpage, to identify the large scale public facilities such as financial institutions, governmental organizations, educational organizations, etc. This can be done by intersecting the NTT Facility point data with the building footprints polygons. Although the geo-coding accuracy of NTT data conversion is building level, 3m tolerance **buffered** was applied to each facility point during the intersecting process.

Step 4: Manual Filtering

This step involves removing of other unpopulated building footprints, which cannot be detected automatically (i.e. footprint is larger than 20 m^2 and height is greater than 2 m), such as large building footprints for multi-storey car parking lots and farming things. In this step, 8 cm **orthoimages** and field survey were applied. Such kinds of non-living building foot prints are sometime larger than 20 m^2 and height is also higher than 2m especially in rural areas where farming practices are abundant. Figure 3-10 shows the way of manual filtering process in rural areas. 8cm **orthoimage** is quite enough to identify the human activities on the ground such as farming, construction and other social activities.

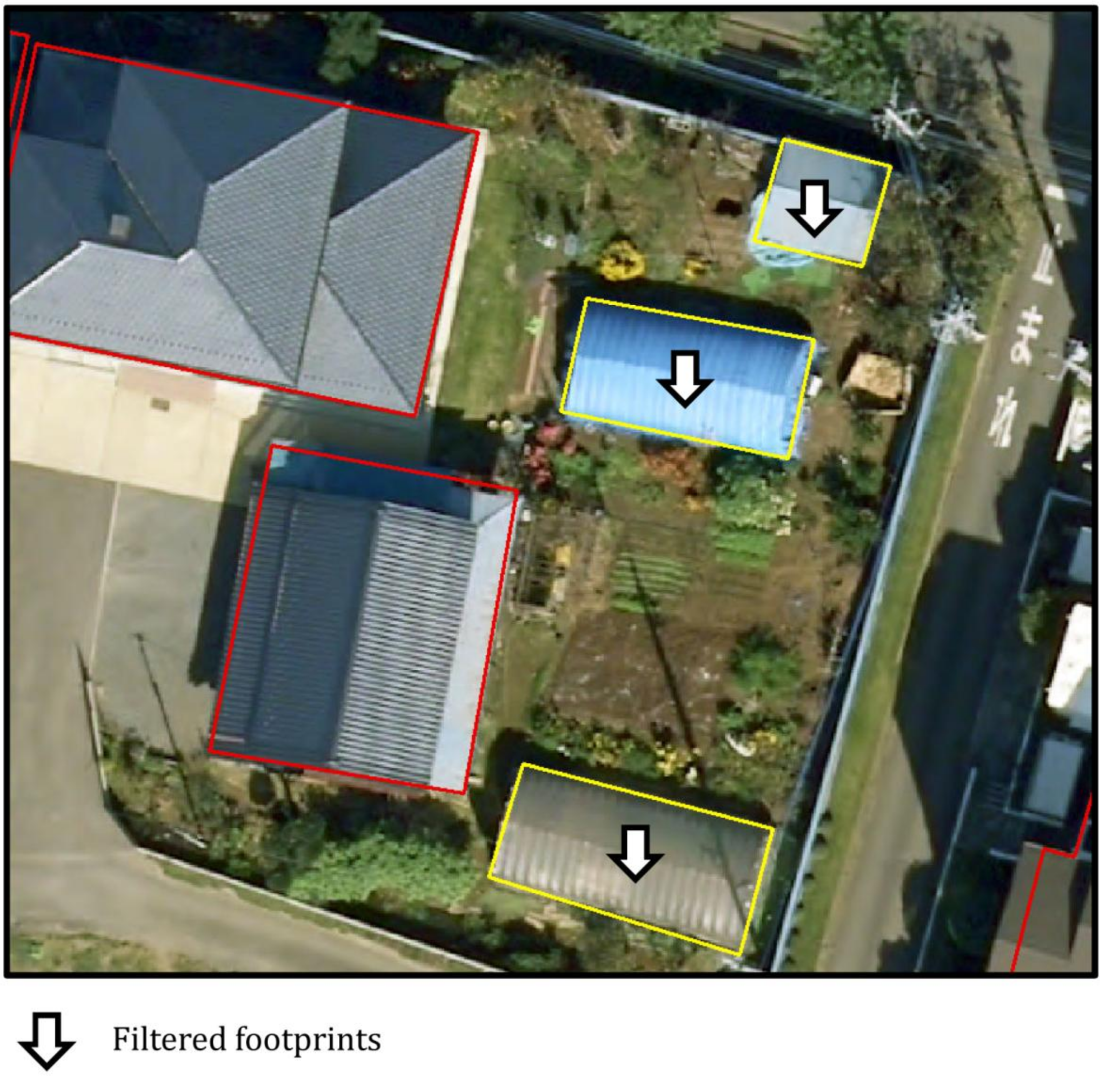


Figure 3-10: Manual filtering on 8cm orthoimage

3.3.6 Adjustment Factor and Mixed Building Use Type

In order to improve the accuracy of the estimation process, mixed building use type is also required to consider (e.g. such buildings use their first floor/ground floor for business activities such as a coffee shop, barber shop, restaurant, real estate agency, etc.) (Figure 3-11). These can be detected from the number of shops per point, sub-business category and business content attribute information from iTownpage data. Calculation of adjusted floor and volume are as follows.

$aF = tF - bF$ Adjusted floor .... (3-1)

$fV = (aF/tF)$ Volume adjustment factor .... (3-2)

$aV = tV*fV$ Adjusted volume ....(3-3)

Where:

aF Adjusted Floor

tF Total Building Floors

bF Number of business used floors

aV Adjusted Volume

tV Total Building Volume

fV Volume Adjustment Factor

For example, the adjustment factor for a four-floor building with one commercial use floor is 0.75 (3/4) and zero for an empty building (abandoned house). Additional field data collection was performed to identify the building use status. iTownpage from NTT data has business address information and this information can be used for identification of business used floor. Figure 3-12 shows the type of building use in different color after separated.

3.3.7 Generation of Building Population Attribute Field

Now ready to compute building population since the building footprint data has building use types (residential, non-residential and mixed), adjusted floor and adjusted volume attribute information (Table 3-3). PopShape **GIS** Tool generates estimated building population attribute field and the result can be analyzed under the map view by labeling the attribute values and overlaying other **GIS** data layers.

Table 3-3: Example of building footprints attribute table

PID	NAME	FLOOR	TYPE	AREA	H_MEAN	H_SOURCE	H_STD	V_SUM	V_SOURCE	BS_USE	ADJ_FLOOR	ADJ_FACTOR	ADJ_VOLUME
PID_720	東大野ハイツA	2	RES	174.5978	5.5873	LIDAR	0.9845	972.1980	LIDAR	0	2	1.00	972.1980
PID_721	東大野ハイツB	2	RES	176.9578	6.0804	LIDAR	0.5258	1076.2400	LIDAR	0	2	1.00	1076.2400
PID_722	エミール・ユキ	2	MIXED	188.9076	6.4752	LIDAR	1.5185	1222.2000	LIDAR	1	1	0.50	611.1000
PID_724	パークシティー春日	2	RES	82.1189	3.9560	LIDAR	1.8953	325.3770	LIDAR	0	2	1.00	325.3770
PID_725		0	OTHERS	43.9445	2.4470	LIDAR	0.5332	107.0560	LIDAR	0	0	1.00	107.0560
PID_726	大曽根タクシー（株）学園中央（営）	1	OTHERS	74.8091	2.9327	LIDAR	0.6126	219.2170	LIDAR	0	1	1.00	219.2170
PID_727		1	RES	144.4233	3.2186	LIDAR	1.2545	466.7000	LIDAR	0	1	1.00	466.7000
PID_728		0	OTHERS	17.2848	3.7575	LIDAR	0.8217	64.8162	LIDAR	0	0	1.00	64.8162
PID_729	住友化学工業（株）かつらぎ寮	5	OTHERS	1134.1210	8.4507	LIDAR	4.7853	9585.2002	LIDAR	0	5	1.00	9585.2002
PID_730	春日グリーンハウスB	2	RES	247.4970	5.3646	LIDAR	1.1407	1327.7400	LIDAR	0	2	1.00	1327.7400
PID_732	プロムナードカスガA	2	RES	144.5287	4.4927	LIDAR	1.8652	650.3230	LIDAR	0	2	1.00	650.3230
PID_733	エステートカスガイチバンカン	2	RES	202.2426	6.1844	LIDAR	1.6201	1252.3500	LIDAR	0	2	1.00	1252.3500
PID_734	エステートカスガサンバンカン	2	RES	125.2634	6.0588	LIDAR	1.5820	761.8950	LIDAR	0	2	1.00	761.8950
PID_735		0	OTHERS	12.0599	2.4802	LIDAR	0.3349	29.7623	LIDAR	0	0	1.00	29.7623
PID_736	エスポワール春日	4	RES	439.6767	10.4085	LIDAR	3.6050	4579.7500	LIDAR	0	4	1.00	4579.7500
PID_737	サニーウェストB	2	RES	216.5759	5.7869	LIDAR	1.8010	1251.4100	LIDAR	0	2	1.00	1251.4100

FLOOR	Total building floor (tF)
TYPE	Building use types
H_MEAN	Average height
H_STD	Height standard deviation
V_SUM	Total building volume (tV)
BS_USE	Number of business used floors (bF)
ADJ_FLOOR	After adjusted floors (aF = tF – bF)
ADJ_FACTOR	Adjustment Factor (fV = aF/tF)
ADJ_VOLUME	Adjusted volume (aV = tV*fV)

Figure 3-11: Example of mixed building use type

Residential buildings

Mixed buildings

Non-residential buildings

Figure 3-12: Identification of building use types

3.4 Results and Validation

In this study area, the building height and shape are normally related to a family unit. These buildings can be grouped into three types, for example, Single Multiple-Unit, Family Multiple-Unit and Family Single-Unit. Single Multiple-Unit (Figure 3-13a) is especially designed for college students, factory workers and other part-time workers. Usually this type of building is not more than three or four floors and the shape of the building is mostly of cubic form (i.e. a regular shape to reduce the construction cost and rental price) and the building height standard deviation is very low. This type of building can be found in surrounding of Tsukuba University campus. Family Multiple-Unit (Figure 3-13b) has more than three floors with either regular or irregular form, so their building height standard deviation is either high or low. Sometimes, additional sections are attached to the main building such as an elevator tower, swimming pool or multi-storey car parking lot. Family Single-Unit (Figure 3-13c) is standalone buildings whose shape is normally irregular and building height standard deviation is normally high. This kind of building has one or two floors.

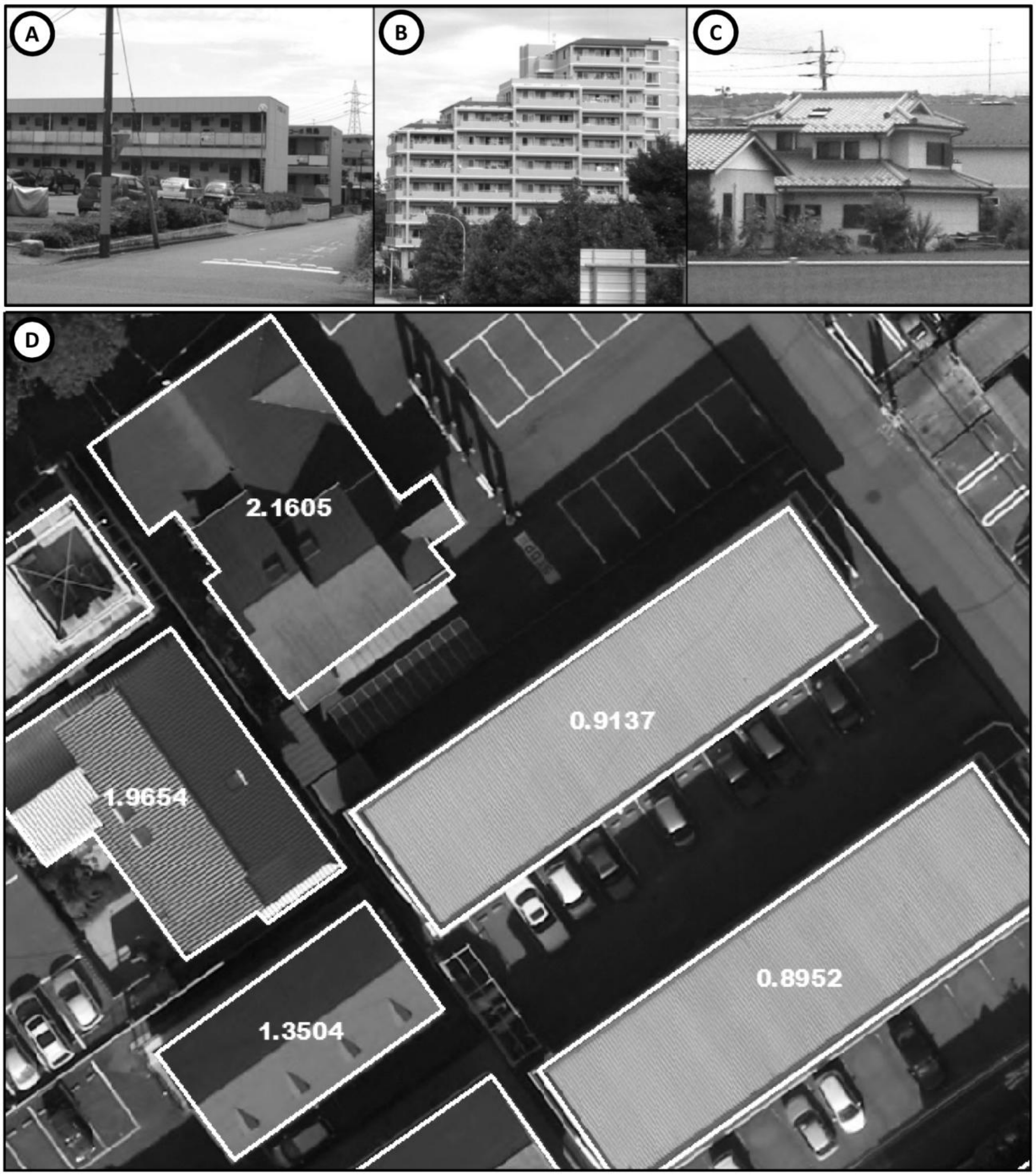
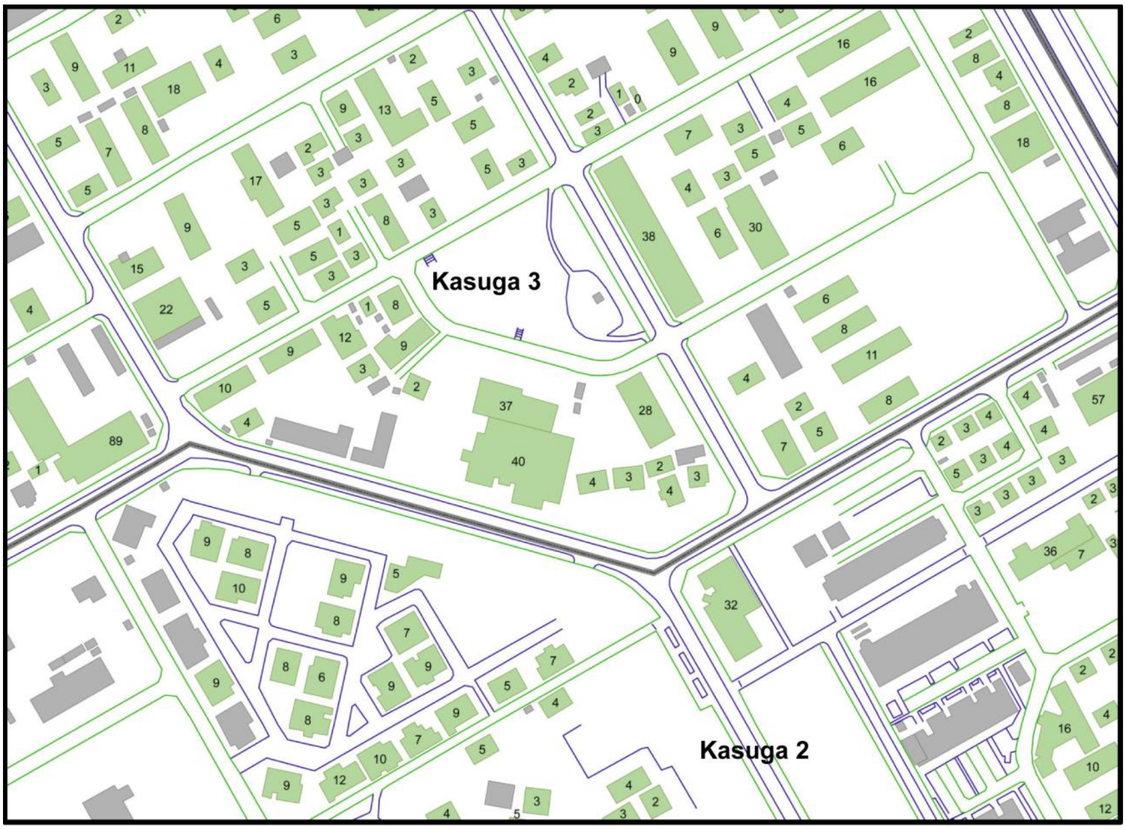


Figure 3-13: Common types of family units and building forms in study area. (A) Single Multiple Unit SMU (very low height standard deviation), (B) Family Multiple Unit FMU and (either high or low height standard deviation), (C) Family Single Unit FSU (high height standard Deviation), (D) Building height standard deviation values on 8 cm orthoimage.

Finally building population attribute field was calculated by averaging two values which estimated from floor and volume approaches (Figure 3-14). Due to privacy concerns, validation process cannot be carried out widely. Moreover, accuracy assessment of building population mapping is unlike other studies such as land use/land cover mapping, vegetation mapping and forest types mapping, because of the dynamic nature of human settlement and the limited accessibility due to privacy issues. Although, some difficulties in validation the results, there is an alternate way to perform validation by checking the mail-box usage condition of Single Multiple-Unit buildings. According to Figure 3-15, the left value of building population was estimated by Floor approach and right building population was estimated from LIDAR derived Digital Volume Model approach. The actual building population of this building is 21.

However, the estimated results between two approaches are almost identical based on correlation analyses. In this step, the correlation coefficients for both approaches (i.e. Building Floor and LIDAR approaches) were calculated (Figure 3-16). Each approach obtained best correlation coefficient $R^2 = 0.946$ in Floor approach and $R^2 = 0.9455$ in LIDAR approach. Moreover, the correlation coefficient between the two results is strongly related which $R^2 = 0.9791$ was obtained (Figure 3-17).

Figure 3-18 shows the dasymetric map of study area based on GIS estimated building population. Dasymetric mapping methods are an effective means of improving the resolution of census data. Dasymetric map can be used for effective urban and town planning, disaster management and emergency preparedness and public facility management.

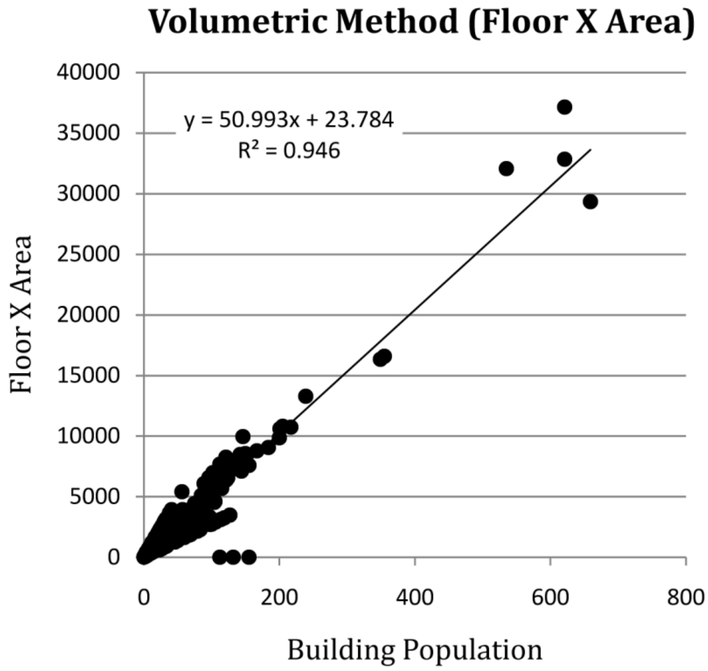


Residential building

Non residential building including height < 2m and area < 20m^2 footprints

Figure 3-14: Estimated building population
(Note: The numbers show the estimated building population)

Figure 3-15: Result validation by Single Multiple Unit mail-box usage condition
Left value is estimated from Floor approach and right value is estimated from LIDAR approach, actual building population is 21



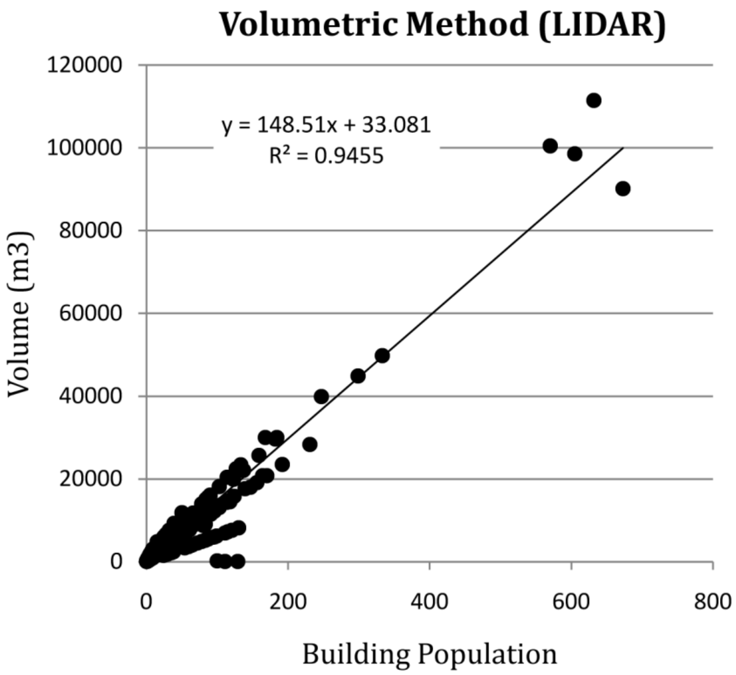


Figure 3-16: Correlation coefficients of two volumetric approaches
(Top) Correlation coefficient of building volume (Floor x Area) and estimated building population; (Bottom) correlation coefficient of building volume from LIDAR and estimated building population

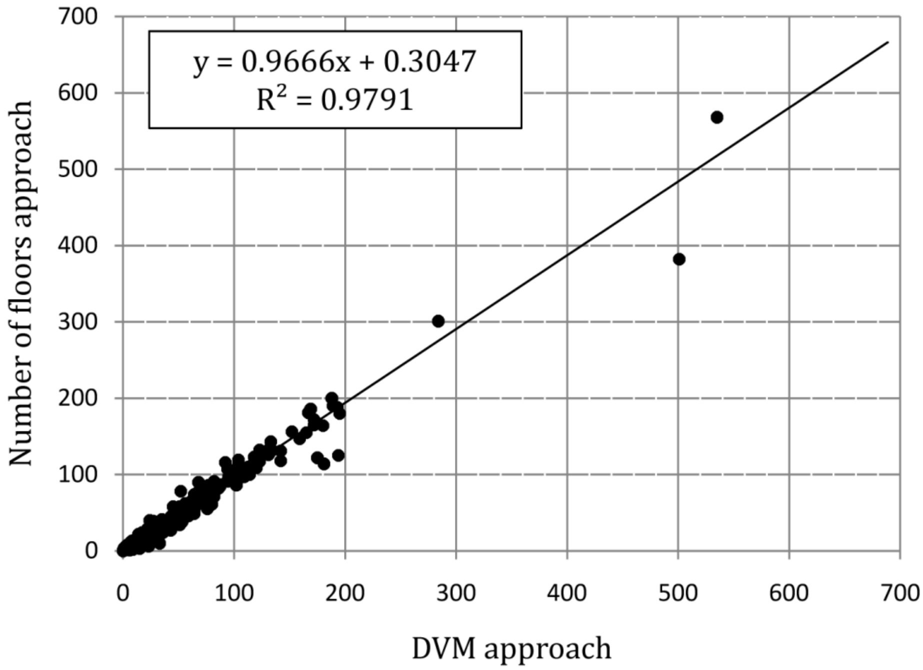


Figure 3-17: Correlation between two approaches

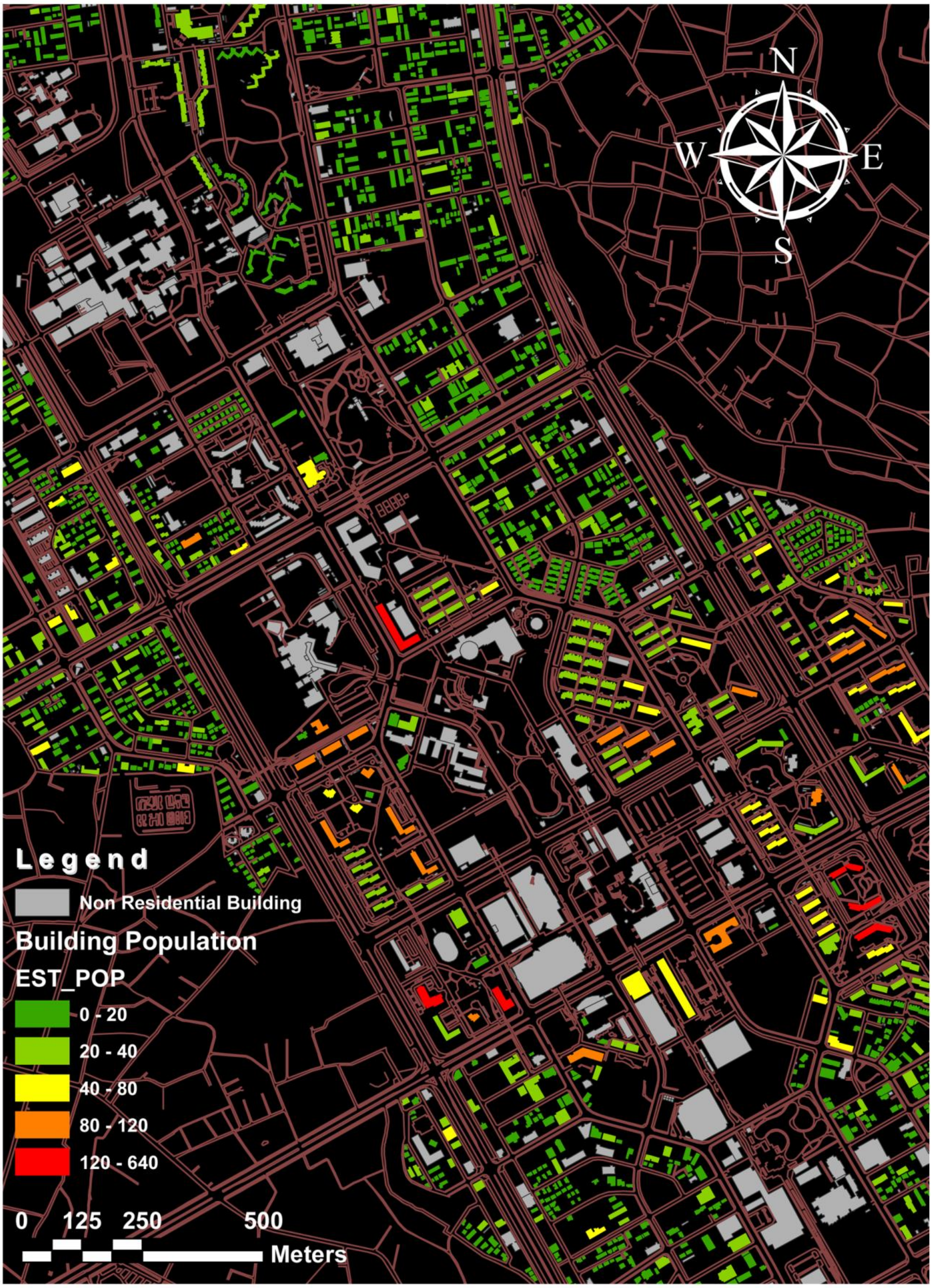


Figure 3-18: Dasymetric map of study area based on estimated building population (Lwin and Murayama, 2010)

Chapter 4

Online Micro-spatial Analysis

4.1 The Purposes of Online Micro-Spatial Analysis

The main purpose of this chapter is to describe about Online Micro-spatial Analysis based on previous estimated building population data of Tsukuba central area. An on-line GIS (also called Web GIS) is a network-based geographic information service that utilizes both wired and wireless Internet to access geographic information and analytical tools in delivering GIS services (Peng and Tsou, 2003; Tsou, 2004). An Internet GIS makes it possible to add GIS functionality to a wide range of networked computer applications. Individuals with an internet connection are able to access GIS applications from their browsers without purchasing proprietary GIS software. The Web-GIS is different from running proprietary GIS software over Local-Area Networks (LANs) or intranets on a limited number of standalone computers.

While Geographic Information Systems (GIS) contain powerful tools for entering, storing, and displaying spatially indexed information, they have been criticized for lacking equally strong spatial analytic capabilities (Anselin *et al.*, 1993; Bailey and Gatrell, 1995; Burrough, 1998). Spatial data analysis can be defined as a set of techniques for analyzing geographically referenced data (Goodchild, *et al*., 1992). These techniques range from simple descriptive measures to complex statistical inference (Anselin and Getis, 1992). The term "spatial analysis" encompasses a wide range of techniques for analyzing, computing, visualizing, simplifying, and theorizing about geographic data. Methods of spatial analysis can be as simple as taking measurements from a map or as sophisticated as complex geo-computational procedures based on numerical analysis. Spatial analysis is statistical description or explanation of either locational or attribute information or both (Goodchild, *et al*., 1992). From Fischer, *et al.* (1996), the spatial analysis includes techniques such as spatial querying, point-in-polygon operation, **buffering**, overlaying, intersection, dissolving, proximity analysis, etc.

However, spatial analysis based on building population is very rare or absent in GIS. Moreover, population data are main input in many planning processes. Spatial analysis based on building population data is a key benefit for disaster management teams in order to prepare humanitarian assistance when disaster occurs. They need specific quantitative data on population with certain geographical units such as 500 m away from along the coast lines in the case of a tsunami strike, or 5 km distant from an earthquake's epicenter. Moreover, customers' physical distribution is one of the key factors in market management and business site selection. Retailers can use the information to learn about opportunities to open new stores who make smarter decisions, enhance customer service and discover new markets and profit opportunities. Understanding the spatial distribution patterns of population, facility locations and road connection are essential for city and urban planners to ensure the establishment of sustainable city or town.

One of the main purposes of Online Micro-spatial Analysis is to measure and evaluate the urban system by performing spatial analytical functions based on distribution of three spatial elements named as population, facility and road network. This study also construct a web-based interactive micro-spatial analytical functions based on estimated building population with other GIS datasets such as public facility locations and the transportation network. This web site can be reached at the following URL (Figure 4-1) and Graphical user interface was shown in (Figure 4-2): http://land.geo.tsukuba.ac.jp/microspa/welcome.aspx

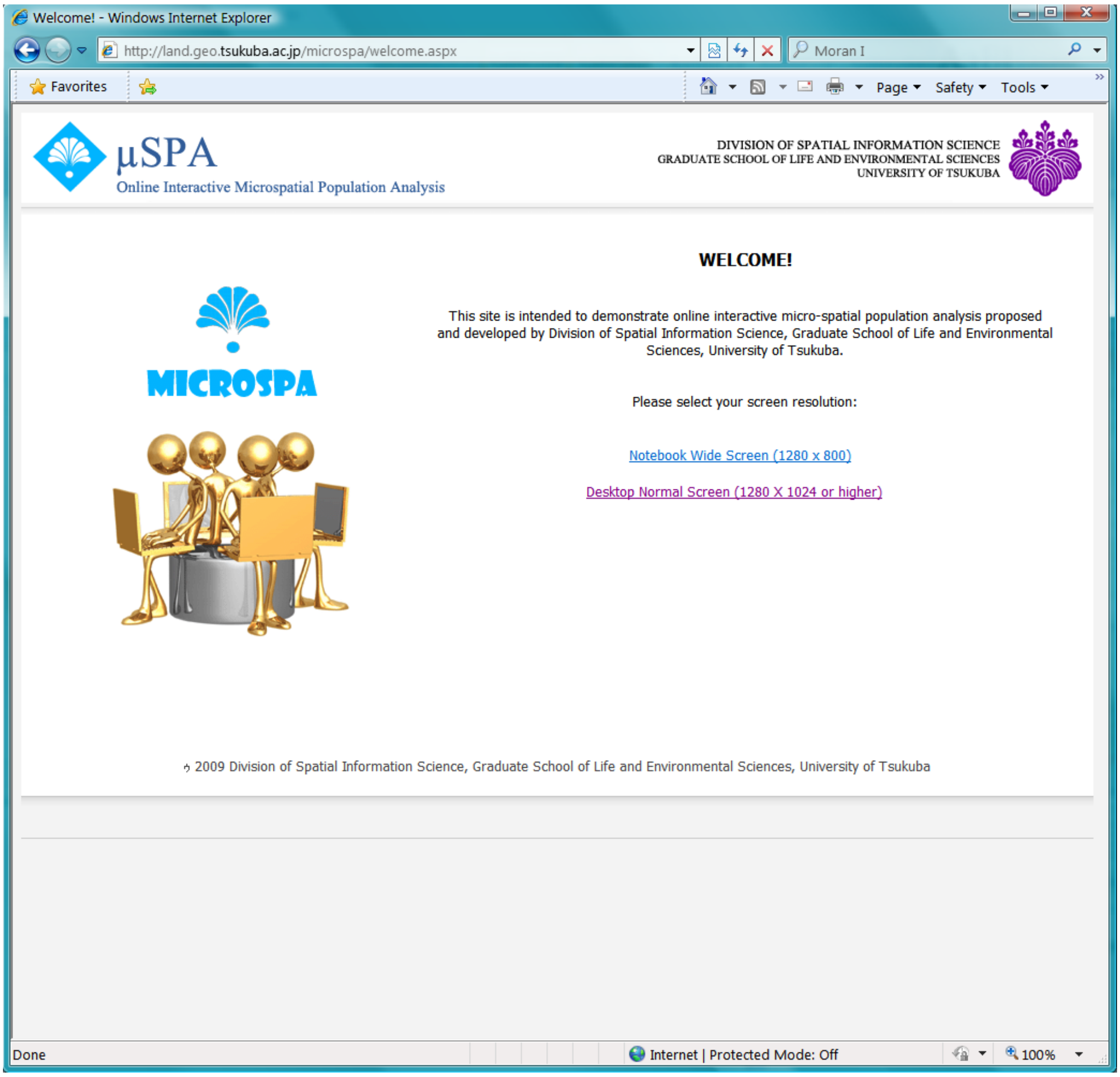


Figure 4-1: Online Micro-spatial Analysis main page

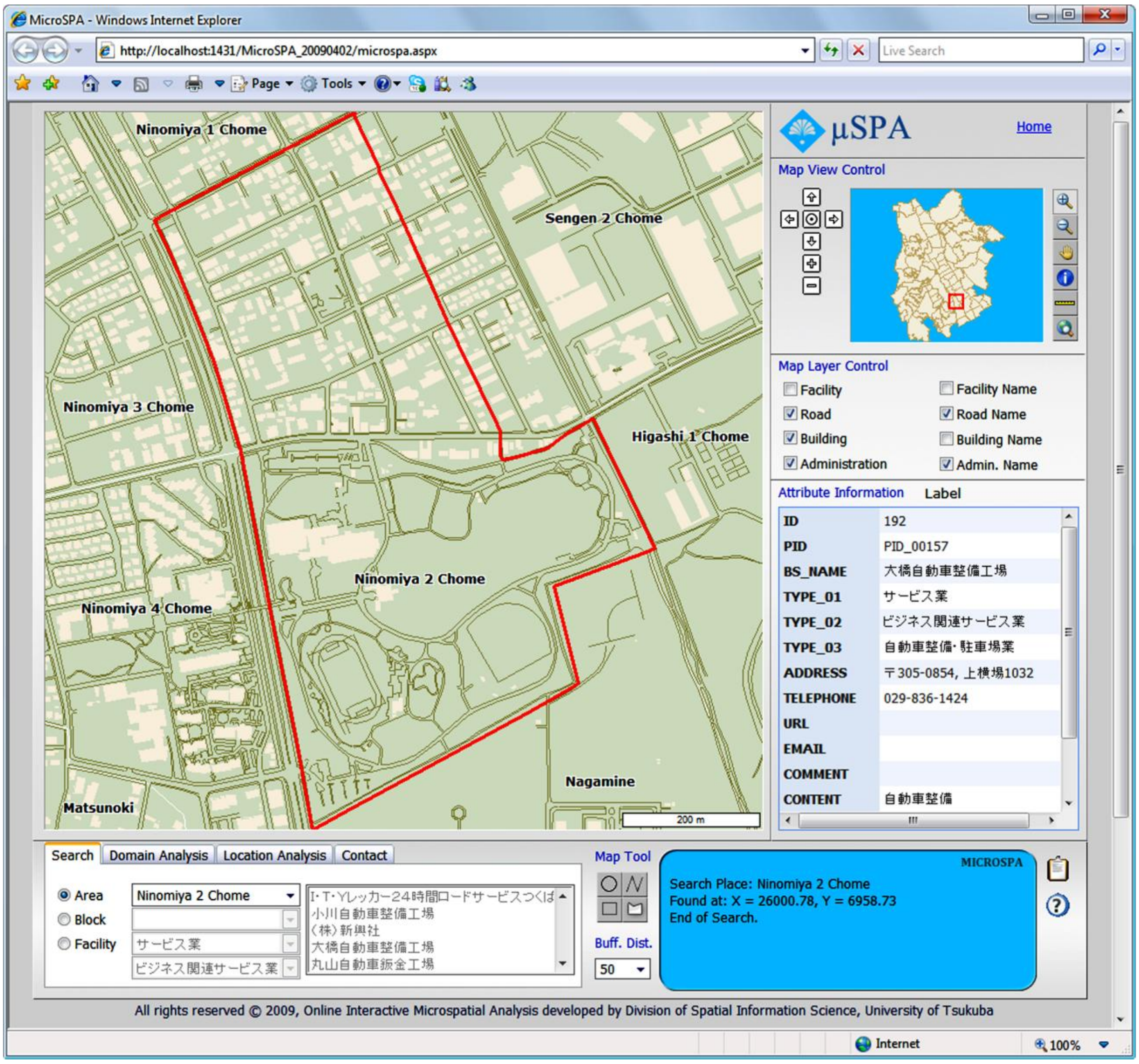


Figure 4-2: Micro-spatial Analysis user interface

4.2 Map Layers and Data Sources

The list of map layers to be used in online micro-spatial analysis is shown in Table 4-1. Road centerline data were acquired from the Geographical Survey Institute (GSI) as line **features** and then real nodes and **dangle** nodes were extracted. The road centerline is one of the important data in connectivity and accessibility analysis. Building data includes estimated building population, building names, block number and building use type. Facility data includes name of the business, main category, sub category, business contents, address and phone number. Figure 4-3 is building footprints and census tracts map layers, Figure 4-4 is road center lines and nodes properties map layers and Figure 4-5 is road outlines on 8 cm orthoimages for landscape visualization purposes.

Road nodes map layer was extracted from road center lines data acquired from Geographical Survey Institute (GSI). Road nodes are classified as real node, which connect each other and dangling node, which does not connect each other (i.e. end point of a road).

Table 4-1: List of map layers and data sources

Layer Order	Layer Name	Description	Feature	Visibility	Source
1	BUILDING	Building footprints with estimated building population attribute information	Polygon	Visible	Zmap-TOWNII
2	FACILITY	Facility locations	Point	Visible	NTT
3	ROAD	Road outlines	Line	Visible	Zmap-TOWNII
4	ROAD_ NODE	Road nodes	Point	Hidden	GSI
5	ADMIN-BND	Administration boundary	Polygon	Visible	Zmap-TOWNII

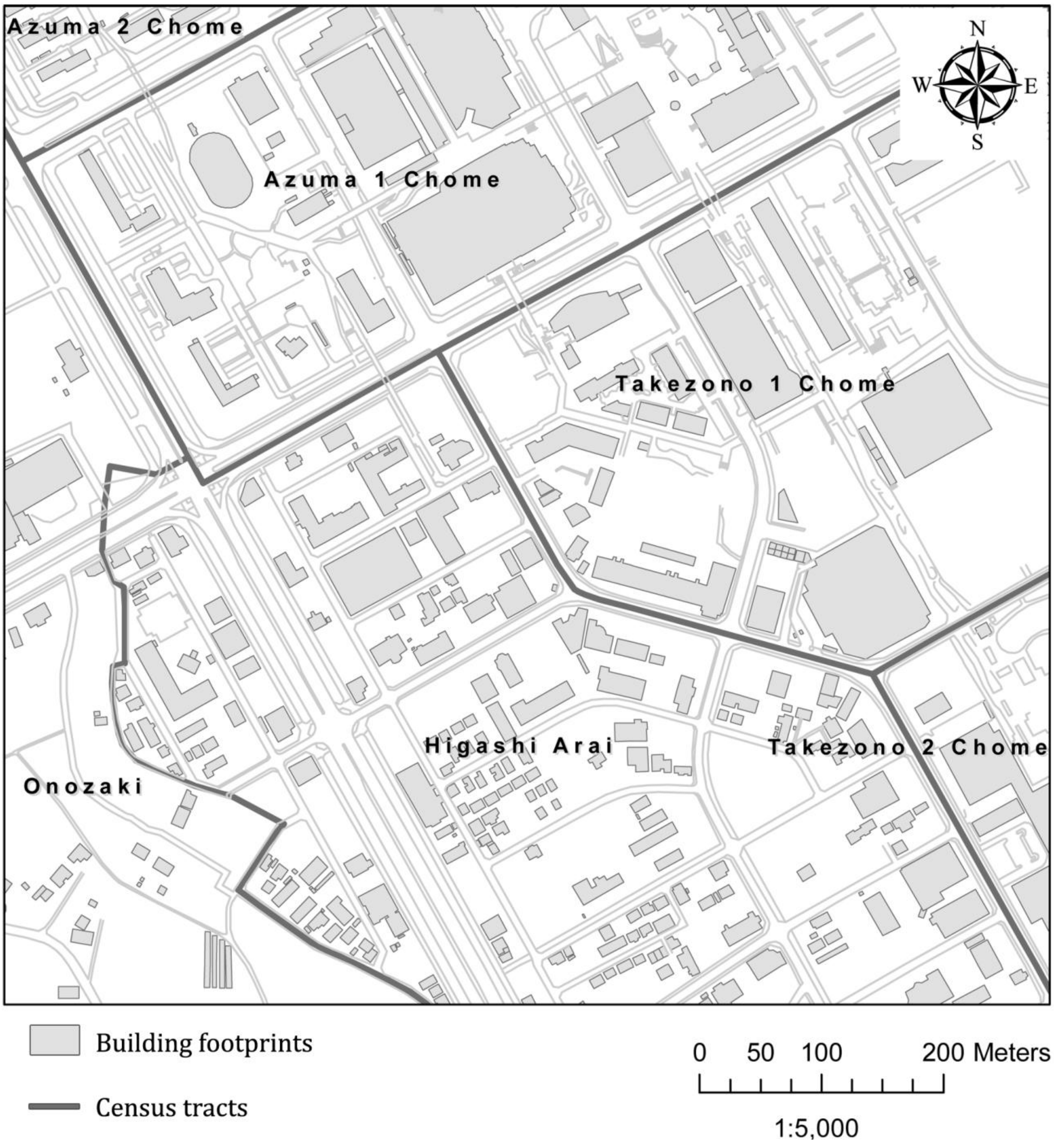


Figure 4-3: Building footprints and census tracts map layers

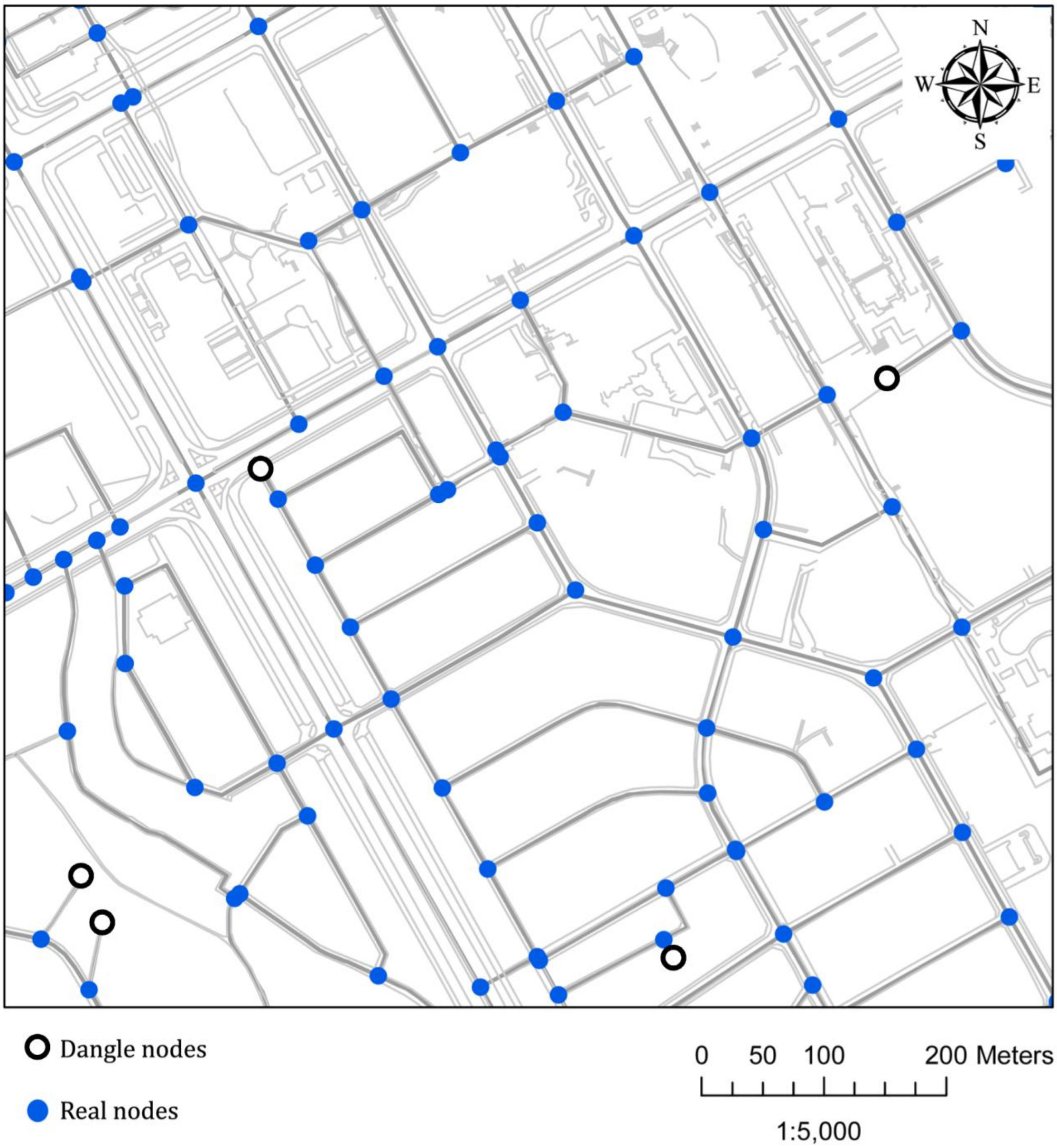


Figure 4-4: Road center lines and road nodes properties map layers

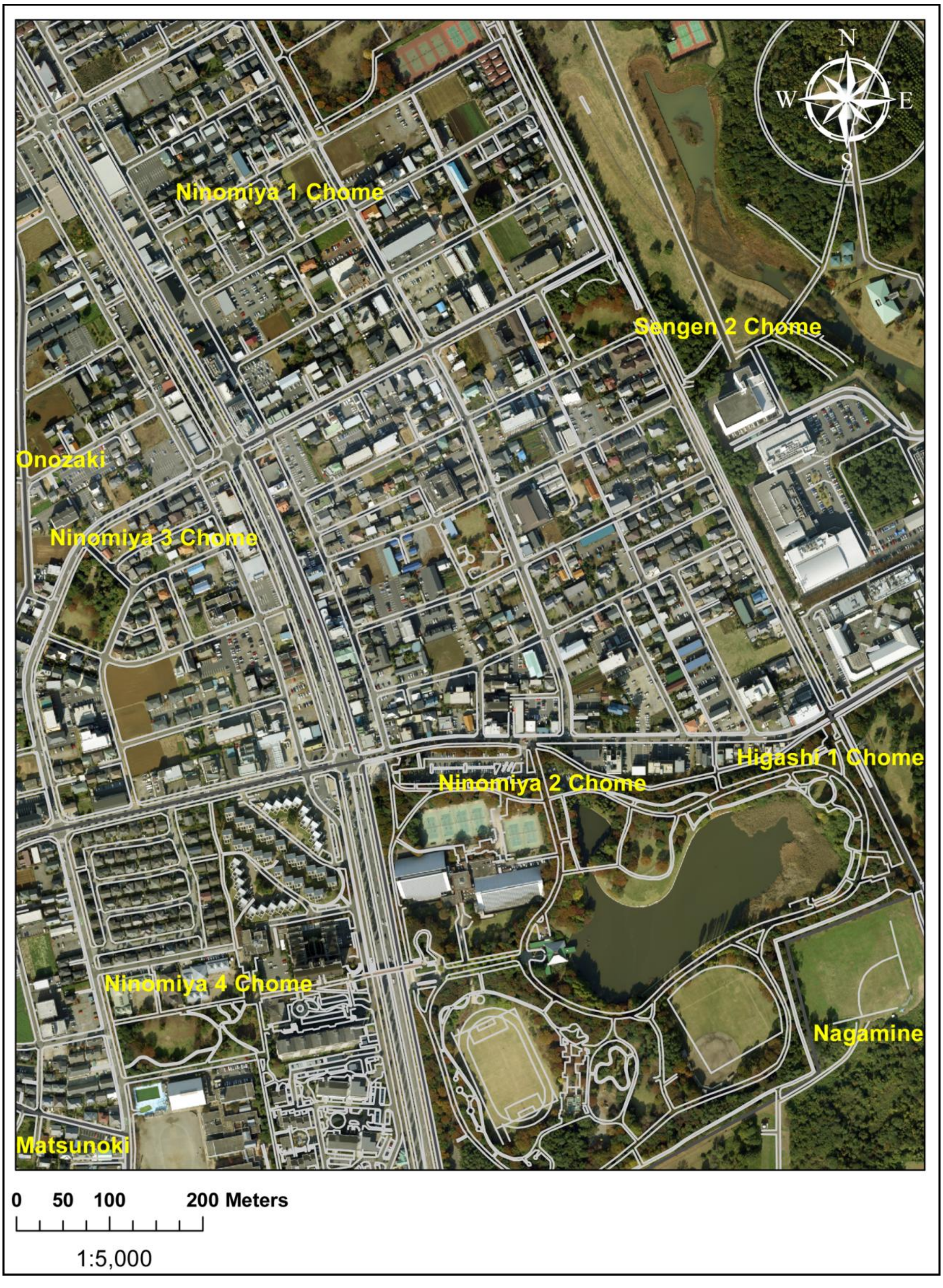


Figure 4-5: Road outlines on 8cm orthoimages

4.3 Measurements

Various quantitative measures have been introduced in the literature to evaluate pedestrian accessibility and measure street connectivity. The distance between origins and destinations for walking and the total length of streets covering an area have been suggested by some authors (Aultman-Hall *et al.*, 1997) to describe how the character of streets differs at neighborhood and regional levels. Pedestrian Route Directness, which measures the ratio between a chosen pedestrian route distance and the 'crow-fly' distance to a particular destination, has been studied (Hess, 1997; Randall and Baetz, 2001) as an indicator of how accessible a neighborhood is to the pedestrians. Some researchers have chosen to calculate the density and pattern of intersections, average block areas and block face lengths per unit area to capture the degree of network connectivity (Southworth and Owens, 1993; Krizek, 2003; Cervero and Kockelman, 1997; Siksna, 1997). Each of these measures is aimed to explain a (slightly or considerably) different aspect of connectivity pertinent to pedestrian accessibility.

Network connectivity is the minimum number of nodes or links that must fail in order to partition the network (or sub-network) into two or more disjoint networks. The larger the connectivity for a network the better the network is able to cope with failures. In the real world network nodes and links do fail (e.g. roads require maintenance or an accident may block a link or junction). When nodes or links fail the network should continue to function with reduced capacity. Network connectivity is a measure of the resiliency of a network and its ability to continue to support traffic flows despite such problems. Several measures have been developed in recent years that attempt to quantify the somewhat abstract idea of connectivity, generally for the auto mode. In an effort to identify the level of connectivity in the metropolitan area of Portland, Oregon, (Dill, 2003) defines and tests several of these measures. Among the most noted of these measures are:

- the Link-Node Ratio, which is measured by dividing the number of links (segments between nodes) in a study area by the number of nodes (intersections plus cul-de-sac termini);
- the Connected Node Ratio, which is a ratio of the number of street intersections to intersections plus the number of cul-de-sacs, thus capturing the number of connected nodes relative to the total number of nodes;
- Intersection Density, which is simply the number of street intersections per unit of area;

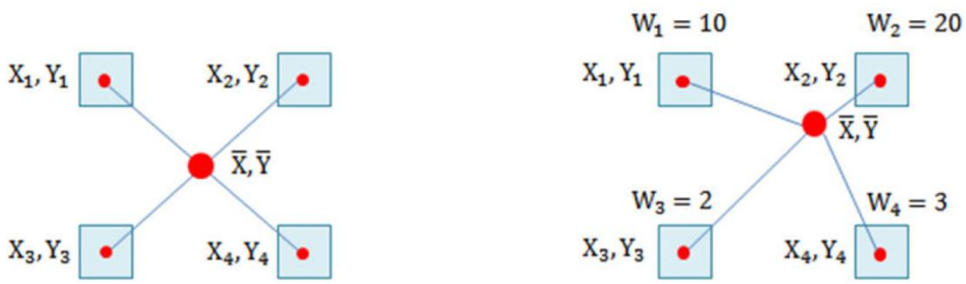
While all of these measures (and other similar ones) provide some method for quantifying connectivity, they fail to take into account the quality of the accommodation provided by the network facilities, an aspect particularly important for the bicycle mode. Of course, there are many other variables to be considered in connectivity and accessibility analysis (e.g. traffic volume, speed limitation, etc.) due to dynamic environment.

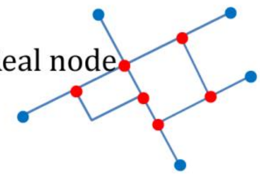
Under the Online Micro-spatial Analysis system, measurements are based on three spatial elements, namely building population, public facility, and transportation network. These three spatial elements are core components in urban systems and interrelated with each other. For example, people need to find the ways to travel from one place to another inside the city. They may need to find the shortest path to buy food and use other utilities. On the other hand, business owners want to know their potential client volume and how they can reach their services. Urban planners or transportation planners need to know the spatial distribution patterns of population and facilities in order to improve the urban transportation system. Each measurement is based on the mean center of three spatial elements such as Weighted Population Mean Center (W-PMC), Weighted Facility Mean Center (W-FMC), Connectivity Mean Center (CMC), Population Index (PIX), Facility Index (FIX), and Connectivity Index (CIX).

4.3.1 Measurement of Mean or Weighted Mean Centers

The mean center is the average x and y coordinate of all the **features** in the study area. It's useful for tracking changes in the distribution or for comparing the distributions of different types of features. Mean center can be used in crime analyst might want to see if the mean center for burglaries shifts if evaluating daytime versus nighttime incidents. This can help police departments better allocate resources, a wildlife biologist can calculate the mean center of elk observations within a park over several years to see where elk congregate in summer and winter to provide better information to park visitors. Providing accessibility to the transport network for resident is vitally important. Weighted mean center is use of one of the numerical attribute fields as a weighted factor such as building population or number of shops per point attribute field (Figure 4-6).

Mean Center | **Weighted Mean Center**





$$\overline{X} = \frac{\sum_i X_i}{n} \; ; \; \overline{Y} = \frac{\sum_i Y_i}{n}$$

$$\overline{X} = \frac{\sum_i \mathbf{W}_i X_i}{\sum_i \mathbf{W}_i} \; ; \; \overline{Y} = \frac{\sum_i \mathbf{W}_i Y_i}{\sum_i \mathbf{W}_i}$$

Weighted Mean Center
- W. Population mean center (W-PMC)
- W. Facility mean center (W-FMC)
- Connectivity mean center (CMC)

Figure 4-6: Measurement of Mean Center and Weighted Mean Center (For example: Weighted Population Mean Center (W-PMC), Weighted Facility Mean Center (W-FMC) and Connectivity Mean Center (CMC)

The Connectivity Mean Center (CMC) was calculated based on X and Y coordinates of road nodes (both real and dangling nodes).

$\overline{X} = \frac{\sum_i X_i}{n}$; $\overline{Y} = \frac{\sum_i Y_i}{n}$ Mean Center .... (4-1)

Where:

X Longitude

Y Latitude

n Number of points

The calculation of Weighted Facility Mean Center (W-FMC) can be selected from specific business type such as a noodle shop or convenience store only. The calculation of Weighted Population Mean Center (W-PMC) and Weighted Facility Mean Center (W-FMC) are as follows:

$\overline{X} = \frac{\sum_i W_i X_i}{\sum_i W_i}$; $\overline{Y} = \frac{\sum_i W_i Y_i}{\sum_i W_i}$ Weighted Mean Center .... (4-2)

Where:

X Longitude

Y Latitude

W Weighted factor (e.g. Building population in the case of Weighted Population Mean Center (W-PMC) and number of shops in the case of Weighted Facility Mean Center (W-FMC))

n Number of points

4.3.2 Measurement of Indices

Ever expanding of urban areas requires effective planning. Designing and maintaining of the cities is important for urban planners to improve the society and community. GIS provides planners, surveyors, and engineers with the tools they need to design and map their neighborhoods and cities.

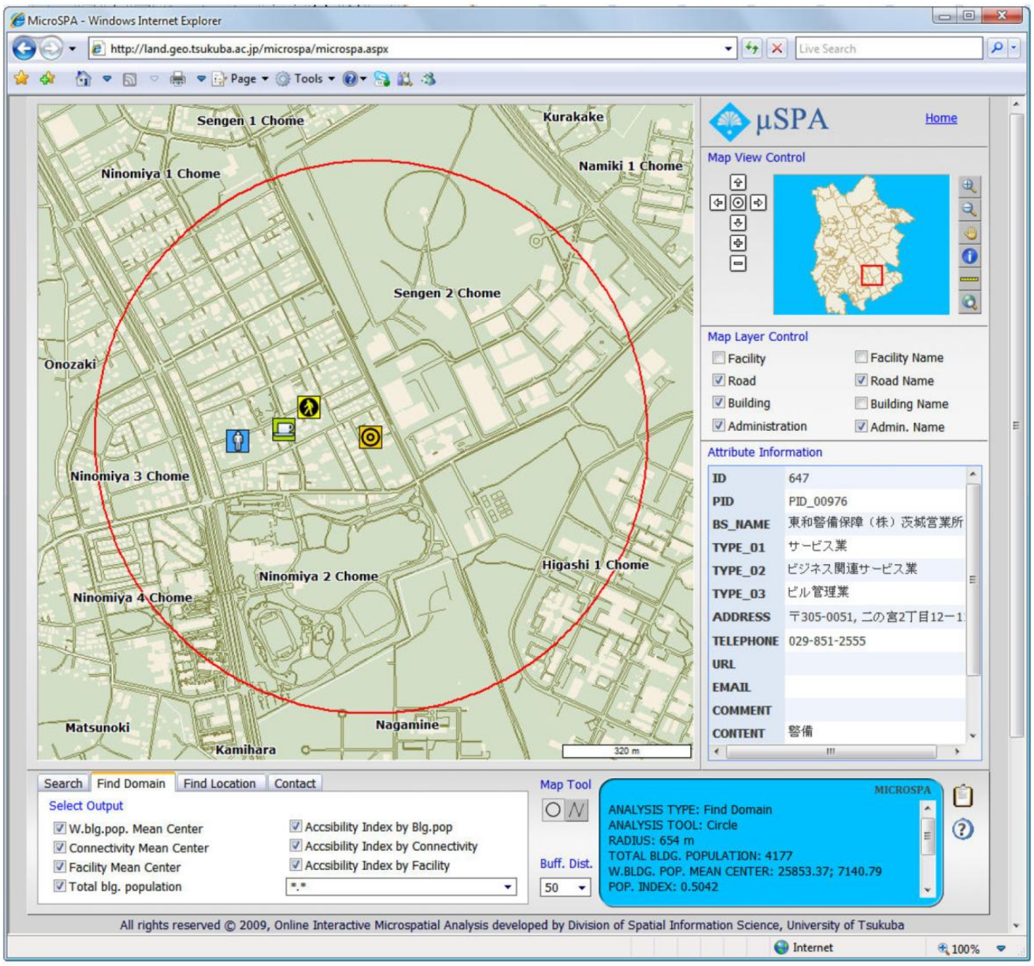
In order to evaluate the availability and quality of basic services for spatial decision makers, this study also measured the Population Index (PIX), Facility Index (FIX) and Connectivity Index (CIX). As generally defined, accessibility reflects the ease of reaching needed or desired activities and thus reflects characteristics of both the land-use system (where activities are located) and the transportation system (how the locations of activities are linked) (Handy and Clifton, 2001).

There are several approaches to measuring accessibility index or degree ranging from simple distance measurement to time and cost integrated measurements. Extensive academic literature on accessibility measures suggests many ways of defining and measuring accessibility, although examples of the actual use of accessibility measures in planning are relatively scarce (Handy and Clifton, 2001).

To quantify the abstract idea of connectivity, Dill (2003) defines and makes several measurements such as link-node ratio, which is measured by dividing the number of links (segments between nodes) in a study area by the number of nodes (intersections plus cul-de-sac termini); the connected node ratio, which is a ratio of the number of street intersections to intersections plus the number of cul-de-sacs (dangling or end node), thus capturing the number of connected nodes relative to the total number of nodes; and intersection density, which is simply the number of street intersections per unit of area. Accessibility is an important concept for urban planners because it reflects the possibilities for activities, such as working or shopping, available to residents of a neighborhood, a city, or a metropolitan area.

In order to measure accessibility index or degree, first find the average distance of all points and divided by circle radius. The circle radius is how far users are willing to travel or size of the service area. Calculation of Facility Index (FIX) can be selected by specific business type such as noodle shop or convenient store only (Figure 4-7).

The index value is between 0 and 1. The interpretation of this value is depends on user's application. For example the lower Facility Index (FIX) is favorable for local residents and nearest competitor for business owners if the facility is same type of business. Local residents can define various location and travel distance to evaluate their potential living place with facilities and connectivity indices. Business owners can assess their potential business location with population and connectivity indices. Urban planners can make future actions based on population and facility indices. Transportation planners can evaluate each bus station or railway station and their serving population by specific **buffered** distance. Indices are only available in "Circle Tool" under find domain analysis mode.



Indices Measurement

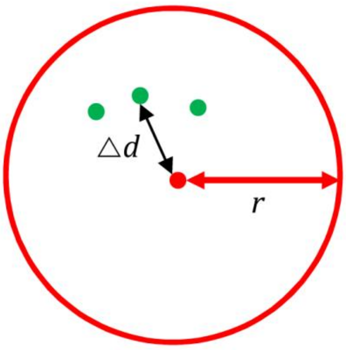
- Population Index (PIX)
- Facility Index (FIX)
- Connectivity Index (CIX)

Weighted Population Mean Center (W-PMC)

Weighted Facility Mean Center (W-FMC)

Connectivity mean center (CMC)

User defined point



Index = △d / r

△d = Average distance of all points

r= Circle radius

Value = 0 ~ 1

Figure 4-7: Measurement of indices

4.4 Map Tools and Analysis Domains

Table 4-2 shows the summary of map tools and analysis domains in Online Micro-spatial Analysis. The Analysis mode can be divided into two groups, Find Domain and Find Location. Circle and Line Tools can be used in Find Domain Analysis mode while Polygon and Rectangle Tools can be used in Find Location Analysis mode (Figure 4-8). Circle Tool draws the circle with user defined radius distance. User can define desire measurement parameters for output display. Under the facility measurement category, user can define desire facility category such as restaurant or grocery shops. Line Tool can measure only building population with specific buffered distance. Indices are not available for Rectangle Tool and Polygon Tool.

With “Circle Tool” (Figure 4-9), user can define the radius size which is how far they are willing to travel or size of the service area in the case of business analysis. Circle tool is useful for disaster planning in the case of earthquake, Frequency Modulation (FM) radio site selection, mobile network planning and so on. With “Line Tool” (Figure 4-10), a user can define a buffer distance from the line and calculate total building population in the case of traffic noise impact studies or local community bus route planning or disaster management. With “Polygon Tool” (Figure 4-11), user can define planning area or service area in irregular shape. This tool can be used for local community center site selection, public facility site sitting and so on.

Table 4-2: Summary of available map tools and analysis domains

Map Tool	Measurement Parameters							Analysis Domain
	TBP	W-PMC	W-FMC	CMC	PIX	FIX	CIX	
Circle	☑	☑	☑	☑	☑	☑	☑	Find Domain
Line	☑	☒	☒	☒	☒	☒	☒	
Polygon	☑	☑	☑	☑	☒	☒	☒	Find Location
Rectangle	☑	☑	☑	☑	☒	☒	☒	

☑ = Available

☒ = Not available

TBP = Total Building Population

W-PMC = Weighted Population Mean Center

W-FMC = Weighted Facility Mean Center

CMC = Connectivity Mean Center

PIX = Population Index

FIX = Facility Index

CIX = Connectivity Index

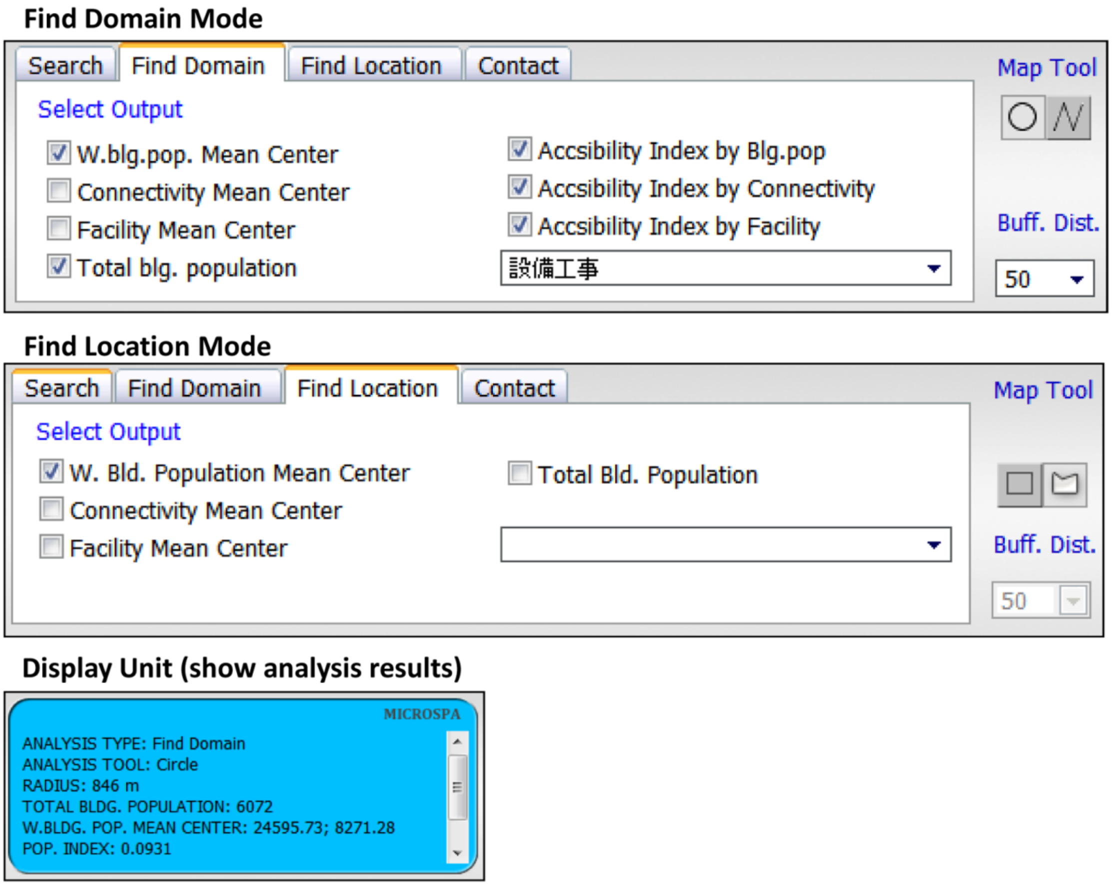


Figure 4-8: MicroSPA analysis panels and results display

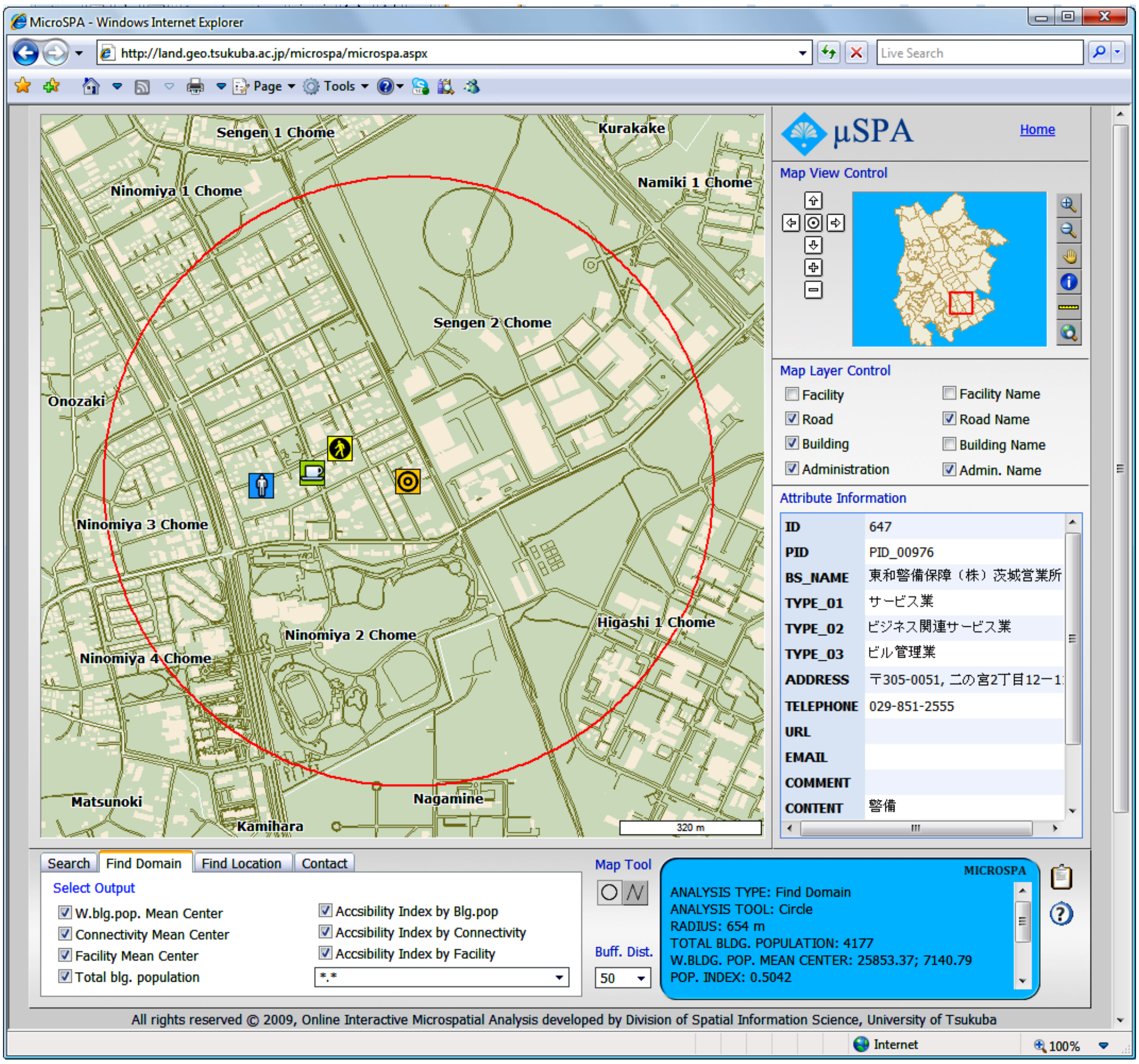


Figure 4-9: Example of "Circle Tool" for business site selection

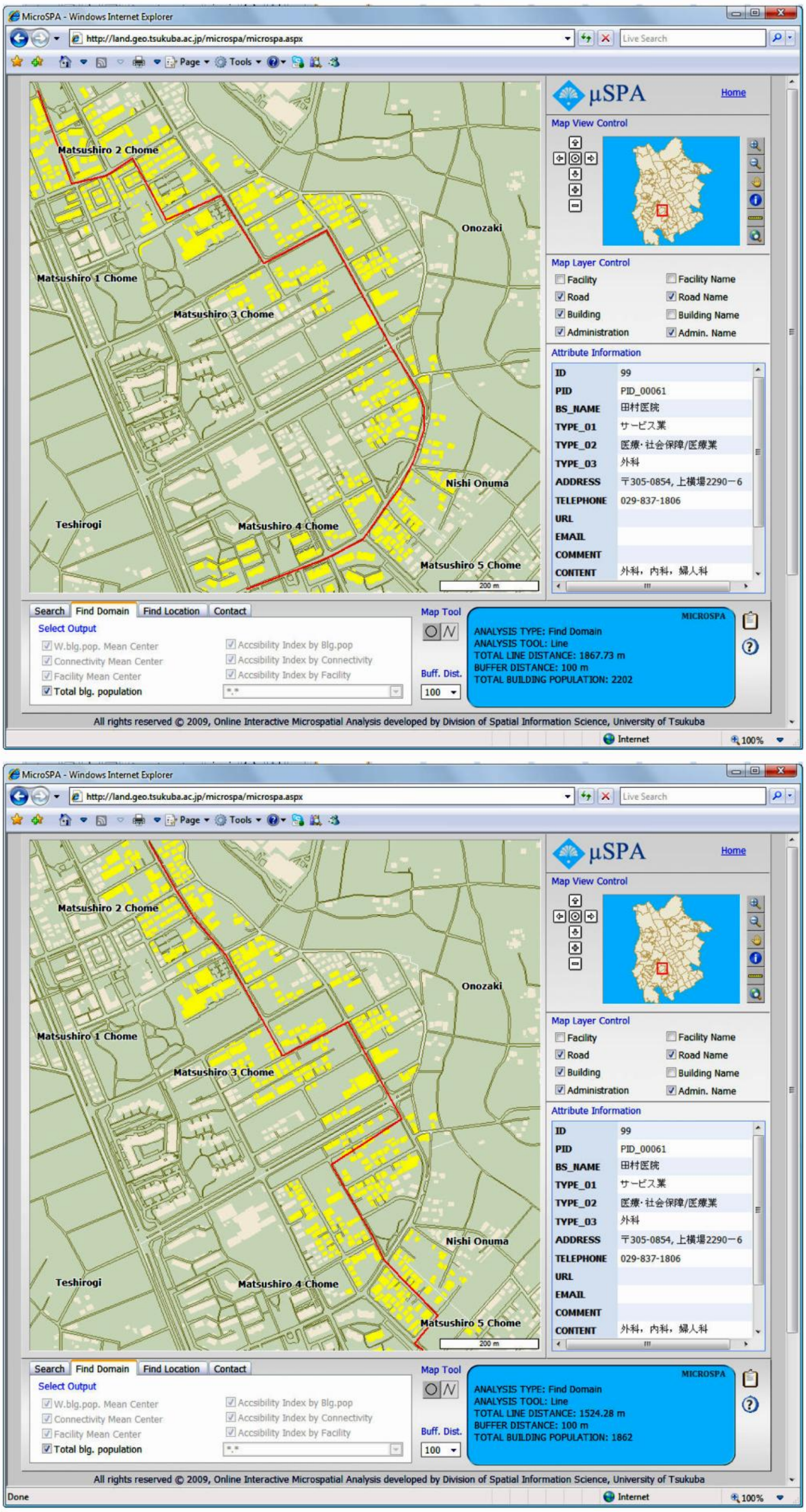


Figure 4-10: Example of “Line Tool” for local community bus route planning (Find shortest path with larger population)

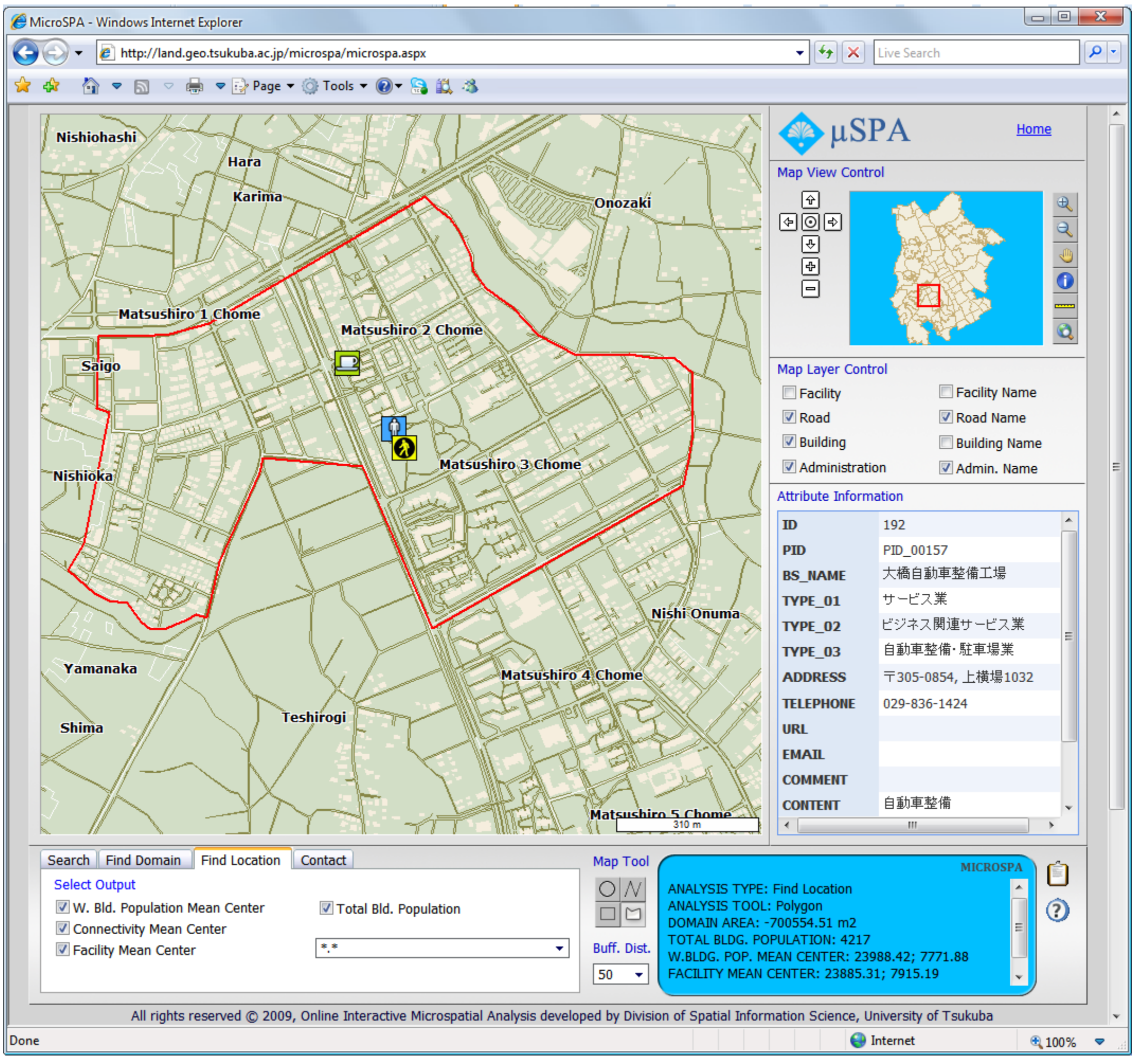


Figure 4-11: Example of "Polygon Tool" for local community center allocation

4.5 Urban Structure and Pattern Analysis

The result patterns will be either dispersed or clustered. Dispersed pattern can interpret as mean center of each elements (i.e., Building Population, Road Nodes and Facility Locations) are away from the user defined point and clustered pattern can interpret as mean center of each elements are closer to the user defined point. The analysis results can vary different urban structures and different buffer distances.

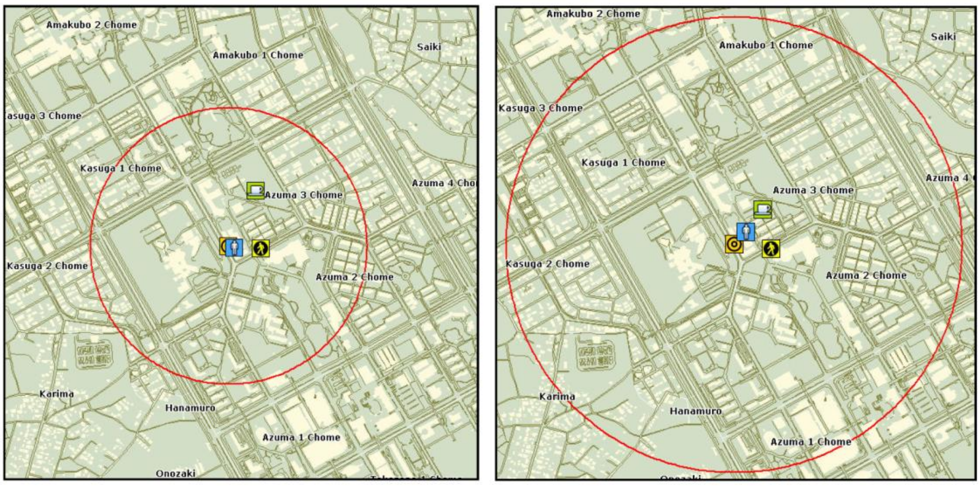
Same buffer distance in different urban structures: Figure 4-12 (Left) shows the dispersed pattern which can interpret as neither good place to live or establish a business. The index of each elements are also large, such as Population Index = 0.6316, Facility Index = 0.7512 and Connectivity Index = 0.1878. Although Connectivity Index is favorable, this location is very difficult to reach all facility locations and also difficult to reach average people to that place. Dispersed pattern is sometime good result for market competition analysis. Clustered pattern are good for potential home buyers. Figure 4-12 (Right) shows the clustered pattern which can interpret as either good place to live or establish a business.

Different buffer distances in same urban structure: The analysis patterns can vary different buffered distances. For example, according to Figure 4-13, Facility Index for this area is 0.4612 within 400m and 0.2074 for 800m buffered zone. According to these indices, people need to travel more than 400m to reach all facilities that not suitable to live for elderly people. Table 4-3 summarizes the potential users and their interpretation based on measurement parameters.

ANALYSIS TYPE: **Find Domain**
ANALYSIS TOOL: **Circle**
RADIUS: **400 m**
TOTAL BLDG. POPULATION: **198**
W.BLDG. POP. MEAN CENTER: **23428.27; 7329.08**
POP. INDEX: **0.6316**
FACILITY X, Y MEAN: **23217.82; 6897.17**
FACILITY INDEX: **0.7512**
CONNECTIVITY X,Y MEAN: **23482.19; 7034.01**
CONNECTIVITY INDEX: **0.1878**

ANALYSIS TYPE: **Find Domain**
ANALYSIS TOOL: **Circle**
RADIUS: **400 m**
TOTAL BLDG. POPULATION: **2583**
W.BLDG. POP. MEAN CENTER: **27285.79; 6187.75**
POP. INDEX: **0.0671**
FACILITY X,Y MEAN: **27305.32; 6203.82**
FACILITY INDEX: **0.1225**
CONNECTIVITY X,Y MEAN: **27303.36; 6168.38**
CONNECTIVITY INDEX: **0.1317**

Figure 4-12 Same buffer distance and different urban structures
(Left: Dispersed pattern and Right: Clustered pattern)



ANALYSIS TYPE: **Find Domain**
ANALYSIS TOOL: **Circle**
RADIUS: **400 m**
TOTAL BLDG. POPULATION: **1709**
W.BLDG. POP. MEAN CENTER: **24755.71; 9663.33**
POP. INDEX: **0.0449**
FACILITY X,Y MEAN: **24850.58; 9801.67**
FACILITY INDEX: **0.4612**
CONNECTIVITY X,Y MEAN: **24820.48; 9701.66**
CONNECTIVITY INDEX: **0.2348**

ANALYSIS TYPE: **Find Domain**
ANALYSIS TOOL: **Circle**
RADIUS: **800 m**
TOTAL BLDG. POPULATION: **7554**
W.BLDG. POP. MEAN CENTER: **24785.62; 9704.93**
POP. INDEX: **0.0806**
FACILITY X,Y MEAN: **24842.35; 9777.7**
FACILITY INDEX: **0.2074**
CONNECTIVITY X,Y MEAN: **24869.77; 9648.41**
CONNECTIVITY INDEX: **0.1713**

Figure 4-13 Same urban structure and different buffer distances
(Left: Dispersed pattern and Right: Clustered pattern)

Table 4-3: Summary of measurement parameters and interpretations

Parameters	Local Residents	Business Owners	Urban Planners
W-PMC		Find most populated area for business site selection	Find most populated area for public facility planning
W-FMC	Find most facilitated area	Find most facilitated area for business activities or find nearest competitors in the case of same business type	Find most facilitated area conjunction with Connectivity Mean Center (CMC) to identify whether additional roads or streets are needed.
CMC	A good place to access	A good place to reach their shops	Combined with Weighted Population Center (W-PMC) and Weighted Facility Mean Center (W-FMC) to evaluate connectivity status
TBP		To define customer volume by specific geographical regions such as total population by specific buffered distance, specific zone, etc.	Find shortest path with larger building population for bus route planning. Planning for public facility site selection such as parks, stadium, spot clubs, etc.
PIX		Evaluate business site 0 = Most people can access 1 = A few people can access	Evaluate public facility site 0 = Most people can access 1 = A few people can access
FIX	0 = A good place to reach all facilities 1 = Difficult to reach all facilities	For all facilities: 0 = Suitable for business 1= Not suitable for business For selected same business: 0 = Nearest competitors are exist 1=Competitors are away from the site	Evaluate facility sites 0 = Most facilitated area 1 = Less facilitated area
CIX		0 = Good place to access 1 = Poor place to access	0 = Good place to access 1 = Poor place to access
Applications	• Choosing a place to live • Neighborhood evaluation	• Business site selection/evaluation • Customers accessibility analysis "How customers can access my shop?" • Find competitors	• Make decision based on population and facility distributions such as: • Where to construct additional roads? • Need additional facility?

4.6 Development Platform

Formerly an individual would have to buy an expensive software package to use and develop Web based GIS applications. However, now the situation is changing. The development of coding (programming) with graphical user interface like Active Server Page ASP.NET technology introduced by Microsoft Cooperation helps developers to develop web-based programs in timely manners. These programs can run on existing web browsers. Moreover, availability of form-based free or commercial GIS components or Engine allows web-GIS developers to develop powerful GIS analytical functions in timely manners.

The overall system was built on Microsoft ASP.NET with **AJAX** Extension and VDS Technologies (Web Mapping Components for ASP.NET). ASP.NET is a web application framework marketed by Microsoft that programmers can use to build dynamic web sites, web applications, and XML web services. **AJAX** (shorthand for asynchronous JavaScript and XML) is a group of interrelated web development techniques used on the client-side to create interactive web applications. With **AJAX**, web applications can retrieve data from the server asynchronously in the background without interfering with the display and behavior of the existing page. The use of **AJAX** techniques has led to an increase in interactive or dynamic interfaces on web pages. AspMap for .NET from VDS Technologies is a set of high-performance, web mapping components and controls for embedding maps in ASP.NET applications (Web Forms).

Recently, a wide variety of web-based or web-deployed tools have become available, enabling datasets to be analyzed and mapped, including dynamic interaction and drill-down capabilities, without the need for local GIS software installation. These tools include the widespread use of Java applets, Flash-based mapping, **AJAX** and Web 2.0 applications, and interactive Virtual Globe explorers, some of which are described in this Guide. They provide an illustration of the direction that many toolset and service providers are taking.

4.7 Potential Applications

The potential application for quantitative building population data is numerous and will increase the accuracy in various spatial decision-making processes at the micro-scale level. Following examples are use of building population data in various spatial decisions makings and planning processes.

4.7.1 Building Population as Market Competition Analysis

Okabe and Kitamura (1997) described about that competition of market allocation for retail stores in a densely inhabited region becomes very hard. For instance, in the central region of Tokyo (23 Wards; 621K m^2) competing their locations. Market analysis involves customer distributions, competitor shops locations and other socio economic factors.

Although there are several physical factors involved in market analysis, building population data can be used for market competition by applying **Voronoi diagram** which is a special kind of decomposition of a metric space determined by distances to a specified discrete set of objects in the space, e.g., by a discrete set of points (i.e. location s of retail shops). Figure 4-14 analyzed the convenience store locations and their potential residential population. Convenience stores are open twenty-four hours a day, seven days a week, including Sundays and public holidays. Therefore, their business is mainly depends on local resident population.

By understanding details about the location of the business and the surrounding population, we can discover where is the best clients are located and where is the most profitable market areas. Nowadays, business companies are using GIS technology to find new market places, make proper decisions, enhance their customer services and find profit opportunities.

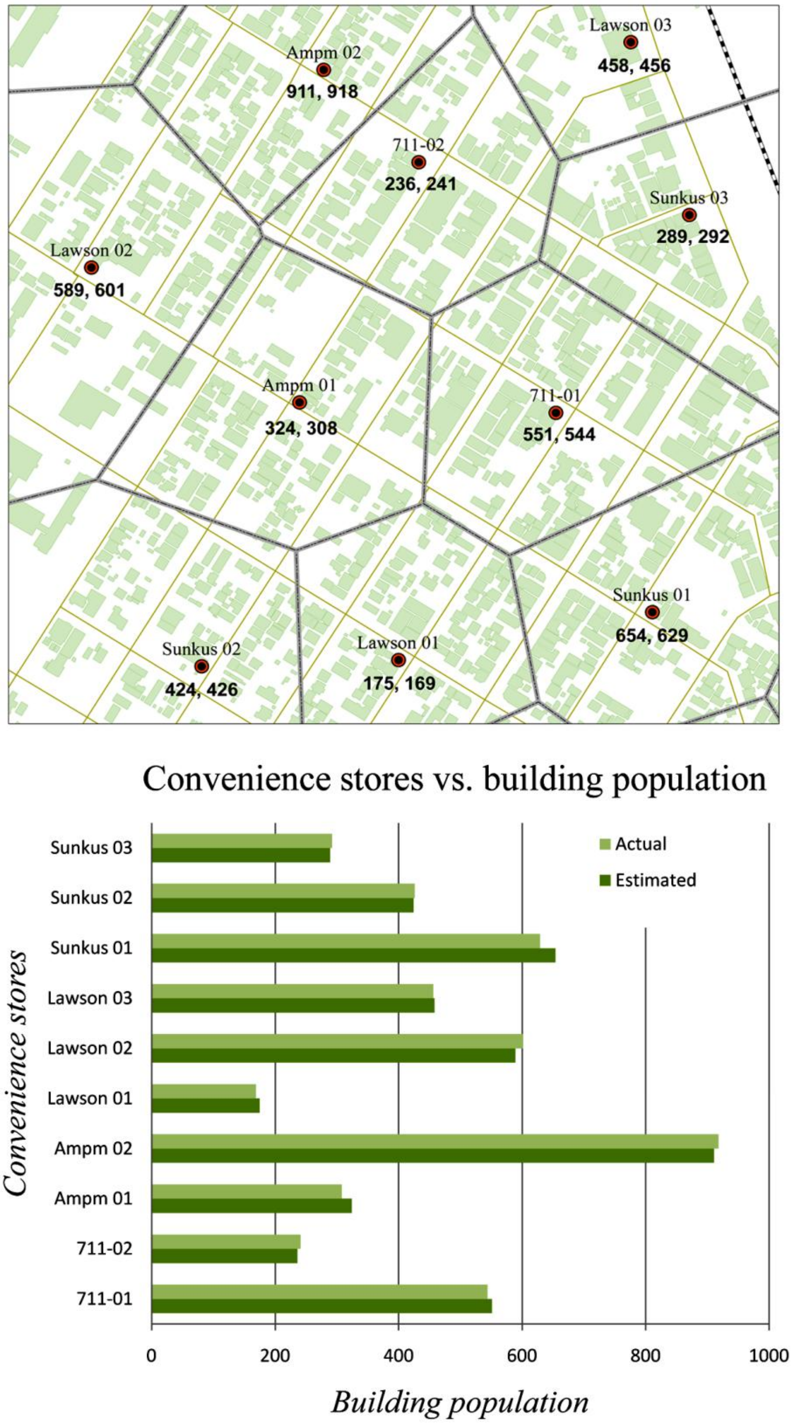


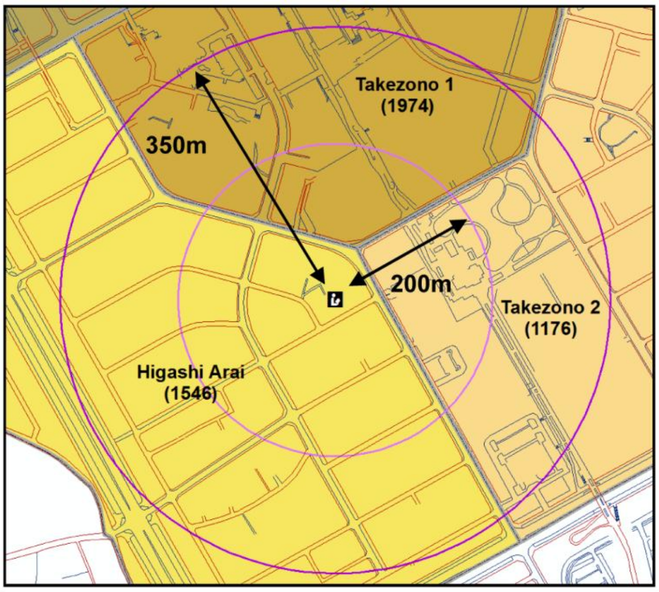
Figure 4-14: Example of building population in market competition analysis

4.7.2 Improved Decision Making

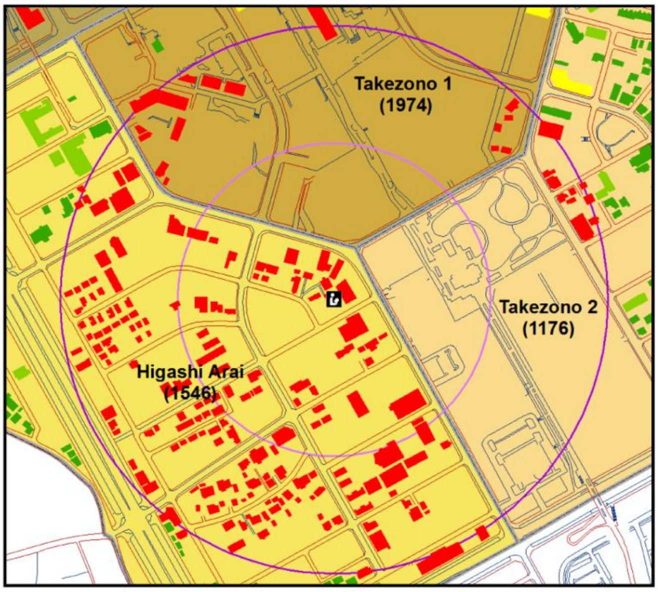
Current population data in GIS limits the many sophisticated decision making processes. Figure 4-15 and Figure 4-16 compare the results between classical approach and using building population approach at micro-scale spatial analysis such as point and line buffering. Population by certain geographical unit is important for disaster and environmental impact studies.

4.7.3 2D Visualization of Building Population Data

Visualization is one of the key components in GIS in terms of mapping the spatial extents and portrait the specific theme known as thematic mapping. Geo-visualization is defined by (MacEachren and Kraak, 2001) as "the integration of visualization in scientific computing, cartography, image analysis, information visualization, exploratory data analysis and GIS, which all together provide theory, methods and tools for visual exploration, analysis, synthesis and presentation of geospatial data". Figure 4-17 shows the visualization of 2 dimensional quantitative map of building population in Tsukuba Central Area.

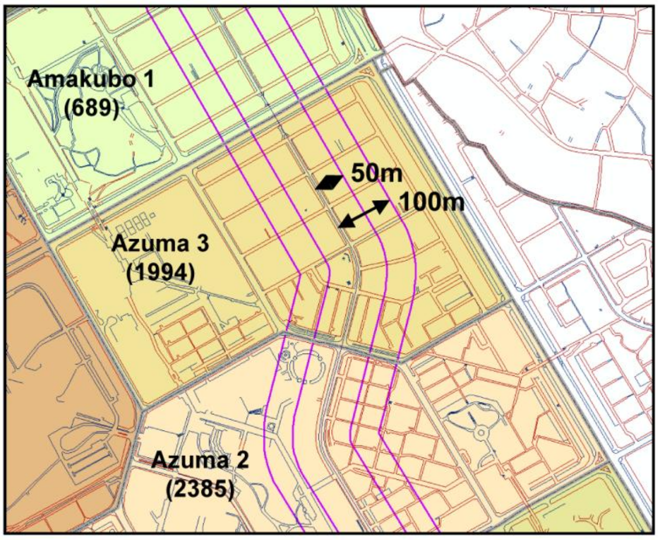


Total population inside 200m : 4696
Total population inside 350m : 4696

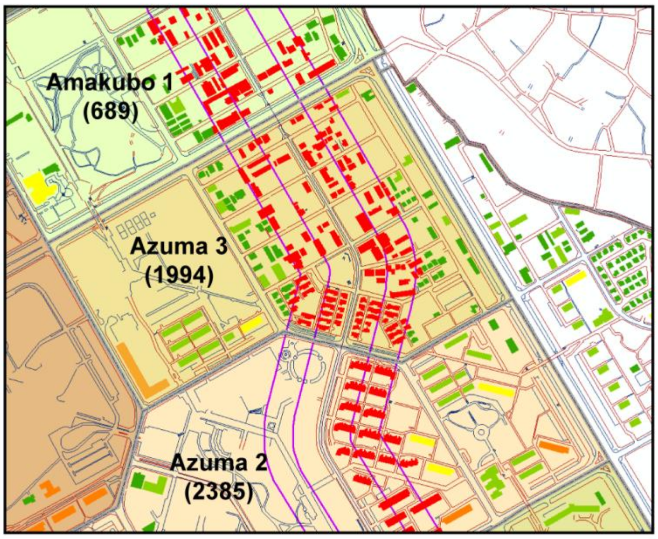


Total population inside 200m : 393
Total population inside 350m : 1443

Figure 4-15: Example of point buffering analysis
(Top: classical approach, Bottom: using building population)



Total population inside 50m : 5068
Total population inside 100m : 5068



Total population inside 50m : 884
Total population inside 100m : 1655

Figure 4-16: Example of line buffering analysis
(Top: classical approach, Bottom: using building population)

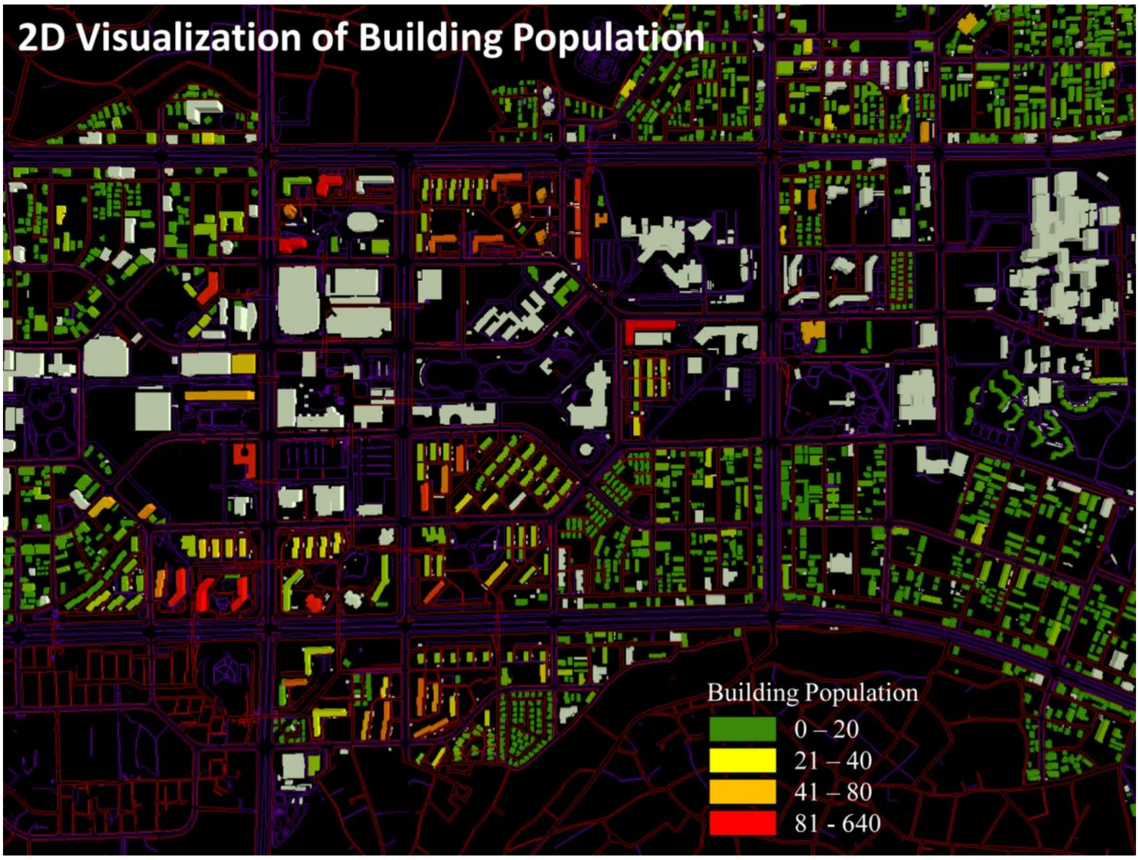


Figure 4-17: 2D Visualization of quantitative building population data

4.7.4 3D Visualization of Building Population Data

The 3D visualization of building population is genuine advantages when it is integrated into the public facility planning process. It shows not only population, but also volume of population. 3D visualization of GIS data is important for community planners, architects, urban designers, and land use planners to visualize the impact of urban design projects and proposed land use and zoning changes or envision the results of smart growth initiatives. 3D visualization models have a variety of applications in geography and urban studies: site location analysis, emergency facilities planning, design review, marketing, etc. While they are generally used to simply visualize the built environment, there are early signs of them being used as 3D interfaces to more sophisticated simulation models. 3D Geo-visualization is closely related to Virtual Reality (Whyte, 2002) and can be applied at different stages in the process of urban planning. Figure 4-18 shows the 3D model of building population of study area.

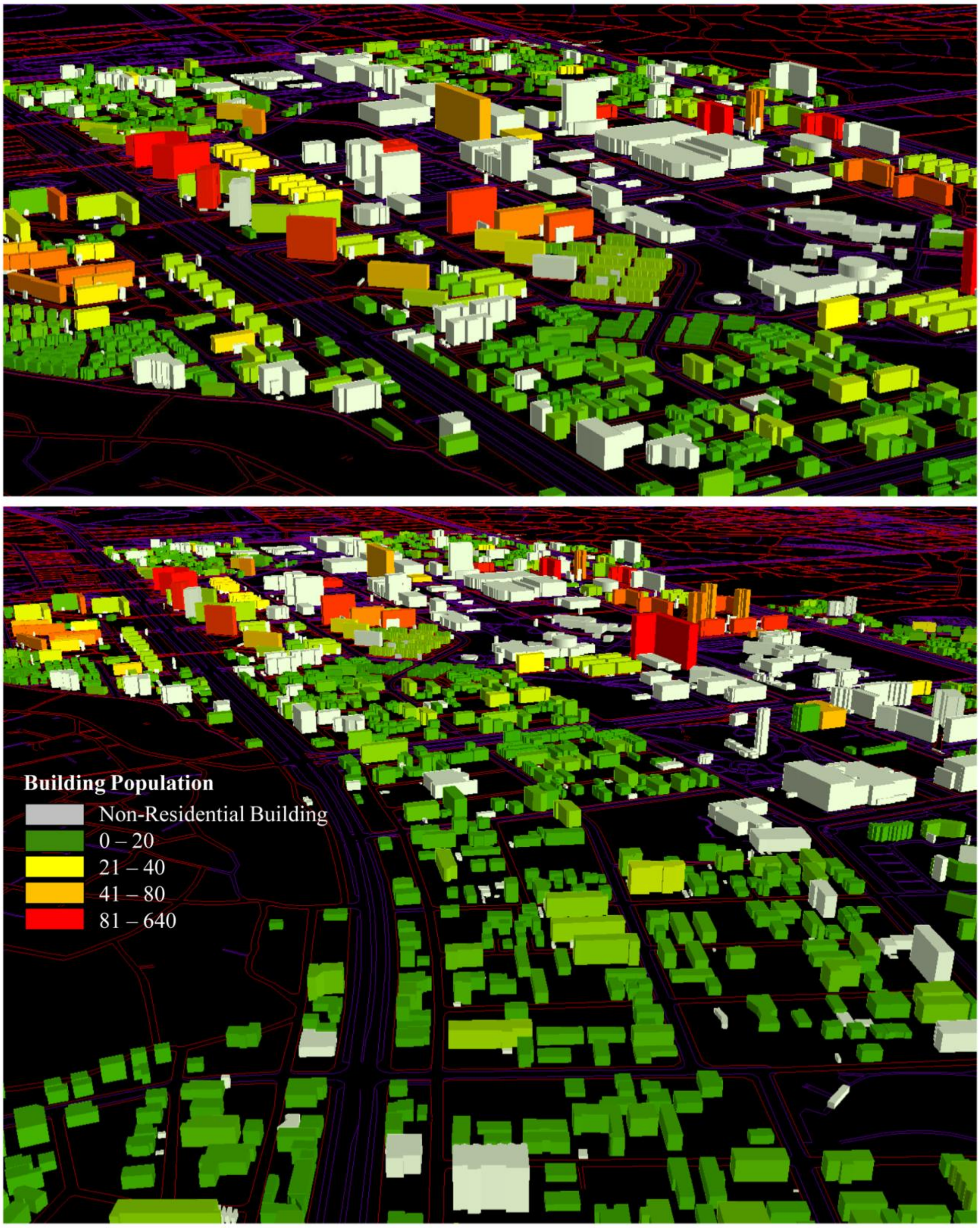


Figure 4-18: 3D Visualization of quantitative building population data

Chapter 5

Conclusion

This research explored and developed the possibilities for extending the spatial analytic capabilities of GIS on the World Wide Web by construction of online micro-spatial analytical functions based on GIS estimated building population by utilizing modern spatial information technologies. Based on proposed building population estimation algorithms, the Volumetric results are closely related to actual building population than Areametric method in terms of visual, statistical and spatial assessments. Volumetric method has achieved reasonable results, confirming model suitability for use in micro-spatial analysis.

This research also developed GIS standalone tool to calculate building population **attribute** field based on census tract and building footprints dataset in ESRI Shape file format for future interested users. This tool can be used for any area where census tract and building footprints are available. Based on data availability, users can choose whether Areametric or Volumetric method. Areametric method does not require number of building floor or building height or building volume information. This is only suitable for use in rural area. However, Volumetric requires either building floor or building height or building volume. In order to improve the accuracy this tool also allows to filter minimum footprint area and building use types. Estimation of day and night time population is also possible to calculate by using all census population and specific building use types such as residential, commercial and educational building use types.

By utilizing modern spatial information data acquisition systems such as LIDAR, the building height and volume can be measured and used to estimate building population. Under the LIDAR data processing, conversion of LIDAR point data to **raster** surface model is time consuming in traditional interpolation methods like IDW, SPLINE and Kriging and require additional mosaicking task for multiple scenes. However, according to this study, TIN process is faster than traditional methods and does not require additional mosaicking task. This is suitable for multiple scenes LIDAR point cloud data processing for large area.

Moreover, iTownpage data from Nippon Telegraph & Telephone Corp. (NTT) plays critical role in identification of building use type and calculation of adjustment factor for mixed building use type. This was improved the accuracy in estimated results.

According to this study, estimation of building population from LIDAR generated Digital Volume Model (DVM) approach and building floors approach are almost identical. However, in order to improve the accuracy in estimation process, we still need LIDAR data for building height measurement in order to separate residential building vs. non-residential building. According to correlation analysis between building floors and estimated building population, most population are accumulated lower building floors regions which show most people are living in low-rise building in this study area. Only a few people are living high-rise floor building.

Based on the development of online micro-spatial analytical functions, the population results were dramatically improved in various GIS analyses compare to classical approaches. Moreover, by providing different map analysis tools (i.e. Line, Circle, Polygon and Rectangle), the GIS users can analyze population by certain geographical units such as planning zone, disaster zone, market area, specific buffer distances and so on. This can perform more realistic spatial decision making processes for advanced GIS users in various applications. The estimation of day time population is also possible by using overall census tracts population and all building use type. This is suitable for such area like Tsukuba city where the most people are living and working inside the city.

By integrating LIDAR derived Digital Volume Model (DVM) with **geodemographic** data is a key benefit for urban and city planners, improving their decision making processes in a timely manner. Although LIDAR data is expensive and advanced technology, number of floor **attribute** information still can be used for estimation of building population. This research also considered the population data as a three-dimensional space unlike previous studies such as segregation based on land use/land cover or cadastral data.

The estimated or quantitative mapping of building population is essential for micro-spatial analysis especially in terms of emergency management. Effective disaster preparedness requires quantitative spatial distribution patterns of population in order to position emergency response centers and prepare food and shelter in the event of disaster. Building population data is also required for improved accuracy in cost estimation of food and shelter for emergency preparedness and other humanitarian assistance. City and urban planners need to know how many local residents will benefit from newly constructed public facilities such as bus centers, railway stations and hospitals. Hydrologists require an estimate on the number of people living on a floodplain. Potential business owners can define their business location and perform consumer analysis. Quantitative building population data can be used as a weighted factor in spatial statistical analysis such as for determining population mean center and standard distance. This is important for decision making related to population such as in selecting a voting site or construction a new public facility. Specific sub-population distribution for urban areas in order to develop an improved "denominator," which would enable the calculation of more correct rates in GIS analyses involving public health, crime, and urban environmental planning. Additionally, the knowledge of accurate population distribution can be extremely valuable in the sphere of urban planning. The understanding of the locational characteristics of target populations would allow for more equitable resource allocation in areas such as community infrastructure development, provision of open space and recreational opportunities, transportation access, and necessary environmental facilities (Maantay *et al*., 2007).

To empower the analysis function of Web GIS in the future, the trend will be to create standard individual function modules or components. These standard function components enable developers to produce 'plug-and-play' type GIS analysis tools. These function components are assembled in the client's web browser in runtime when they are needed. This componentized or modularized approach is consistent with the direction of the world of computing, which is moving toward component-ware and network-based computing. This approach demands a standard specification to ensure interoperability (Peng, 1999).

Through advancing in spatial web technologies, it is possible to automate population data. For example, by utilizing monthly updated city office resident registration data, Online Micro-spatial Analysis system can be configured as middleware between city office and micro-spatial information users. Under the MicroSPA system, use of Tsukuba City monthly updated resident registrations information can feed into to building population estimation model via File Transfer Protocol (**FTP**). This process will carry out by automatically and refresh the building population information at monthly intervals.

By utilizing modern spatial web sensor network technologies, real-time monitoring of population flow and movement can be measured, for example, utilizing train ticket boarding pass machine as population movement sensors and monitor the population flow. This is especially possible in Tokyo metropolitan areas where most people are using well connected sub-way systems.

As a summary, this study has three significant approaches in terms of methodology, data processing and application development. As a methodology, I proposed two building population estimation methods (Areametric and Volumetric methods) by applying GI Science theory and practice. Most studies on population estimation by remote sensing and GIS is only consider population as a two dimensional space, for example, using Land/Use Cover data, high resolution optical images and Cadastral data which do not show actual population volume.

This may introduce some errors in population data analysis at high-rise building area.

As a data processing, using LIDAR data and integrated with other tabular data (i.e. NTT iTownpage) to estimate the building population is absent in GIS researches. Generating of Digital Height Model from LIDAR point cloud data is time consuming and requires additional computational powers in order to process thousands of points per scene. In this study, Triangulated Irregular Network TIN process was used to generate TIN-DSM layer and then converted into raster layer. This process dramatically reduced the processing time and eliminated the mosaicking task which requires in classical surface generation process like IDW, SPLINE and Kriging interpolations.

Until now, spatial analysis by utilizing building population data is absent in GIS due to unavailable population data at building level due to privacy concerns. This study demonstrates the possibilities of future population data analysis scenario by providing common GIS analysis tools to make decision support for local residents, business owners and urban planners based on three spatial elements named as building population, facility locations and road connectivity nodes. The analysis results will depend on various urban structures and buffered distances. Nowadays, Internet is part of society in terms of searching information on the Web, Shopping, Chatting, Mailing and other entertainment activities. Moreover, Web GIS is part of our daily life by means of searching a place to go, finding a facility, planning a driving direction, making interactive spatial analysis and so on.

Acknowledgements

I would like to gratefully acknowledge the enthusiastic supervision of Professor Yuji MURAYAMA who has supported me throughout my thesis with his patience, knowledge and experience. I would also like to include my gratitude to Professors Akira TABAYASHI, Akira TEZUKA, Kiyomi YAMASHITA, and Assistant Professor Takehiro MORIMOTO for their extremely valuable comments and useful suggestions. Furthermore I am deeply grateful to my Assistant/Associate Professors and colleagues from Division of Spatial Information Science, Human Geography and Regional Geography departments where I can expand my Remote Sensing and GIS knowledge to human geography, tourism and economic geography.

References

Anselin, L., Dodson, R. and Hudak, S., 1993, Linking GIS and spatial data analysis in practice, *Geographical Systems* **1**(1):3-23.

Anselin, L. and Getis, A., 1992, Spatial statistical analysis and geographic information systems, *The Annals of Regional Science* **26**(1):19-33.

Aultman-Hall, L., Roorda, M. and Baetz, B. W., 1997, Using GIS for evaluation of neighborhood pedestrian accessibility, *Journal of Urban Planning and Development* **123**(1):10-17.

Bailey, T. C. and Gatrell, A. C., 1995, *Interactive spatial data analysis*, Longman, Harlow.

Baller, R. D., Anselin, L., Messner, S. F., Deane, G. and Hawkins, D. F., 2001, Structural covariates of US county homicide rates: Incorporating spatial effects, *Criminology* **39**(3):561-590.

Baudot, Y., 2001, Geographical analysis of the population of fast-growing cities in the third world, in: *Remote Sensing and Urban Analysis* (J. P. Donnay, Barnsley, M. J. and Longley, P. A., ed.), Taylor and Francis, London, pp. 225-241.

Ben-Moshe, E., 1997, Integration of a national GIS project within the planning and implementation of a population census, in: *Proceedings of the Euro-Mediterranean Workshop on New Technologies for the 2000 Census Round: Sharing the Israeli Experience from the 1995 Census.*

Bielecka, E., 2005, A dasymetric population density map of Poland, in: *Proceedings of the International Cartographic Conference, July*, A Coruna, Spain, pp. 9-15.

Boston, A. N. and Stockwell, D., 1995, Interactive species distribution reporting, mapping and modelling using the World Wide Web, *Computer Networks and ISDN Systems* **28**(1-2):231-238.

Bowers, K. and Hirschfield, A., 1999, Exploring links between crime and disadvantage in North-West England: an analysis using geographical information systems, *International Journal of Geographical Information Science* **13**(2):159-184.

Browning, C. R., Cagney, K. A. and Wen, M., 2003, Explaining variation in health status across space and time: Implications for racial and ethnic disparities in self-rated health, *Social Science & Medicine* **57**(7):1221-1235.

Burrough, P. A., McDonnell, R. A., Burrough, P. A. and McDonnell, R., 1998, *Principles of geographical information systems*, Oxford University Press, Oxford.

Cervero, R. and Kockelman, K., 1997, Travel demand and the 3Ds: density, diversity, and design, *Transportation Research Part D* **2**(3):199-219.

Chang, N. B., Lin, Y. T. and Chang, Y. H., 1998, A client/server computer framework for solid waste management decision analysis in Taiwan, *Journal of Hazardous Materials* **58**(1-3):15-31.

Chang, Y. C. and Chang, N. B., 2002, The design of a web-based decision support system for the sustainable management of an urban river system, *Water Science & Technology* **46**(6):131-139.

Chang, Y. C., Chang, N. B. and Ma, G. D., 2001, Internet Web-based information system for scrap vehicle disposal in Taiwan, *Environmental Modeling and Assessment* **6**(4):237-248.

Chen, H., Smith, T. R., Larsgaard, M. L., Hill, L. L. and Ramsey, M., 1997, A geographic knowledge representation system for multimedia geospatial retrieval and analysis, *International Journal on Digital Libraries* **1**(2):132-152.

Chomitz, K. M. and Gray, D. A., 1996, Roads, land use, and deforestation: a spatial model applied to Belize, *The World Bank Economic Review* **10**(3):487.

Coleman, D. J., McLaughlin, J. D., 1997, Information access and network usage in the emerging spatial information marketplace, *URISA Journal* **9**(1):8-19.

Cook, L., 2004, The quality and qualities of population statistics, and the place of the census, *Area* **36**(2):111-123.

Dent, B. D., 1993, *Cartography: Thematic map design*, William C. Brown, Oxford.

Dill, J. and Portland, O. R., 2003, Measuring network connectivity for bicycling and walking, in: *Joint Congress of ACSP-AESOP, Leuven, Belgium*.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C. and Worley, B. A., 2000, LandScan: a global population database for estimating populations at risk, *Photogrammetric Engineering and Remote Sensing* **66**(7):849-857.

Dorling, D. and Fairbairn, D., 1997, *Mapping: Ways of representing the world*, Longman, London.

Duncan, C., Jones, K. and Moon, G., 1993, Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain, *Social Science and Medicine* **37**(6):725-733.

Eicher, C. L. and Brewer, C. A., 2001, Dasymetric mapping and areal interpolation: implementation and evaluation, *Cartography and Geographic Information Science* **28**(2):125-138.

Fischer, M. M., Scholten, H. J., and Unwin, D., 1996, *Spatial analytical perspectives on GIS*, Taylor & Francis, London.

Flowerdew, R. and Green, M., 1992, Developments in areal interpolation methods and GIS, *The Annals of Regional Science* **26**(1):67-78.

Forster, B. C., 1985, An examination of some problems and solutions in monitoring urban areas from satellite platforms, *International Journal of Remote Sensing* **6**(1):139-151.

Förstner, W., 1999, 3-D City Models: Automatic and semiautomatic acquisition methods, in: *Photogrammetric Week '99* (D. a. S. Fritsch, R., ed.), Herbert Wichmann Verlag, Heidelberg, pp. 291-303.

Fox, J., Rindfuss, R. R., Walsh, S. J. and Mishra, V., 2003, *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*, Kluwer Academic Publishers, Boston.

Frew, J., Carver, L., Fischer, C., Goodchild, M., Larsgaard, M., Smith, T. and Zheng, Q., 1995, The Alexandria Rapid Prototype: Building a digital library for spatial information, in: *Proceedings of the 1995 ESRI User Conference*.

Goodchild M.F. and Proctor, J., 1997, Scale in a digital geographic world, *Geographical and Environmental* **1**(1):5-23.

Goodchild, M.F., Haining, R., Wise, S. and 12 others, 1992, Integrating GIS and spatial data analysis: problems and possibilities, *International Journal of Geographical Information Systems* **6**(5):407-424.

Goodchild, M. F., Lam, N. S., 1980, Areal interpolation: a variant of the traditional spatial problem, *Geoprocessing* **1**:297-312.

Gruen, A., K_bler, O. and Agouris, P., 1997, *Automatic extraction of man-made objects from aerial and space images (II)*, Birkhauser Verlag, Basel-Boston-Berlin.

Haala, N. and Brenner, C., 1999, Extraction of buildings and trees in urban environments, *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):130-137.

Hall, S. S., 1993, *Mapping the next millennium: How computer-driven cartography is revolutionizing the face of science*, Vintage Books, New York.

Handy, S. L. and Clifton, K. J., 2001, Local shopping as a strategy for reducing automobile travel, *Transportation* **28**(4):317-346.

Hess, P. M., 1997, Measures of connectivity, *Places* **11**:58-65.

Holloway, S. R., Schumacher, J. and Redmond, R. L., 1999, People and place: dasymetric mapping using ARC/INFO, in: *GIS solutions in natural resource management: balancing the technical-political equation* (S. Morain, ed.), OnWord Press, Santa Fe, New Mexico, pp. 283-291.

Holt, J. B., Lo, C. P. and Hodler, T. W., 2004, Dasymetric estimation of population density and areal interpolation of census data, *Cartography and Geographic Information Science* **31**(2):103-121.

Hug, C., 1997, Automatic extraction of man-made objects from aerial and space images (II), in: *Extracting artificial surface objects from airborne laser scanner data* (A. Gruen, Baltsavias, E. P., and Henricsson, O., ed.), Birkhauser Verlag, Basel-Boston- Berlin.

Iisaka, J. and Hegedus, E., 1982, Population estimation from Landsat imagery, *Remote Sensing of Environment* **12**(4):259-272.

Jelalian, A., 1992, *Laser Radar Systems*, Artech House, Boston.

Jensen, J. R. and Cowen, D. C., 1999, Remote sensing of urban/suburban infrastructure and socio-economic attributes, *Photogrammetric Engineering and Remote Sensing* **65**(5):611-622.

Kletzli, R. and Peterson, J., 1998, Inertial measurement and LIDAR meet digital ortho photography: A sensor fusion boon for GIS, in: *Proceedings of the ESRI International Users Conference*.

Krizek, K. J., 2003, Residential relocation and changes in urban travel: Does neighborhood-scale urban form matter?, *Journal of the American Planning Association* **69**(3):265-282.

Langford, M. and Unwin, D. J., 1994, Generating and mapping population density surfaces within a geographical information system, *The Cartographic Journal* **31**(1):21-26.

Langford, M., Maguire, D. J. and Unwin, D. J., 1991, The areal interpolation problem: estimating population using remote sensing in a GIS framework, in: *Handling Geographical Information: Methodology and Potential Applications* (I. M. Blakemore, ed.), Longman Scientific and Technical, New York, pp. 55-77.

LeClere, F. B., Rogers, R. G. and Peters, K., 1998, Neighborhood social context and racial differences in women's heart disease mortality, *Journal of Health and Social Behavior* **39**(2):91-107.

Li, G. and Weng, Q., 2005, Using Landsat ETM imagery to measure population density in Indianapolis, Indiana, USA, *Photogrammetric Engineering & Remote Sensing* **71**(8):947-958.

Liverman, D. M., 1998, *People and pixels: Linking remote sensing and social science*, National Academy Press, Washington, DC.

Lo, C. P., 1995, Automated population and dwelling unit estimation from high-resolution satellite images: a GIS approach, *International Journal of Remote Sensing* **16**(1):17-34.

Loh, D. K., Choo, Y. K. and Hsieh, Y. T., 1995, Global resource and environmental information management and exchange, in: *Proceedings of 1995 International Symposium on Sustainable Development*, pp. 154-166.

Longley, P., Goodchild, M. F., Maguire, D. J. and Rhind, D. W., 2005, *Geographical information systems and science*, John Wiley & Sons, New York.

Lwin, K. K. and Murayama, Y., 2009, A GIS approach to estimation of building population for micro-spatial analysis, *Transactions in GIS* **13**(4):401-414.

Lwin, K. K. and Murayama, Y., 2010, Development of GIS tool for dasymetric mapping, *International Journal of Geoinformatics* **6**(1):11-18.

Maantay, J. A., Maroko, A. R. and Herrmann, C., 2007, Mapping population distribution in the urban environment: the cadastral-based expert dasymetric system (CEDS), *Cartography and Geographic Information Science* **34**(2):77-102.

Maas, H. G. and Vosselman, G., 1999, Two algorithms for extracting building models from raw laser altimetry data, *ISPRS Journal of Photogrammetry and Remote Sensing* **54**(2-3):153-163.

MacEachren, A. M. and Kraak, M. J., 2001, Research challenges in geovisualization, *Cartography and Geographic Information Science* **28**(1):3-12.

Matuschak, B. J., 1996, Commentary: GIS is being redefined by current computing trends, in: *The Electronic Atlas Newsletter*.

Maune, D., Daniel, C. and Damron, J., 2000, LIDAR and IFSAR: Pitfalls and opportunities for our future, in: *Proceedings of the American Society for Photogrammetry and Remote Sensing Conference, Washington, DC*.

Mayer, H., 1999, Automatic object extraction from aerial imagery-a survey focusing on buildings, *Computer Vision and Image Understanding* **74**(2):138-149.

Mennis, J., 2003, Generating surface models of population using dasymetric mapping, *The Professional Geographer* **55**(1):31-42.

Moran, E. F. and Brondizio, E., 1998, Land-use change after deforestation in Amazonia, in: *People and Pixels: Linking Remote Sensing and Social Science* (D. Liverman, Moran, E. F., Rindfuss, R. R., and Stern, P. C., ed.), National Academy Press, Washington, DC, pp. 94-120.

Morenoff, J. D. and Sampson, R. J., 1997, Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970-1990, *Social Forces* **76**(1):31-64.

Morenoff, J. D., Sampson, R. J. and Raudenbush, S. W., 2001, Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence, *Criminology* **39**(3):517-559.

Mouw, T., 2000, Job relocation and the racial gap in unemployment in Detroit and Chicago, 1980 to 1990, *American Sociological Review*:730-753.

Mrozinski, R. D. and Cromley, R. G., 1999, Singly- and doubly-constrained methods of areal interpolation for vector-based GIS, *Transactions in GIS* **3**(3):285-301.

Murayama, Y., 2000, Internet GIS for Malaysian population analysis, Sci.Rep. Inst. Geosci., Univ. Tsukuba, Sec. A., **21**,131-146.

Nelson, G. C. and Hellerstein, D., 1997, Do roads cause deforestation? Using satellite images in econometric analysis of land use, *American Journal of Agricultural Economics* **79**(1):80-88.

Okabe, A. and Kitamura, M., 1997, A computational method for market area analysis on a network, *Location Science* **5**(3):198-198.

Okabe, A. and Okunuki, K., 2001, A computational method for estimating the demand of retail stores on a street network and its implementation in GIS, *Transactions in GIS* **5**(3):209-220.

Openshaw, S., 1989, Learning to live with errors in spatial databases, in: *The Accuracy of Spatial Databases*, Taylor & Francis, London, pp. 263-276.

Pebley, A. R., 1998, Demography and the environment, *Demography* **35**(4):377-389.

Peng, Z. R., 1999, An assessment framework for the development of Internet GIS, *Environment and Planning B* **26**(1):117-132.

Peng, Z. R. and Nebert, D. D., 1997, An Internet-based GIS data access system, *URISA Journal* **9**(1):20-30.

Peng, Z. R. and Tsou, M. H., 2003, *Internet GIS: distributed geographic information services for the Internet and wireless networks*, John Wiley & Sons, New York.

Plewe, B., 1997, GIS online: Information retrieval, mapping, and the Internet, OnWord Press, New York.

Poulsen, E. and Kennedy, L. W., 2004, Using dasymetric mapping for spatially aggregated crime data, *Journal of Quantitative Criminology* **20**(3):243-262.

Preobrazenski, A., 1954, Dorewolucjonnyje i sovietskije karty razmieszczenija nasielenia, *Woprosy Geografii Kartographia* **34**:134-149.

Raisz, E., 1948, General cartography, McGraw-Hill Book Co., New York and London.

Randall, T. A. and Baetz, B. W., 2001, Evaluating pedestrian connectivity for suburban sustainability, *Journal of Urban Planning and Development* **127**(1):1-15.

Rees, G., 1999, *The remote sensing data book*, Cambridge University Press, Cambridge.

Reibel, M. and Bufalino, M. E., 2005, Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems, *Environment and Planning A* **37**(1):127-139.

Rhind, D. W., 1991, Counting the people: the role of GIS, *Geographical Information Systems* **2**:127-137.

Robert, S. A., 1998, Community-level socioeconomic status effects on adult health, *Journal of Health and Social Behavior* **39**(1):18-37.

Rossmo, D. K., 1995, Crime analysis through computer mapping (C. R. Block, Dabdoub, M. and Fregly, S., ed.), Police Executive Research Forum Washington, DC, pp. 65-97.

Sahar, L. and Krupnik, A., 1999, Semiautomatic extraction of building outlines from large-scale aerial images, *Photogrammetric Engineering and Remote Sensing* **65**(4):459-466.

Siksna, A., 1997, The effects of block size and form in North American and Australian city centres, *Urban Morphology* **1**(1):19-33.

Sleeter, R., 2004, Dasymetric mapping techniques for the San Francisco Bay region, California, in: *Proceedings of the Urban and Regional Information Systems Association Annual Conference*, Reno, Nevada.

Slocum, T. A., 1999, *Thematic cartography and visualization*, Prentice Hall, New Jersey.

Southworth, M. and Owens, P. M., 1993, The evolving metropolis: Studies of community, neighborhood, and street form at the urban edge, *Journal of the American Planning Association* **59**(3):271-284.

Sutton, P., Roberts, D., Elvidge, C. and Baugh, K., 2001, Census from Heaven: an estimate of the global human population using night-time satellite imagery, *International Journal of Remote Sensing* **22**(16):3061-3076.

Tsou, M. H., 2004, Integrated mobile GIS and wireless internet map servers for environmental monitoring and management, *Cartography and Geographic Information Science* **31**(3):153-165.

Vosselman, G. and Dijkman, S., 2001, 3D building model reconstruction from point clouds and ground plans, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **34**(3/W4):37-44.

Wever, C. and Lindenberger, J., 1999, Experiences of 10 years laser scanning, *Photogrammetric Week* **99**:125-132.

Whyte, J., 2002, *Virtual reality and the built environment*, Architectural Press, Oxford.

Wood, C. H. and Skole, D., 1998, Linking satellite, census, and survey data to study deforestation in the Brazilian Amazon, in: *People and Pixels: Linking Remote Sensing and Social science* (D. Liverman, Moran, E. F., Rindfuss, R. R., and Stern, P. C., ed.), National Academy Press, Washington, DC, pp. 70-93.

Wright, J. K., 1936, A method of mapping densities of population: With Cape Cod as an example, *Geographical Review* **26**(1):103-110.

Yen, I. H. and Kaplan, G. A., 1999, Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study, *American Journal of Epidemiology* **149**(10):898-907.

Internet

3Di, LLC, 2000, Services: Remote Sensing: LIDAR: At http: //www.3Dillc.com/rem-lidar.htm.

Appendix A: List of Equations

Calculation of Building Population

Areametric Method

$BP_i = \left(\frac{CP}{\sum_{k=1}^{n} BA_k}\right) BA_i$ Using building footprint surface area. . . . . .(2-1)

Volumetric Method

$BP_i = \left(\frac{CP}{\sum_{k=1}^{n} BA_k . BF_k}\right) BA_i . BF_i$ Using number of floors information . . . . . . (2-2)

$BP_i = \left(\frac{CP}{\sum_{k=1}^{n} BA_k . BH_k}\right) BA_i . BH_i$ Using average building height . . . . (2-3)

$BP_i = \left(\frac{CP}{\sum_{k=1}^{n} BV_k}\right) BV_i$ Using total building volume . . . . (2-4)

Where:

BP_i Population of building *i*

CP Census tract population

BA_i Footprint area of building *i*

BF_i Number of floors of building *i*

BH_i Average height of building *i* (from LIDAR data)

BV_i Total volume of building *i* (from LIDAR data)

i, k Summation indices

n Number of buildings that meet user-defined criteria and fall inside the CP polygon

Root Mean Square Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\text{Actual} - \text{Estimated})^2}{n}}$$ Root Mean Square Error .... (2-5)

Calculation of Adjusted Floor and Adjusted Volume

aF = tF – bF Adjusted floor .... (3-1)

fV = (aF/tF) Volume adjustment factor .... (3-2)

aV = tV*fV Adjusted volume .... (3-3)

Where:

aF Adjusted Floor

tF Total Building Floors

bF Number of business used floors

aV Adjusted Volume

tV Total Building Volume

fV Volume Adjustment Factor

Mean Center

$\overline{X} = \frac{\sum_i X_i}{n}\ ;\ \overline{Y} = \frac{\sum_i Y_i}{n}$ Mean Center . . . . (4-1)

Weighted Mean Center

$\overline{X} = \frac{\sum_i W_i X_i}{\sum_i W_i}\ ;\ \overline{Y} = \frac{\sum_i W_i Y_i}{\sum_i W_i}$ Weighted Mean Center . . . . (4-2)

Where:

X Longitude

Y Latitude

W Weighted factor (e.g. Building population in the case of Weighted Population Mean Center (W-PMC) and number of shops in the case of Weighted Facility Mean Center (W-FMC))

n Number of points

Appendix B: Glossary

AJAX

1. [Internet] Acronym for *Asynchronous JavaScript and XML*. A programming technique for creating fast, interactive Internet applications. **AJAX** adds a small application to part of the software user's browser for fast loading and display. (ESRI)

accessibility

[business] An aggregate measure of the degree of ease with which a place, person, or thing can be reached, depending on factors such as slope, traffic, distance, and so on.

[usability] The degree to which Web sites, software, or computers provide equivalent information and functionality to a variety of people, including those with disabilities or visual impairment. (ESRI)

attribute

1. [data models] Nonspatial information about a geographic feature in a GIS, usually stored in a table and linked to the feature by a unique identifier. For example, attributes of a river might include its name, length, and sediment load at a gauging station.
2. [data models] In raster datasets, information associated with each unique value of a raster cell.
3. [graphics (map display)] Information that specifies how features are displayed and labeled on a map; for example, the graphic attributes of a river might include line thickness, line length, color, and font for labeling.
4. [ESRI software] In MOLE, aspatial information about a geographic feature in a GIS, usually stored in a table and linked to the feature by a unique identifier. For example, attributes of a force element might include its name

and speed. Most MOLE attributes are what some military specifications refer to as labels or modifiers. (ESRI)

Azimuth

1. The horizontal direction of a vector, measured clockwise in degrees of rotation from the positive Y-axis, for example, degrees on a compass.(AGI)

buffer

1. [spatial analysis] A zone around a map feature measured in units of distance or time. A buffer is useful for proximity analysis.
2. [spatial analysis] A polygon enclosing a point, line, or polygon at a specified distance.
3. [computing] Space on a computer disk or RAM that has been allocated for temporary storage. This temporary storage may also be called a spooler when it is used to hold data in memory before the data is sent to another machine, such as a printer. (ESRI)

Choropleth map

1. [cartography] A thematic map in which areas are distinctly colored or shaded to represent classed values of a particular phenomenon. (ESRI)

dangle node

1. [data capture] The endpoint of a dangling arc.(ESRI)

dasymetric mapping

1. [data analysis] A technique in which attribute data that is organized by a large or arbitrary area unit is more accurately distributed within that unit by the overlay of geographic boundaries that exclude, restrict, or confine the attribute in question. For example, a population attribute organized by census tract might be more accurately distributed by the overlay of water bodies, vacant land, and other land-use boundaries within which it is reasonable to infer that people do not live.(ESRI)

DBMS (Database Management System)

[database structures] A set of software applications used to create and maintain databases according to a schema. Database management systems provide tools for adding, storing, changing, deleting, and retrieving data.

dissolve

1. [ESRI software] A geoprocessing command that removes boundaries between adjacent polygons that have the same value for a specified attribute.(ESRI)

feature

1. Frequently used within GIS referring to point, line (including polyline and mathematical functions defining arcs), polygon and sometimes text (annotation) objects (see also, vector)

filter

1. [spatial analysis] On a raster, an analysis boundary or processing window within which cell values affect calculations and outside which they do not. **Filters** are used mainly in cell-based analysis where the value of a center cell is changed to the mean, the sum, or some other function of all cell values inside the filter. A filter moves systematically across a raster until each cell has been processed. Filters can be of various shapes and sizes, but are most commonly three-cell by three-cell squares.
2. [data analysis] A desktop GIS operation used to hide (but not delete) features in a map document or attribute table.
3. [data analysis] A constraint used to define a subset of data. In ArcWeb Services, for example, users can filter the list of content to show only content related to an industry, such as transportation, or a provider, such as Tele Atlas. (ESRI)

geocoding

1. [geocoding] A GIS operation for converting street addresses into spatial data that can be displayed as features on a map, usually by referencing address information from a street segment data layer.(ESRI)

georeferencing

1. [coordinate systems] Aligning geographic data to a known coordinate system so it can be viewed, queried, and analyzed with other geographic data. Georeferencing may involve shifting, rotating, scaling, skewing, and in some cases warping, rubber sheeting, or orthorectifying the data. (ESRI)

Geodemographics

1. The analysis of people by where they live, in particular by type of neighborhood. Such localized classifications have been shown to be powerful discriminators of consumer behavior and related social and behavioral patterns.

Geospatial

1. Referring to location relative to the Earth's surface. "Geospatial" is more precise in many GI contexts than "geographic," because geospatial information is often used in ways that do not involve a graphic representation, or map, of the information. (OGC)

Geostatistics

1. Statistical methods developed for and applied to geographic data. These statistical methods are required because geographic data do not usually conform to the requirements of standard statistical procedures, due to spatial autocorrelation and other problems associated with spatial data (AGI).

   The term is widely used to refer to a family of tools used in connection with spatial interpolation (prediction) of (piecewise) continuous datasets and is widely applied in the environmental sciences. Spatial statistics is a term more commonly applied to the analysis of discrete objects (e.g. points, areas) and is particularly associated with the social and health sciences

Geovisualization

1. A family of techniques that provide visualizations of spatial and spatio-temporal datasets, extending from static, 2D maps and cartograms, to

representations of 3D using perspective and shading, solid terrain modeling and increasingly extending into dynamic visualization interfaces such as linked windows, digital globes, fly-through animations, virtual reality and immersive systems. Geovisualization is the subject of ongoing research by the International Cartographic Association (ICA)

GIS

1. [GIS technology] Acronym for *geographic information system*. An integrated collection of computer software and data used to view and manage information about geographic places, analyze spatial relationships, and model spatial processes. A GIS provides a framework for gathering and organizing spatial data and related information so that it can be displayed and analyzed.(ESRI)

GI Science

1. [social context of GIS] Abbreviation for *geographic information science*. The field of research that studies the theory and concepts that underpin GIS. It seeks to establish a theoretical basis for the technology and use of GIS, study how concepts from cognitive science and information science might apply to GIS, and investigate how GIS interacts with society. (ESRI)

GPS

1. [GPS] Acronym for *Global Positioning System*. A system of radio-emitting and -receiving satellites used for determining positions on the earth. The orbiting satellites transmit signals that allow a GPS receiver anywhere on earth to calculate its own location through trilateration. Developed and operated by the U.S. Department of Defense, the system is used in

navigation, mapping, surveying, and other applications in which precise positioning is necessary. (ESRI)

GPS/DGPS

1. Global positioning system; Differential global positioning system — DGPS provides improved accuracy over standard GPS by the use of one or more fixed reference stations that provide corrections to GPS data

hillshading

1. [map design] Shadows drawn on a map to simulate the effect of the sun's rays over the varied terrain of the land.
2. [map design] The hypothetical illumination of a surface according to a specified **azimuth** and altitude for the sun. Hillshading creates a three-dimensional effect that provides a sense of visual relief for cartography, and a relative measure of incident light for analysis. (ESRI)

interpolation

[mathematics] The estimation of surface values at unsampled points based on known surface values of surrounding points. Interpolation can be used to estimate elevation, rainfall, temperature, chemical dispersion, or other spatially-based phenomena. Interpolation is commonly a raster operation, but it can also be done in a vector environment using a TIN surface model. There are several well-known interpolation techniques, including SPLINE and Kriging. (ESRI)

[ESRI software] In the context of linear referencing, the calculation of measure values for a route between two known measure values.

Kernel

1. Literally, the core or central part of an item. Often used in computer science to refer to the central part of an operating system, the term kernel in geospatial analysis refers to methods (e.g. density modeling, local grid analysis) that involve calculations using a well-defined local neighborhood (block of cells, radially symmetric function).

layer

1. [data structures] The visual representation of a geographic dataset in any digital map environment. Conceptually, a layer is a slice or stratum of the geographic reality in a particular area, and is more or less equivalent to a legend item on a paper map. On a road map, for example, roads, national parks, political boundaries, and rivers might be considered different layers.
2. [ESRI software] In ArcGIS, a reference to a data source, such as a shapefile, coverage, geodatabase feature class, or raster, that defines how the data should be symbolized on a map. Layers can also define additional properties, such as which features from the data source are included. Layers can be stored in map documents (.mxd) or saved individually as layer files (.lyr). Layers are conceptually similar to themes in ArcView 3.x. (ESRI)

LIDAR

1. [remote sensing] Acronym for *light detection and ranging*. A remote-sensing technique that uses lasers to measure distances to reflective surfaces. (ESRI)

mashup

1. A recently coined term used to describe websites whose content is composed from multiple (often distinct) data sources, such as a mapping service and property price information, constructed using programmable interfaces to these sources (as opposed to simple compositing or embedding).

orthophotograph (orthoimage)

1. [aerial photography] An aerial photograph from which distortions owing to camera tilt and ground relief have been removed. An orthophotograph has the same scale throughout and can be used as a map. (ESRI)

pixel/image

1. Picture element — a single defined point of an image. Pixels have a "color" attribute whose value will depend on the encoding method used. They are typically either binary (0/1 values), grayscale (effectively a color mapping with values, typically in the integer range [0,255]), or color with values from 0 upwards depending on the number of colors supported. Image files can be regarded as a particular form of raster or grid file.

polygon

1. A closed figure in the plane, typically comprised of an ordered set of connected vertices, $\mathbf{v_1,v_2,\ldots v_{n-1},v_n=v_1}$ where the connections (edges) are provided by straight line segments. If the sequence of edges is not self-crossing it is called a simple polygon. A point is inside a simple polygon if traversing the boundary in a clockwise direction the point is always on the right of the observer. If every pair of points inside a polygon can be joined

by a straight line that also lies inside the polygon then the polygon is described as being convex (i.e. the interior is a connected point set).

2. The OGC definition of a polygon is “a planar surface defined by 1 exterior boundary and 0 or more interior boundaries. Each interior boundary defines a hole in the polygon”

polyline

1. An ordered set of connected vertices, $\boldsymbol{v_1,v_2,...v_{n-1},v_n{}^{1}v_1}$ where the connections (edges) are provided by straight line segments. The vertex $\boldsymbol{v_1}$ is referred to as the start of the polyline and $\boldsymbol{v_n}$ as the end of the polyline. The OGC specification uses the term LineString which it defines as: a curve with linear interpolation between points. Each consecutive pair of points defines a line segment

raster/grid

1. A data model in which geographic features are represented using discrete cells, generally squares, arranged as a (contiguous) rectangular grid. A single grid is essentially the same as a two-dimensional matrix, but is typically referenced from the lower left corner rather than the norm for matrices, which are referenced from the upper left. Raster files may have one or more values (attributes or bands) associated with each cell position or pixel.

resolution

1. [cartography] The detail with which a map depicts the location and shape of geographic features. The larger the map scale, the higher the possible resolution. As scale decreases, resolution diminishes and feature

boundaries must be smoothed, simplified, or not shown at all; for example, small areas may have to be represented as points.

2. [graphics (computing)] The dimensions represented by each cell or pixel in a raster.
3. [graphics (computing)] The smallest spacing between two display elements, expressed as dots per inch, pixels per line, or lines per millimeter.
4. [ESRI software] In ArcGIS, the smallest allowable separation between two coordinate values in a feature class. A spatial reference can include x, y, z, and m resolution values. The inverse of a resolution value was called a precision or scale value prior to ArcGIS 9.2.

(Spatial) Autocorrelation

1. The degree of relationship that exists between two or more (spatial) variables, such that when one changes, the other(s) also change. This change can either be in the same direction, which is a positive autocorrelation, or in the opposite direction, which is a negative autocorrelation. The term autocorrelation is usually applied to ordered datasets, such as those relating to time series or spatial data ordered by distance band. The existence of such a relationship suggests but does not definitely establish causality. (AGI)

TIN

1. [data structures] Acronym for *triangulated irregular network*. A vector data structure that partitions geographic space into contiguous, nonoverlapping triangles. The vertices of each triangle are sample data points with x-, y-, and z-values. These sample points are connected by lines to form Delaunay triangles. TINs are used to store and display surface models. (ESRI)

2. Triangulated irregular network. A form of the tesseral model based on triangles. The vertices of the triangles form irregularly spaced nodes. Unlike the grid, the TIN allows dense information in complex areas, and sparse information in simpler or more homogeneous areas. The TIN dataset includes topological relationships between points and their neighboring triangles. Each sample point has an *X,Y* co-ordinate and a surface, or *Z*-Value. These points are connected by edges to form a set of non-overlapping triangles used to represent the surface. TINs are also called irregular triangular mesh or irregular triangular surface model (AGI)

Topology

1. The relative location of geographic phenomena independent of their exact position. In digital data, topological relationships such as connectivity, adjacency and relative position are usually expressed as relationships between nodes, links and polygons. For example, the topology of a line includes its from- and to-nodes, and its left and right polygons (AGI). In mathematics, a property is said to be *topological* if it survives stretching and distorting of space.

vector

1. Within GIS the term vector refers to data that are comprised of lines or arcs, defined by beginning and end points, which meet at nodes. The locations of these nodes and the topological structure are usually stored explicitly. Features are defined by their boundaries only and curved lines are represented as a series of connecting arcs. Vector storage involves the storage of explicit topology, which raises overheads, however it only stores those points which define a feature and all space outside these features is "non-existent" (AGI)
3. In mathematics the term refers to a directed line, i.e. a line with a defined origin, direction and orientation. The same term is used to refer to a single

column or row of a matrix, in which case it is denoted by a bold letter, usually in lower case.

Voronoi diagram

1. [Euclidean geometry] A partition of space into areas, or cells, that surround a set of geometric objects (usually points). These cells, or polygons, must satisfy the criteria for Delaunay triangles. All locations within an area are closer to the object it surrounds than to any other object in the set. **Voronoi diagrams** are often used to delineate areas of influence around geographic features. Voronoi diagrams are named for the Ukrainian mathematician Georgy Fedoseevich Voronoi (1868-1908).(ESRI)

Zonal Statistics

1. [ESRI software] In ArcGIS Spatial Analyst, the calculation of a statistic for each zone of a zone dataset based on values from another dataset, a value raster. A single output value is computed for each cell in each zone defined by the input zone dataset. (ESRI)