POPULATION DISTRIBUTION DURING THE DAY

Budhendra Bhaduri[*]
Geographic Information Science & Technology Group
Oak Ridge National Laboratory

SYNONYMES

Daytime Population; Mobile Population; Non-residential Population

DEFINITION

Population distribution during the day can be defined as distribution of population in an area during the daytime hours. However, a precise definition of daytime hours is challenging given the geographic variability in the length of a day or daylight hours. The US Census Bureau used "normal business hours" as the span of time to describe daytime population [1]. Given that Censuses typically estimates residential population, it represents a nighttime population distribution. In that respect, daytime population in an area may be broadly defined as distribution of population at times other than when they are expected to be at their residences at night which extends the duration from business hours to include the evening hours as well.

HISTORICAL BACKGROUND

Population data has served as a fundamental backbone for planning sustainable development. There is evidence from the early 1600s of a population census in Virginia where people were counted in nearly all of the British colonies that became the United States at the time of the Revolutionary War [2]. Historically it has been used to meet a variety of long term socio-economic and political planning needs. For example, the first US Census of 1790, that counted 3.9 million residents, helped raise the membership in the U.S. House of Representatives from an original 65 to 105. Modern censuses that include not only population count but also its demographic and socio-economic characteristics have had a tremendous impact on many aspects of our society covering , among others, urban planning and housing development, transportation planning, energy demand and infrastructure planning, healthcare planning, environmental impact assessment, emergency preparedness and response, and scientific research. However, majority of these planning activities aimed at medium to long-term solutions over a

number of years and consequently a general geographic assessment of population, described through their residential locations, was adequate to address such planning processes. Movement of population during a day results directly from people traveling to their locations of daytime activities (employment, business, educational institutions, and recreational locations) away from their residences [3]. The patterns of such population displacements depend on the relative geographic distribution of residential and business areas. In most modern societies, these two activity locations are distinctly separated in space and employment or business locations contain fewer residences compared to businesses, Consequently, a large number of people move into these areas while only a few leave resulting in a substantial swelling in the daytime population of that area. Motivation to formalize the concept of non-residential and daytime population distribution roots predominantly in two areas. First, it has been well perceived that understanding of the daytime population distribution provides a very competitive economic advantage as businesses are enabled to target specific consumer bases depending on their locations and convenience of access during majority of the 24-hour period when people are out of their residences. In recent years, a stronger requirement for understanding daytime population has emerged from the emergency preparedness and response community to assess the at risk population from the threats of technological and natural disasters and deliberate attacks on human lives such as terrorist events.

SCIENTIFIC FUNDAMENTALS

Population movement is a function of both geographic space as well as time. Mobility of population during the day is driven by people's need to temporarily relocate to the locations of daytime activities that include places of education (schools, colleges, universities), employment, businesses (shopping, post offices, restaurants, and others), or recreational (parks, museums, other tourist attractions). In general daytime population distribution of an area can be conceptually expressed as:

Daytime Population = Workers + School children + Tourists + Business travelers + Residual Nighttime Residential Population

Or,

Daytime Population = Nighttime Population + Daytime incoming population – Daytime outgoing population

Deriving a quantitative estimate from the above qualitative expressions involves further analyses of population data which can be represented as:

Daytime Population = Nighttime Residential Population – Workers leaving during the day + Workers moving in during the day – School children leaving during the day + School children moving in during the day + Tourists visiting during the day + Business travelers coming into the area

Although this may not be the most accurate representation, it is largely accepted as the expression leading to the best available daytime population estimates. It follows that, developing quantitative estimates of daytime population distribution involves two distinct components. The first component involves identification of daytime activity locations such as businesses, schools, and other recreational activities. The second component covers the identification and distribution of the mobile population that are at those locations. Usually, it is easier to gather information on the first as these are static geographic features and are commonly captured in public and commercial databases for various infrastructures, or can be derived from remote sensing based land cover data, high resolution satellite and aerial photographs, or state and local government data. However, it is extremely challenging to obtain information on the number and nature of movement of people during the day that comprehensively captures the net displacement of nighttime residential population during daytime. Although, detailed population movement data sets may be available for isolated local communities, they are not available at a national scale. In fact, the US Census Bureau's compilation of journey to work data is the only readily available and nationally consistent data set for the US that describes people's movement from residences to employment locations. Consequently, the US Census Bureau's estimate of daytime population based on the 2000 Census only reflect populations based on travel to work. Similarly, it does not limit the work related commuting to specific hours. All worker-related travel, irrespective of what time of the day it occurs, has been used to derive these estimates of daytime population [1].

An important aspect of daytime population distribution is the geospatial scale at which it is estimated. Theoretically, the finest spatial resolution achievable through the map algebra technique described above is directly tied to the finest scale of the available input data. For example, the US Census Bureau collects worker commuting data at the census tract level and reports national daytime population distribution at the county level. It also reports estimates of daytime population for key cities in each state. Similar city level estimates of daytime population from government and commercial sources are available for Japan [4], Canada, and the US. All these data sets appear to be heavily focused on worker population movement during the day and the data is presented through vector data models (points and polygons). For example, daytime population fluxes are restricted to individual county and city boundary polygons. Some commercial databases represent individual activity locations as points which potentially offer high spatial accuracy but mostly account for worker population at individual business locations. In reality, the data sets necessary to comprehensively estimate daytime population exist in the forms of points and polygons which makes it challenging to create a high resolution population distribution through simple map algebra analysis. It requires integration of disparate spatial data and advanced geospatial modeling where the spatial model enables decomposition of the input data into finer spatial resolutions and represented through uniform raster or gridded dataset.

Decomposition of population distribution estimates has been a well known problem. Several interpolation and decomposition methods have been developed to address this issue with census (polygonal) population data. They include areal weighting, pycnophylactic interpolation, dasymetric mapping, and various smart interpolation

techniques. Aerially weighted interpolation is the simplest of the methods where a regular grid is intersected with the Census polygon and each grid cell is assigned a value based on the proportion of the polygon contained in each cell [5, 6, 7]. This method implies an assumption of uniform distribution of population which is not a realistic solution for decomposition of population data. Pycnophylactic interpolation extends aerial weighting methodology by applying a smoothing function to the raster cell values, with the weighted average of its nearest neighbors, iteratively while preserving the total population count of the polygon [8]. This method creates a continuous surface which contradicts the obvious discontinuous nature of population distribution. Dasymetric modeling is analogous to areal interpolation but uses ancillary spatial data to aid in the interpolation process. The ancillary spatial data is at a finer spatial resolution and the variability in its values enables an asymmetric allocation of population values. Land cover/land use is the best example in this respect where different land cover or land use categories for each cell can be used as weighting function for population distribution such as urban areas will have a higher weight than forested areas (Figure 1) [7, 9, 10]. Smart interpolation, in principle, is a multidimensional version of dasymetric model where the allocation refinement comes from more than one ancillary data sources which are at a finer resolution than the population polygon [10, 11]. Utility of such interpolation techniques at local scales are well documented. In fact, there are two well known publicly available data sets, Gridded Population of the World (GPW) [16] and LandScan Global Population database [12], which employ such method to produce global population distribution data. While GPW is a product of simple aerial weighting interpolation at 2.5 arc-minute or approximately 5 km cell size, LandScan, at 30 arc-second or approximately 1 km cell size the finest resolution global population data available to date, is derived through advanced spatial data integration and multidimensional dasymetric modeling or smart interpolation.

Figure 1 Goes Here

[Figure 1: Example illustrating how land-cover data is used in the LandScan daymetric model.]

Although both GPW and LandScan datasets are developed using Census information, GPW depict a nighttime residential population (i.e. directly decomposed Census data) while the LandScan database represent an "ambient" or average of the 24 hour period population (because the model assigns some part of the populations based on non-residential activities). Both databases are publicly available for non-commercial usage from Columbia University [18] and Oak Ridge National Laboratory [19] respectively. GPW is updated periodically, while LandScan Global Population database has been updated and released annually since 2000.

Development of daytime population distribution models and databases is significantly more challenging as it requires further integration and modeling of activity based datasets into the residential population distribution model. In 2004, the US Census Bureau released the following three daytime population distribution data tables based on the 2000 census [1]:

- Table 1. Leading Places on Percent Change in Daytime Population, by Size (202 highly populated cities)
- Table 2. The United States, States, Counties, Puerto Rico and Municipalities
- Table 3. Selected Places by State (6524 communities)

However, these data sets only take into account commuting worker population in an area. The best spatial resolution of these data is still at the community level (small cities) and thus is appropriate for general purpose planning. Expanding on their LandScan Global Population research program, since early 2000s, the US Department of Energy's Oak Ridge National Laboratory has played a pioneering role in developing an advanced scalable daytime population distribution model for the nation called LandScan USA [14, 15]. At an unprecedented resolution of 3 arc-second or approximately 90m cell size, LandScan USA demonstrates a consistent methodology for developing daytime population distribution. This enhancement in resolution comes from the incorporation of a large number of high resolution ancillary data sets used in the LandScan USA dasymetric model. Some of these ancillary data sets include:

- Population
    - Census Block population; Census Tract-to-track worker flow; Bureau of Labor Statistics quarterly updates for worker population.
- Roads
    - Tele Atlas North America (formerly known as Geographic Data Technology) Dynamap; US Census TIGER data; Navigational Technologies (NAVTEQ) roads;
- Land Cover/Land Use
    - National Land Cover Data (NLCD); State and local level GIS databases;
- Slope
    - National Elevation Data (NED)
- Academic Institutions
    - Department of Education; Environmental Systems Research Institute(ESRI); Tele Atlas North America;
- Prisons
    - Department of Justice National Jail Census
- Hospitals
    - American Hospital Association (AHA)
- Business Employment
    - ESRI Business Database (Info USA)
- Ortho Imagery
    - Google Earth, Earth Viewer; Microsoft Terra Server

In addition to the worker population in an area, LandScan USA database accounts for children at K-12 school age, university students, institutional population (jails and prisons), and a mobile daytime residential population at various activity locations such as shopping malls, post offices, cultural attractions, and recreational facilities (parks). Development of LandScan USA version 1.0 has been completed for the 50 US States and

Puerto Rico but the data is not publicly available yet. It should be noted that both LandScan Global and LandScan USA are evolving databases and new ancillary input data sets are continuously added to the model as they become available.

Figure 2 Goes Here

[Figure 2: Difference in population distribution between nighttime and daytime as illustrated by LandScan USA data for Washington DC.]

A similar approach [20] has also been adopted for estimating daytime population at a lower resolution of 250 meter grid cells. The coarser resolution of this data set has been attributed to coarser resolution of input variables in the model such as county-to-county worker mobility data from Census as compared to tract-to-tract worker mobility data used in LandScan USA. Moreover, this estimation is solely based on worker and residential population and does not account for population at academic institutions, commercial retail locations, and recreational areas.

KEY APPLICATIONS

Missions of national priority ranging from socio-environmental studies to homeland security utilize population data as one of the critical elements. High resolution population distribution data during daytime hours is even more significant for successfully addressing research and practical applications that require an estimation of mobile population. Applications of daytime population distribution are numerous and can be broadly divided into two categories:

Estimating Population at Risk from Disasters

Large number of human lives become at risk from natural and technological disasters. Volcanic eruptions, earthquakes, hurricanes, floods, wildfires, blizzards, droughts, and tornadoes are examples of natural disasters that are slow on their onset, predictable in most cases (except earthquakes and volcanic eruptions), and typically geographically restricted. However, natural disasters are unpreventable and in most part uncontrollable. Technological disasters, on the other hand, occur suddenly (unpredictable) but can be controlled and their impact minimized through effective disaster planning and management. Examples of technological disasters include explosions, electrical blackout, nuclear accidents, and bioterrorism. Critical application domains for estimating population at risk include national and homeland security where improved knowledge of where people live relative to sites of potential terrorist activities and refine estimates of potential populations exposed [20] or injured for rapid risk assessment; emergency preparedness and response where daytime population can support emergency response resource planning, emergency evacuation planning, and disaster relief delivery.

Public Health and Socioeconomic Analysis

Public health is probably one of the most promising areas that can take great advantage of daytime population distribution. Disease epidemiology with short and long term exposure assessment; evaluating access to health care facilities and locating future health care facilities can be done effectively with an understanding of daytime population distribution. Assessment of the mobility daytime population with respect to their residences also facilitates understanding of contagious disease propagation patterns. Exposure and risk assessment from environmental pollutants for work related activities can be performed very effectively using daytime distribution of population. Given that workers are likely to spend 50% or less of their time at home during work days, the daytime population distribution provides a tremendous advantage for occupational exposure analysis over using traditional census data. High resolution daytime population data also reduces population distribution errors around point and area sources and can be of significant help for environmental justice analysis. Other socio-economic applications include demographic analysis to evaluate socioeconomic disparity patterns of a region in terms of work related commuting patterns for different demographic groups and estimating rates and trends of urban sprawl.

FUTURE DIRECTIONS

Accurate estimation and representation of daytime population distribution poses significant challenges. First, any region witnesses an influx of tourists (or visitors) and business travelers during the day, particularly in large urban areas and cultural/natural attractions (such as national parks). In addition, a large number of people travel along roads driving through major urban areas. Such transitional population is not effectively captured in any consistent and organized databases and will require advanced data integration and modeling techniques to be effectively included in a daytime population distribution. Current spatial modeling and population distribution techniques only locate population at specific activity locations and do not account for the commuting time when the mobile population is on the transportation networks. Thus a true average daytime population distribution requires details of worker commuting patterns and non-worker travel habits, which includes data for the number of people at necessary service locations such as post offices, banks, shops, and parks. Another important aspect of daytime population distribution is to characterize the temporal variability. The nature of daytime population distribution can be significantly different depending upon whether the data represents a working or a weekend day or a holiday. Moreover, there is also a seasonal and weather impact on the daytime population distribution. Summer days do not have students in academic institutions and have more people at outdoor locations. In contrast, a larger population tends to be indoors during days with weather extremes (such as extreme heat, cold, or storms). Thus assessment of a true "representative" daytime population of a region will require development of an average distribution from such different daytime population distribution scenarios.

CROSS REFERENCES

- Geodemographic Segmentation
- Spatial Data Analysis
- Land Use, Land Cover Classifications
- Visualization and Modeling of Geographic Dynamics
- Movement Patterns in Spatio-Temporal Data
- Temporal GIS and Applications
- GeoSpatial Intelligence
- Homeland Security and Spatial Data Mining
- Fitness of Use of GeoSpatial Data

RECOMMENDED READING

1. Census 2000 PHC-T-40. Estimated Daytime Population and Employment-Residence Ratios: 2000; Technical Notes on the Estimated Daytime Population; (http://www.census.gov/population/www/socdemo/daytime/daytimepoptechnotes.html).

2. US Census Bureau's Fact Finer for the Nation: History and Organization. May 2000; (http://www.census.gov/prod/2000pubs/cff-4.pdf).

3. Quinn, J (1950) The Daytime Population of the Central Business District of Chicago. Review by Breese Gerald W. American Sociological Review. 15 (6) 827-828.

4. Japanese Statistics Bureau, Ministry if Internal Affairs and Communications; (http://www.stat.go.jp/english/data/kokusei/2000/jutsu1/00/01.htm).

5. Goodchild, M, Anselin, L and Deichmann, U (1993) A Framework for the Aerial Interpolation of Socioeconomic Data. Environment and Planning. A 25. 383–397.

6. Goodchild, M and Lam, N (1980) Aerial Interpolation: A Variant of the Traditional Spatial Problem. Geo-Processing. 1. 297–312.

7. Mennis, J (2003) Generating Surface Models of Population Using Dasymetric Mapping. Professional Geographer. 55 (1) 31-42.

8. Tobler, W (1979) Smooth Pycnophylactic Interpolation for Geographical Regions. Journal of the American Statistical Association. 74 (367) 519-530.

9. Wright, J (1936) A Method of Mapping Densities of Population: With Cape Cod as an Example. Geographical Review. 26 (1) 103-110.

10. Langford M, Unwin D (1994) Generating and Mapping Population Density Surfaces Within a Geographical Information System. Cartogr Journal. 31 (1) 21-26.

11. Cohen* J and Small C (1998) Hypsographic Demography: The Distribution of Human Population by Altitude. Proceedings of the National Academy of Science. 95 (24) 14009-14014.

12. Dobson J, Bright E, Coleman P, Durfee R and Worley B (2000) LandScan: A Global Population Database for Estimating Populations at Risk. Photogrammetric Engineering & Remote Sensing. 66 (7) 849–857.

13. Dobson J, Bright E, Coleman P and Bhaduri B (2003) LandScan2000: A New Global Population Geography, Remotely-Sensed Cities. Taylor and Francis. 267-279.

14. Bhaduri B, Bright E, Coleman P and Dobson J (2002) LandScan: Locating People is What Matters. Geoinformatics. 5 (2) 34-37.

15. Cai Q, Rushton G, Bhaduri B, Bright E and Coleman P (2006) Estimating Small-Area Populations by Age and Sex Using Spatial Interpolation and Statistical Inference Methods. Transactions in GIS. 10 (4) 577–598.

16. Deichmann U, Balk D and Yetman G (2001) Transforming Population Data for Interdisciplinary Usages: From Census to Grid; http://sedac.ciesin.columbia.edu/plue/gpw/GPWdocumentation.pdf.

17. McPherson T and Brown M (2004) Estimating Daytime and Nighttime Population Distributions in U.S. Cities for Emergency Response Activities. Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone. 84th AMS Annual Meeting, Seattle WA.

18. Gridded Population of the World (GPWv3). Center for International Earth Science Information Network. Columbia University, New York, New York; http://sedac.ciesin.org/gpw/.

19. LandScan Global Population Project. Oak Ridge National Laboratory. Oak Ridge, Tennessee; http://www.ornl.gov/sci/landscan/.

20. McPherson T and Brown M (2004) Estimating Daytime and Nighttime Population Distributions in U.S. Cities for Emergency Response Activities. Symposium on Planning, Nowcasting, and Forecasting in the Urban Zone. 84th AMS Annual Meeting, Seattle WA.