Journal of Korea Spatial Information Society Vol.22, No.3 : 89-99, June 2014
http://dx.doi.org/10.12672/ksis.2014.22.3.089
ISSN 2287-9242(Print)
ISSN 2287-9250(Online)

대시메트릭 매핑 기법을 이용한 서울시 건축물별 주거인구밀도의 재현

Representation of Population Distribution based on Residential Building Types by using the Dasymetric Mapping in Seoul

이석준* · 이상욱** · 홍보영*** · 엄홍민**** · 신휴석***** · 김경민******

Sukjoon Lee · Sang Wook Lee · Bo Yeong Hong · Hongmin Eom · Hyu-Seok Shin · Kyung-min Kim

요 약 도시 내 거주하는 인구의 분포는 다양한 연구에 활용될 수 있는 기초자료로서 중요한 의미를 지닌다. 그러나 이와 관련한 데이터는 정보기밀성 등을 이유로 일정한 공간단위로 집계되어 제공된다. 때문에 제공되는 것과 다른 공간단위나, 더 세밀한 공간분석을 요하는 연구와 활용이 제한된다. 이에 본 연구는 인구분포 데이터가 가지는 한계를 극복하여, 행정구역을 기준으로 한 인구데이터 제공단위보다 세밀한 인구분포를 현실적으로 재현하고자 대시매트릭 매핑 기법을 활용하였다. 대시메트릭 매핑은 주 데이터와 연관된 공간적 보조 데이터를 사용하여 주 데이터의 공간적 분포를 세밀하게 나타내는 지도화 방법이다. 본 연구에서는 서울시 집계구별 인구센서스 자료를 주 데이터로, 건축물 대장을 보조 데이터로 사용하였으며, 모든 인구가 주거용 건축물에 거주한다고 가정하여 주거용 건축물 단위에서 인구분포를 재현하였다. 주거용 건축물을 공간적으로 추출하기 위해 바이너리기법이 활용되었으며. 이후 회귀분석 기법을 활용하여 건물 용도와 층수를 고려한 가중치를 부여해 입체적인 거주인구밀도를 재현하는데 성공하였다. 이를 통해 전체 토지 및 용도구역을 기반으로 한 기존의 방법론보다 현실적용오차를 줄였다. 또한 주거가능 건축물의 주택유형과 층수가 반영된 3차원적 공간자료를 활용하여 주거용 건축물에 거주하는 인구를 추정하는 모델을 도출하여, 정확도가 높고 세밀화 된 인구분포지도를 제작하였다. 본 연구를 통해 도출된 개별 건물 단위의 인구 데이터는 일정한 공간단위로 집계를 재가공한 것으로 정보기밀성의 문제가 없는 세밀한 단위의 인구분포 정보를 제공한다. 따라서 이후 도시 내 재난재해 발생에 따른 영향 인구 파악과 같은, 지역인구를 활용한 다양한 연구에 활용될 수 있는 중요한 자료로써 의미를 지닌다.

키워드 : 대시매트릭 매핑, 단계구분도, 회귀분석 기법, 집계구 인구, 건축물대장

Abstract The aim of this study is to represent the residential population distribution in Seoul, Korea more precisely through the dasymetric mapping method. Dasymetric mapping can be defined as a mapping method to calculate details from truncated spatial distribution of main statistical data by using ancillary data which is spatial data related to the main data. In this research, there are two types of data used for dasymetric mapping: the population data (2010) based on a output area survey in Seoul as the main data and the building footprint data including register information as ancillary spatial data. Using the binary method, it extracts residential buildings as actual areas where residents do live in. After that, the regression method is used for calculating the weights on population density by considering the building types and their gross floor areas. Finally, it can be reproduced three-dimensional density of residential population and drew a detailed dasymetric map. As a result, this allows to extract a more realistic calculating model of population distribution and draw a more accurate map of population distribution in Seoul. Therefore, this study has an important meaning as a source which can be applied in various researches concerning regional population in the future.

Keywords : Dasymetric mapping, Chropoleth map, Regression method, Census track population, Building footprint data

† This article is based on research, Simulation model for spatio-temporal distribution of urban population (2012R1A1A1015571), supported by Researcher Supporting Program of National Research Foundation of Korea.

* Sukjoon Lee, Researcher, Environmental Planning Institute, Seoul National University. sukjoon@snu.ac.kr
** Sang Wook Lee, Doctor's Student, GSES, Seoul National University. bestswlee@snu.ac.kr
*** Bo Yeong Hong, Master's Student, GSES, Seoul National University. promise0ct@snu.ac.kr
**** Eom Hongmin, Master's Student, GSES, Seoul National University. djaghdals@naver.com
***** Hyu-Seok Shin, Visiting Researcher, Institute for Korean Regional Studies, Seoul National University. gisci.hyu@gmail.com
****** Kyung-min Kim, Associate Professor, GSES, Seoul National University. kkim2@snu.ac.kr (Corresponding Author)

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서 론

최근 들어 사회경제적 빅데이터의 활용이 증가하면서 이것의 역할과 중요성, 구현방법 및 응용방법에 대한 논의가 대두 되고 있다. 하지만, 이러한 데이터는 방대한 양을 집약적으로 나타내고, 정보제공자의 프라이버시와 정보의 기밀성(confidentiality)을 유지하기 위해(Fisher 외[5]), 일정한 공간단위를 기준으로 집계되는 경우가 많다. 임의적인 공간을 기준으로 집계된 사회경제적 데이터는 도시 내 소규모 공간에서 발생하는 이벤트에 대해 영향을 받는 세부인구의 규모 및 공간적인 영향 패턴, 주어진 공간 경계에 걸쳐 조성된 사회적 클러스터와 같은 현실의 사회경제적 현상을 분석하는데 제약이 되어왔으며, 이를 해결하기 위한 대안들이 논의되어 왔다.

공간적으로 집계된 데이터를 별도의 가공 없이 공간단위 면적으로 나누어 지도화(mapping)한 것을 코로폴레스 맵(Choropleth map, 이하 단계구분도)이라고 한다. 이는 특정 경계 내의 합산된 데이터만 보여주기 때문에, 데이터를 지도화하는 과정에서 특정 공간의 평균적인 밀도데이터로 대표(homogeneous)되어 세부적인 데이터의 특성을 잃어버리게 된다. 즉, 인구밀도를 재현할 경우 면적이 큰 지역은 낮은 인구밀도를, 면적이 작은 지역은 높은 인구밀도를 나타내는 경향을 보인다(Langford외[10]; Holt 외[7]). 이러한 문제점을 보완하고 실제 데이터의 이질성(heterogeneous) 표현을 위하여 대시메트릭 매핑(Dasymetric mapping)의 개념이 사용되는데, 이는 주 데이터와 연관된 공간적 보조 데이터를 사용하여 더 세밀한 밀도 분포를 재현하는 방법이다(Eicher 외[4]; Munnis 외[14]). 대시메트릭 매핑을 이용할 경우, 보조 정보를 토대로(예를 들어 토지 이용용도) 데이터의 공간 단위를 더 작은 크기로 분할하여 각 구역 내의 밀도 분포를 실제와 더 가깝게 추정할 수 있다.

인구 데이터는 도시 공간 내의 인구분포를 알 수 있는 중요한 지표이며, 더 나아가 지역계획이나 정책적 의사결정에 중요한 토대가 된다. 따라서 인구밀도를 얼마나 세밀하게 추정하고 재현하느냐에 따라 다양한 실제 상황에 대한 의사결정의 적합도가 결정된다. 이러한 이유로 인구밀도분포의 정확한 추정을 위한 대시메트릭 매핑기법이 주목받고 있다.

대시메트릭 매핑을 활용한 기존 연구들에서는 집계된 자료를 세부 공간단위로 할당하기 위한 방법론 구축 측면의 연구가 주로 이루어졌으며, 이때 활용된 보조 데이터로는 토지이용용도가 주를 이루고 있다. 그러나 토지이용용도를 사용한 기존의 방법론에서는 인구가 거주하지 않는 건축물 외 지역에서까지 인구가 추정되는 문제와 동일한 용도의 토지에서도 건축연면적과 같이 지역에 따라 다르게 나타나는 실제적인 공간이용 등이 고려되지 못해 세밀한 인구추정에는 한계를 보이고 있다. 이에 본 연구는 2차원적 토지이용 중심의 대시메트릭 매핑의 한계를 보완하고자 실제인구가 거주하는 건축물 단위에서 인구분포를 나타내고자 한다. 또한 개별 건물의 층수와 유형별 특징, 지역적 특성을 반영하여 현실에 가까운 건축물별 거주인구밀도 모델을 제시하는 것이 본 연구의 목적이다.

2. 이론적 배경

2.1 대시메트릭 매핑의 개념

대시메트릭 매핑은 단계구분도[1]과 함께 대표적인 지역 기반의 지도 제작 기법중 하나이다(McCleary[12]; Holt 외[7]). 대시메트릭 매핑은 18세기 초반, 근대 지도학 기법이 성립되면서 인구 분포를 지도에 나타내는 기법으로 등장했으나, 코로플레스 매핑과 달리 1900년대에 이르러서는 잘 사용되지 않았다(Eicher외[4]). 이후 Wight[16]에 의하여 구체적인 인구 밀도 재현 방법론으로 다시 대두되었고, 데이터의 공간 분포를 보여주는 효율적인 방법으로 인정받아 현재의 연구에 이르고 있다.

코로플레스 매핑은 일반적으로 행정적 관리나 사회경제적 데이터의 집계 기준으로 설정한 구역만을 기반으로 하는 반면에, 대시메트릭 매핑은 이와 다른 위계를 가지는 공간 데이터를 보조적으로 이용함으로써 어떠한 현상의 분포를 작은 공간 단위에서 나타낼 수 있다. 즉, 대시메트릭 매핑은 주 데이터를 기반으로 이와 관련된 공간적인 보조 데이터를 활용하여 주 데이터의 구역 경계를 보다 더 세밀한 공간 단위로 나누어 주 데이터의 세부내용을 재현하는 방법으로, 통계적 표면의 가장 적합한 표현이라는 목적을 갖는다(Eicher 외[4]).

이러한 대시메트릭 매핑은 데이터 분해(Data deaggregation)와 에어리얼 인터폴레이션(Areal interpolation) 등과 밀접한 관련을 갖는다(Bajat 외[3]). Fisher 외[5]에 의하면 에어리얼 인터폴레이션은 하나의 공간 단위를 바탕으로 하는 데이터를 다른 공간 단위의 데이

1) 단계구분도는 행정적 구역 경계를 기준으로 집계된 데이터의 값을 공간단위의 전체 면적으로 나누어준 밀도분포 산출 방식을 의미한다.

터로 전환하는 과정을 의미하는 것으로, 한정된 공간 단위에서 벗어나 더 세부적인 공간 또는 다른 특징을 갖는 공간 단위의 데이터로 주 데이터를 가공하는 방법을 말한다. 이러한 점에서 에어리얼 인터폴레이션은 대시메트릭 매핑을 포괄하는 상위개념으로 특정 공간 단위로 집계된 데이터를 다른 종류의 구역 체계로 전환할 필요가 있는 경우 사용된다(Eicher 외[4]).

대시메트릭 매핑은 에어리얼 인터폴레이션의 과정에서 주 데이터과 관련 있는 보조정보를 활용해 기존 자료를 더 작은 공간으로 구분한다는 점에서 일반적인 에어리얼 인터폴레이션과 구분된다. 이러한 개념의 연장선상에서 대시메트릭 매핑은 연속적인 데이터의 표현과 단계 구분적인 데이터의 표현의 중간 형태로 나타나게 된다(MacEachren[12]).

2.2 대시메트릭 매핑의 유형 및 원리

대시메트릭 매핑 기법은 지형도를 보조 데이터로 이용하여 주거지역과 비주거지역으로 나눈 공간에 인구밀도를 재분포하는 방법으로 처음 제안되었다(Wright[16]). 이후 다양한 대시메트릭 매핑의 기법들이 연구되었으며, 대표적으로 바이너리 기법(Binary Method), 다중-클래스 기법(Multi-Class Method), 제한변수 기법(Limiting Variable Method)등이 사용되고 있다(Eicher 외[4], 이상일 외[2]).

바이너리 기법은 보조 데이터를 활용해 불필요한 영역들을 제거하기 위한 방법으로 유의미한 지역과 무의미한 지역의 두 가지로 속성을 구분하여 각각 1과 0의 가중치를 부여하는 방식으로 적용된다. 다중-클래스 기법(Multi-Class Method)은 바이너리 기법의 연장선에서 속성을 세분화 하여 적용하는 방법이다. 예를 들어, Eicher 등[4]은 토지이용에 따른 인구분포 연구에서 다중 클래스기법을 활용하면서, 인구가 거주 가능한 지역을 도시(0.7), 농경지(0.2), 산림지(0.1)로 구분하여 각각의 거주 비중에 차이를 두었으며, 수계 공간(0.0)에는 인구가 거주하지 않는 것으로 설정하였다.

제한변수 기법(Limiting Variable Method)은 실제에 가까운 특성을 반영하여 개별 속성별 한계치를 미리 설정하고, 이를 초과하는 사례에 대해서 속성을 바꾸어 적용하는 방법이다. Eicher 등[4]의 연구에서는 인구가 거주할 수 있는 토지의 용도(도시, 농지, 산림지)를 정하고 동일한 인구밀도로 설정한 한 후, 특정 토지용도에 대해서 전체 인구에 대한 최대 밀도 한계를 지정하고 그보다 높은 값을 보이는 구역은 다른 토지용도로 전환하여 연구에 이용하고 있다.

토지의 용도에 따른 가중치를 반영하는 것은 대시매트릭 매핑의 정확도를 결정한다는 측면에서 중요하다. 대시메트릭 매핑 기법을 이용한 기존의 인구분포 추정 연구에서는 주로 토지의 용도에 따른 가중치를 바탕으로 인구 분포를 작성하고 있다. 이때 용도별 가중치를 결정하는 방법에는 인구 비중법(Population proportion method), 표준 밀도법(Standard density method), 회귀분석법(Regression method)등이 주로 사용되고 있다. 인구 비중법은 전체 인구가 토지용도에 따라 얼마정도 살고 있는 지를 선험적인 자료를 바탕으로 추산하여 각 토지용도에 인구를 분포시키는 방식이다. 앞서 살펴본 Eicher 등[4]은 토지용도에 따른 가중치를 산정에 있어서 Holloway 등[6]의 선험적 연구에 기반을 두었다. 이와 유사하게 표준 밀도법은 각 토지용도에 따른 인구밀도를 선험적 자료를 바탕으로 산출하고 전체 토지용도의 인구밀도의 총합에 대한 비율을 계산하여 토지용도별 가중치를 구한다. Mennis 등[14]은 토지용도를 구분하고 각 토지용도가 잘 나타나는 전형적인 구역을 선정하여, 그 구역의 인구밀도를 산정하여 이를 각 토지용도별 표준인구밀도로 산출하여 가중치를 만들었다. 회귀분석법은 센서스 인구를 종속변수로 두고 토지용도별 거주하는 인구수를 독립변수로 두어 상수항이 없는 OLS 회귀분석을 실시한다. 이 과정을 통해 구한 회귀계수는 각 토지용도별 인구밀도로 해석할 수 있다(이상일 외[2]).

이러한 대시메트릭 매핑 기법은 계속 연구되고 있으며 기존의 기법들을 응용하여 연구자만의 새로운 대시메트릭 매핑 기법이나 하이브리드 매핑 기법이 만들어지기도 한다. 복합적인 매핑기법은 기존의 세 가지 방식의 대시메트릭에 의한 인구분포를 보다 연속적이고 부드럽게 하여 실제에 가까운 인구분포를 나타내 정확도를 높이고자 하는 노력이다. Kim 등[9]은 바이너리 기법과 체적보존 보간법(Pycnophylatic method)을 복합적으로 사용하여 단위지역 내 총 인구를 유지하면서도 주변지역의 인구를 고현한 인구분포를 작성하였다. Su 등[15]은 작은 공간 유닛에서 보다 실제에 가까운 인구추정을 위해 다중-클래스 기법과 함께 바이너리기법, 토지용도에 따른 가중치를 반영하고 각 토지용도에 기반시설 밀도 등의 세밀한 정보들을 접목시킨 다층(Multi-layer)적인 기법을 동시에 활용한 MLMCD (Multi-Layer Multi-Class Dasymetric)을 고안하였다.

한편, 대부분의 연구들이 대시매트릭 매핑과정에서 토지용도 데이터를 보조데이터로 사용하고 있지만, 반드시 그럴 필요는 없다. 이병길[1]은 사업체의 분표

Table 1. Main and Ancillary data for dasymetric mapping

Classify	Data	Items	Year	Note
Main data	Census (Census Track)	Population	2010	census track 16,232 total population 9,409,255
Ancillary data	Building Ledger	Main Usage of Building, Building Floor Area, Number of floors	2011	678,595 buildings

Table 2. Summary of Ancillary data

Main Usage of Building	Number of buildings	Sum of bldg floor area (m^2)	Sum of bldg gross floor area (m^2)	Average of floors	/ Area of Seoul	/ Developable Area of Seoul
Residential	463,375	**55,288,729**	**251,494,165**	**4.5**	**9.1%**	**13.9%**
Detached	142,352	10,522,345	20,514,566	1.9	1.7%	2.6%
Multi-family	216,540	20,534,995	62,861,743	3.1	3.4%	5.2%
Multiplex	68,381	9,448,649	28,620,663	3.0	1.6%	2.4%
Low-rise apt	9,557	2,402,084	7,596,924	3.2	0.4%	0.6%
Apartment	26,545	12,380,656	131,900,269	10.7	2.0%	3.1%
Non-residential	215,220	44,754,520	168,830,780	3.8	7.4%	11.2%
Total	678,595	100,043,249	420,324,945	4.2	16.5%	25.1%
Area of Seoul		605,207,639	-	-	100%	-
Developable Area		398,249,497	-	-	-	100%

현황을 파악하기 위해 간선도로망정보를 사용하였고, Holt 등[7]과 Su 등[15]은 토지용도와 함께 교통 및 접근성 정보를 보조데이터로 이용하였다. Lwin 등[11]은 건물높이와 용도를 포함하고 있는 건축물정보(Building footprint dataset)를 보조데이터로 활용하여 더 작은 단위에서 효과적인 인구 측정이 가능토록 하였다.

3. 연구범위 및 분석자료

본 연구의 공간적 범위는 우리나라 인구의 20.16%가 거주하는 서울특별시로 한정한다. 본 연구에서는 데시매트릭 매핑을 위해 인구주택총조사(통계청, 2010)의 집계구별 센서스 인구자료를 주 데이터로, 서울시 건축물 현황도(2011) 자료를 보조 데이터로 이용하였다(Table 1).

인구센서스 데이터는 서울특별시 25개 구, 424개 행정동을 총 16,232개로 나눈 집계구[2)]를 바탕으로 인구데이터가 집계되어 있다. 서울시의 2010년 기준 총 인구수는 9,409,255명, 총 가구수는 3,470,025호로 집계되었다.

건축물 현황도는 서울시 내 총 678,595개 개별건축물 정보를 포함하는 자료로써, 건축물 주용도, 층수, 면적, 행정코드 등의 정보를 포함한다. 이 자료는 보조자료로 활용되어 건축물 주용도와 면적을 바탕으로 서울시 인구의 각 건축물별 분포를 결정한다(Table 2).

4. 연구방법

본 연구는 집계구 단위 인구센서스(2010)과 건축물 현황도(2011)을 활용하여 데시매트릭 매핑 기법을 이용하여 인구분포를 개별건물 단위에서 추정하는 것을 목적으로 한다. 이러한 정밀한 인구분포 모델을 도출하기 위하여 바이너리 기법, 연면적 적용기법(Volumetric method), 회귀분석 기법이 차례로 적용되었다. 각 단계는 인구분포에 대한 다음과 같은 가정으로부터 출발하여 발전되었다.

가정1. 모든 인구는 주거용 건축물에 거주한다.
가정2. 건축물에서의 인구밀도는 주택유형에 따라 다르게 나타난다.
가정3. 지역별 주택유형에 따른 인구밀도는 지역적 특성을 반영하여 다르게 나타난다.

2) 집계구(소지역)이란 통계조사의 편의상 자료를 집계하기 위한 구역으로 인구 약 500명 정도의 규모, 읍면동 크기의 약 1/25로 지정한 구역이다.(통계청 통계정보국 공간정보서비스과)

Avg. Pop. of census track: 580
Population of Census Track

Every Population live in buildings
Building Area

Removing Non-residential Bldgs
Residential Building Area

Figure 1. Concept diagram of Binary method

4.1 바이너리 기법

데시매트릭 매핑의 첫 번째 단계로 바이너리 기법을 적용하여 인구분포 재현 대상지를 주거가능 건축물로 한정하였다. Figure 1과 같이 현재 제공되는 인구통계 자료는 집계구를 기준으로하고 있다. 그러나 가정1 에서와 같이 현실에서 실제 인구가 거주하는 지역은 집계구 영역 내에서도 건축물에 한정된다. 본 연구에서는 인구가 주거용도 건축물에만 거주하는 것으로 가정하였다. 따라서 주 데이터인 총인구수의 분포영역을 건축물 현황도를 바탕으로 건축물과 그 외 지역으로 구분하고, 비거주지를 제거하였다. 즉, 주거용도의 건축물=1, 그 외(비주거 건축물 및 기타 지역)=0의 가중치를 부여하여 모든 인구가 주거용 건축물에만 분포되어 있는 모델을 추출하였다. 바이너리 기법에 의해 추출된 인구분포 재현 대상지는 서울시 면적의 약 9%, 개발가능지[3]만을 기준으로 할 경우에도 약 14% 정도에 해당하는 면적이다. 이는 인구밀도가 주거용 건축물 이외 지역에서까지 예측되었던 토지이용도 등을 활용한 기존 방법론에 비하여 현실에 가까운 결과를 제공한다.

4.2 연면적 적용 기법

앞서 바이너리 기법에 의하여 한정된 주거가능 면적에 해당 집계구의 인구가 균일하게 거주하고 있다고 가정하면, 주거용 건축물에 대한 단계구분도를 얻을 수 있다. 그러나 주거용 건축물의 바닥면적만으로 코로플레스 매핑을 할 경우에는 건축연면적이 고려되지 않아 실제 거주점유면적(사용면적) 규모에 따른 인구배분이 이루어지지 않는다. 건축물 바닥면적과 실제 사용면적의 합인 건축연면적 사이에 차이가 있으며, 동일 집계구 내에도 다양한 층수의 건축물이 존재하기 때문에 실제에 근접한 인구 배분을 위해서는 3차원적인 공간이 고려되어야 한다. 따라서 본 연구에서는 건물의 층수를 반영하여, 기존의 바닥면적만을 고려한 2차원 지도상에서 나타나지 않는 입체적 면적을 모델에 반영하였다.

4.3 회귀분석 기법(Regression method)

코로플레스 매핑의 경우, 모든 건축물에 동일한 밀도로 인구가 거주하고 있다는 것을 전제로 하고 있다. 그러나 현실에서는 모든 주거용 건축물에 균일한 밀도로 인구가 분포하지 않는다. 각 주택유형(단독주택, 다가구주택, 다세대주택, 연립주택, 아파트)은 주거특성에 따라 거주점유면적이 다르게 나타난다.

주거용 건축물의 유형별 인구밀도 가중치를 서로 다르게 부여하기 위하여 본 연구에서는 다중회귀분석을 이용하였다. 또한 이러한 인구밀도패턴은 각 지역 특성으로 인하여 지역적으로 차이가 있다고 전제하여 각 지역별로 다중회귀분석을 재차 진행하여 인구분포 예측모델을 정교화 하였다.

각 주택유형별 특성을 반영한 가중치를 도출하기 위한 모델을 다음과 같다.

3) 환경부(2009) 토지피복도 자료에 근거하여 하천, 호수, 해수등의 수계 지역과 골프장, 활엽수림, 침엽수림, 혼효림, 기타초지 등의 녹지를 제외한 행정동별 면적을 계산하여 개발 가능 지역(Developable area)로 정의하였다.

$$CPop_i = \sum_{j=\text{단독주택}}^{\text{아파트}} (PDen_j \cdot \sum GFA_{hj}) + 0$$

h : 건물 ID
i : 집계구 ID
j : 주택유형(단독주택, 다가구주택, 다세대주택, 연립주택, 아파트)
$CPop_i$: i 집계구의 인구수
$PDen_j$: 주택유형 j 의 평균 인구밀도
$\sum GFA_{hj}$: i 집계구 내 j 유형 건물 연면적의 합

이는 다음과 같이 회귀모델의 형태로 표현될 수 있다.

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + 0$$

Y : 16,232개 집계구별 인구수
$X_1 \sim X_5$: 집계구 주택유형별 연면적의 합
$\beta_1 \sim \beta_5$: 주택유형별 인구밀도

종속변수는 서울시 16,232개의 집계구 인구수, 독립변수는 각 집계구 주택유형별 연면적의 합으로 설정한다. 이러한 다중회귀분석을 통하여 도출된 베타값($\beta_1 \sim \beta_5$)은 각 주택유형별 특성을 반영한 가중치로써 해당 주택유형에 살고 있는 단위면적당 인구수, 즉 인구밀도를 의미한다. 회귀분석을 통해 보조 데이터(건

Table 3. Result table of population density(β)

Region	Population Density (pop/m^2)					number of cases	R^2
	단독주택	다가구주택	다세대주택	연립주택	아파트		
Seoul	**0.012054*** **	**0.047081*** **	**0.052808c**	**0.029546*** **	**0.014227*** **	16232	0.7242
Jongno	0.003589	0.062685***	0.06102***	0.026574***	0.014129***	260	0.8178
Jung	0.034783**	0.067484***	0.047878***	0.022593	0.012365***	222	0.7186
Yongsan	-0.00136	0.04595***	0.06229***	6.48E-05	0.01268***	377	0.7448
Seongdong	0.044055***	0.045004***	0.051813***	0.036973***	0.023694***	503	0.7995
Gwangjin	0.026876*	0.044047***	0.043447***	0.039308***	0.006601***	605	0.8174
Dongdeamun	0.047823***	0.043495***	0.041148***	0.039329***	0.019009***	611	0.7922
Jungnang	0.027179***	0.04578***	0.045357***	0.040041***	0.010102***	677	0.8094
Seongbuk	0.004254*	0.061172***	0.058008***	0.03739***	0.012769***	747	0.6922
Gangbuk	0.030918***	0.043504***	0.063956***	0.027847*	0.018029***	526	0.8262
Dobong	0.053639***	0.043975***	0.049849***	0.063717***	0.02533***	609	0.8069
Nowon	0.073315***	0.048543***	0.048386***	0.044814**	0.02455***	1023	0.7328
Eunpyeong	0.028636***	0.043873***	0.042712***	0.031308***	0.018729***	718	0.8600
Seodeamun	0.010221***	0.054804***	0.036554***	0.032485***	0.021212***	521	0.8438
Mapo	0.029919***	0.046001***	0.044065***	0.020465*	0.015114***	614	0.7610
Yangcheon	0.030613***	0.042416***	0.053006***	0.062459***	0.009772***	778	0.6898
Gangseo	0.028451***	0.040045***	0.060367***	0.039725***	0.021445***	921	0.8085
Guro	0.031815**	0.04313***	0.056024***	0.032646***	0.021804***	963	0.7838
Geumcheon	0.056117**	0.037693***	0.053037***	0.032841***	0.029839	392	0.8487
Yeongdeungpo	0.077645***	0.03617***	0.05325***	0.040458**	0.015002***	705	0.7042
Dongjak	0.033013***	0.04637***	0.04254***	0.037649**	0.013321***	621	0.8050
Gwanak	0.04822***	0.041675***	0.049212***	0.059378***	0.021983***	824	0.8244
Seocho	0.01493***	0.040528***	0.055002***	0.014942***	0.014409***	663	0.7286
Gangnam	0.014151**	0.041959***	0.031335***	0.030306***	0.016925***	858	0.6733
Songpa	0.038603	0.050103***	0.05333***	0.070774***	0.006598***	1038	0.6002
Gangdong	0.025811	0.041358***	0.058477***	0.034703***	0.021143***	729	0.7775

p-value: '***'(< 0.001) '**' (< 0.01) '*' (< 0.05)

Table 4. Summary of population density

Region	Regression level	Number of Cases / Population (avg.)	R2	Note
Seoul (Si/do level)	Census track	16,232 / 580	0.724	
Gu (Si/gun/gu level)	Census track	660 (avg.) / 580	0.771 (avg.)	Max: 0.849 (Geumcheon) min: 0.600 (Songpa)

축물 현황)에서 도출한 각 집계구의 주택유형별 연면적의 합에 따른 주 데이터(집계구의 인구수)를 설명하는 주택 유형별 인구분포 모델을 도출하였다. 여기서, 특정 집계구 내의 주택유형에 따른 유형별 가중치(주택유형별 평균 인구밀도)와 유형별 연면적의 곱을 모두 더한 값이 그 집계구의 인구수가 된다.

본 연구는 건물단위 인구수(*BPop*)를 반영하는 인구분포 모델을 추정하는 것이다. 이는, 각 건축물의 연면적(*GFA*)에 회귀분석을 통해 도출한 해당 주택유형별 인구밀도($PDen_j$)를 곱하여 알 수 있다. 이렇게 추정된 개별 주거용 건축물의 거주 인구를 GIS프로그램을 이용하여 건축물과 연계하여 서울시 전체 개별 주거용 건축물 단위의 인구(또는 인구밀도)를 매핑할 수 있다.

앞서 회귀분석 기법 주택유형별 가중치를 도출, 주택유형사이에 존재하는 특징적인 인구밀도의 차이를 반영하였다. 그러나 동일한 주택유형이라 할지라도 지역적인 특성에 따라 그 연면적에 대한 인구밀도의 차이가 존재한다. 예를 들어 서울시 단독주택의 평균 인구밀도는 0.0121(인/m^2)로 4인가구라면 평균 330m^2 면적의 단독주택에 살고 있는 것으로 예측할 수 있으나, 현실세계에서는 그보다 더 넓거나 좁은 단독주택에 살고 있는 경우가 존재하며, 동일한 면적에 더 많거나 적은 수의 인구가 거주하기도 한다. 따라서 본 연구에서는 이 같은 차이가 지역적 특성에 따른 것으로 가정하고, 앞서 정의한 회귀분석 기법을 각 지역별로 개별 적용하여 결과를 도출하였다.

$$CPop_i = \sum_{j=\text{단독주택}}^{\text{아파트}} (PDen_{jk} \cdot \sum GFA_{hj}) + 0$$

h : 건물 ID
i : 집계구 ID
j : 주택유형(단독주택, 다가구주택, 다세대주택, 연립주택, 아파트)
$CPop_i$: i 집계구의 인구수
$PDen_{jk}$: k 지역 주택유형 j의 평균 인구밀도
$\sum GFA_{hj}$: i 집계구 내 j 유형 건물 연면적의 합

주택유형별 인구밀도의 지역별 차이를 반영하기 위해서는 지역적 특성을 공유하는 보다 작은 공간 단위에 있는 집계구들을 대상으로 각 사례에 맞는 회귀분석이 진행되어야 한다. 집계구보다 상위의 행정경계로는 행정동과 자치구 경계가 존재한다. 행정동을 회귀분석의 공간단위(모집단)로 삼을 경우 평균 38개의 집계구만이 포함되어 사례수 부족으로 결과에 대한 신뢰도가 떨어지는 문제가 발생한다. 따라서 본 연구에서는 서울시 내 25개 자치구를 기준으로, 자치구별로 회귀분석을 시행하여 총 25개의 주택유형에 따른 인구분포 모델을 도출하였다.

5. 연구결과

5.1 주택 유형별 인구밀도

서울시 16,232개의 집계구의 주택유형에 따른 인구밀도 값은 앞서 시행한 회귀모델분석의 결과로 Table 3과 같이 도출 되었다. 서울시의 건축물연면적에 대한 인구밀도는 주택유형에 따라 단독주택 0.0121(인/m^2), 다가구주택 0.0471(인/m^2), 다세대주택 0.0528(인/m^2), 연립주택 0.0295(인/m^2), 아파트 0.0142 (인/m^2)로 나타났다. 이러한 각 회귀계수들은 모두 높은 신뢰수준에서 산출되었다.

결과에 따르면, 인구분포패턴에 대한 초기의 가정과 같이 주택유형에 따른 인구밀도의 차이가 존재함을 확인할 수 있다. 주택 연면적에 대한 인구밀도는 경험적으로 예상되는 바와 같이, 다가구주택과 다세대주택에서 높게 나타나 단독주택이나 아파트의 경우보다 집약적으로 이용되고 있음이 확인되었다(Table 4).

이에 더하여, 각 자치구별로 시행한 두 번째 회귀분석의 결과는 Table 3의 자치구별 결과값과 같이 각 지역별로 주택유형별 인구밀도의 값(β 값)이 다르게 나타났다. 따라서 모델에 지역적 특성을 고려했을 경우가 그렇지 않았을 경우보다 현실에 근접한 인구분포 재현 모델을 설정하는데 도움이 됨을 알 수 있다. 이것은 주택유형별 인구밀도의 차이가 지역적 특성에

따라 달라진다는 초기의 가정을 반증하고 있다. 예를 들어, 단독주택의 경우 서울시 전체 집계구를 대상으로 시행한 회귀분석의 결과에서는 0.0121(인/m^2)로 나타나 연면적 330m^2의 단독주택에 약 4명이 거주하고 있는 것으로 예상할 수 있다. 그러나 영등포구의 경우 약 26명, 성북구의 경우 약 2명이 거주하고 있을 것으로 예측되는 등 지역에 따른 차이가 두드러진다. 또한 아파트의 경우에도 서울시의 평균인구밀도가 0.0142(인/m^2)로 4인 가족이 건축물 연면적을 기준으로 평균 280m^2의 공간을 사용하고 있는 것으로 예상할 수 있으나, 광진구와 송파구의 경우 약 600m^2, 금천구의 경우는 약 134m^2로 차이를 보인다. 자치구 내에 특정 주택유형이 너무 적게 분포하거나, 인구밀도의 편차가 큰 소수의 경우를 제외한 대부분의 경우에서 높은 신뢰수준에서 지역적 특성을 반영한 주택유형별 연면적 대비 인구밀도가 도출되었다. 모델의 설명력에 있어서도 서울시 전체를 대상으로 한 인구분포 추정 모델의 경우(R^2=0.724)보다 서울시 25개 자치구별로 나누어 시행한 경우(R^2=0.771)가 평균적으로 더 우수한 것으로 나타났다. 따라서 본 연구에서는 자치구를 중심으로 지역별 가중치를 고려한 모델을 최종적으로 선택하여 인구분포를 매핑 하였다.

5.2 인구 매핑(Mapping) 및 결과비교

본 연구의 결과로 도출된 최종적인 인구분포모델에 의해 추정된 인구는 지도상에서 Figure 6과 같이 표현된다. 결과의 비교를 위하여 방법론의 각 단계를 시행한 결과를 함께 함께 제시하였으며, 각 인구분포지도에 대한 인구분포 추정모델은 다음과 같다.

(2D Choropleth)

$$BPop_h = PDen_{average} \cdot A_h$$

(3D Choropleth)

$$BPop_h = PDen_{average} \cdot GFA_h$$

(지역별 차이를 반영하지 않은 인구분포 모델):

$$BPop_h = PDen_j \cdot GFA_h$$

(지역별 가중치를 반영한 인구분포 모델)

$$BPop_h = PDen_{jk} \cdot GFA_h$$

$BPop_i$: 주거용 건축물 h 에 거주하는 인구수
$PDen_j$: h 건물의 주택유형 j 의 인구밀도
$PDen_{jk}$: k 지역 h건물의 주택유형 j 의 인구밀도
A_h : h 건물의 바닥면적
GFA_h : h 건물의 연면적

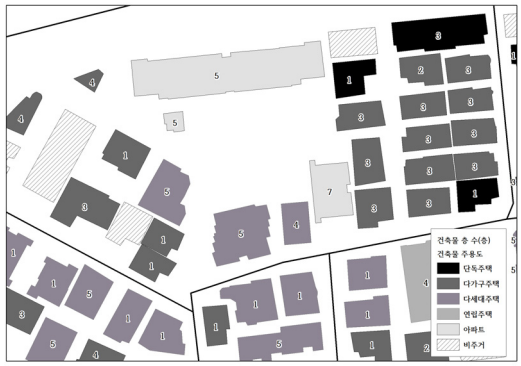

Figure 2. Housing types and floors (part of Ssangmun3 dong, Dobong)

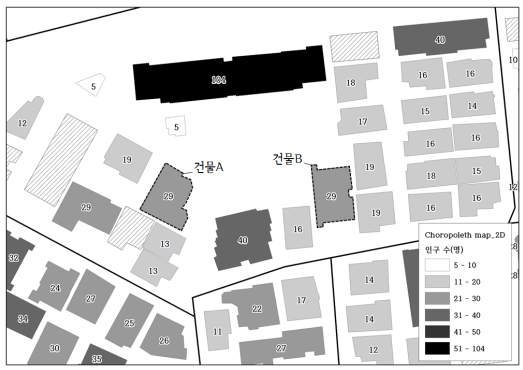

Figure 3. Choropleth map (using floor area)

Figure 2은 주자료인 집계구 단위의 2010년 인구센서스를 지도에 표현한 것으로 집계부보다 세밀한 공간 영역에 존재하는 인구를 다루는 연구에서는 한계를 지닌다. Figure 3는 보조자료를 활용하여 바이너리 기법으로 인구거주 영역을 주거가능 건축물에 한정시킨 단계구분도이다.

Figure 3는 인구의 공간적인 분포 위치에 있어서는 기존의 인구통계자료에 비해 현실에 가까운 결과를 제공하지만, 현실세계의 입체적인 공간 사용이 무시된 결과이기 때문에 개별 위치에서 추정되는 인구의 수가 왜곡되어있을 가능성이 크다. 이러한 왜곡의 존재는 연면적 적용기법를 이용하여 건물의 층수를 건축물 바닥면적에 대한 가중치로 적용하여 건물의 입체적인 활용을 고려한 Figure 4와의 차이를 통해 확인할 수 있다. 예를 들어, Figure 5와 Figure 6는 '건물A'와 '건물B'에 대해 서로 다른 거주인구 예측결과를 보이고 있다. 건축물의 바닥면적은 유사하지만 각각

Figure 4. Choropleth map (using gross floor area)

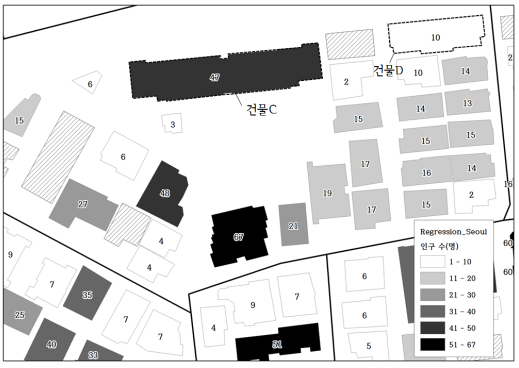

Figure 5. Dasymetric map (using variable of housing type)

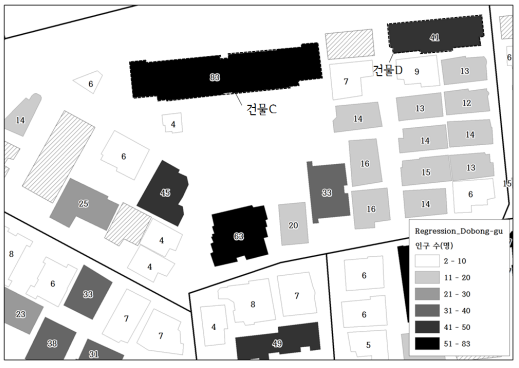

Figure 6. Dasymetric map (using variable of housing type & regional variable)

5층과 7층으로 연면적에서 차이를 보이는 두 건물에 대해 29명으로 동일한 인구예측을 보인 Figure 4의 경우 연면적을 반영하여 각각 40명과 57명이라는 인구예측결과를 보인 Figure 5에 비해 왜곡이 크게 발생할 수 있음을 알 수 있다.

Figure 5는 회귀분석 기법을 통하여 Figure 4의 단계구분도로부터 주택유형별 가중치를 적용한 결과이다. 더 나아가, Figure 6는 지역적 특성에 따른 차이를 추가로 반영한 결과이다. 따라서 Figure 6는 본 연구의 최종 결과물로 지역적 특성이 반영된 주택 유형별 인구밀도가 고려되어 가장 현실에 가깝게 예측된 인구분포 결과를 보여준다. 예를 들어, 아파트 '건물C'와 단독주택 '건물D'에 거주하는 인구는 Figure 5에서 각각 47명과 10명으로 예측되었다. 이 같은 예측 값은 건물이 위치한 도봉구의 주택유형별 인구밀도 특성을 반영하지 못한 것으로 현실과는 차이를 보인다. 도봉구의 단독주택 연면적에 대한 인구밀도는 서울시 평균의 4배 이상으로, 서울시 평균에 비해 단독주택이 집약적인 형태로 사용되고 있는 특성이 있다. 이와 유사하게 아파트인 '건물C'의 경우에도 Figure 5의 예측값은 현실과 거리를 보인다. 따라서 본 연구에서는 Figure 6와 같이 주택유형과 지역적 특성을 모두 반영한 경우에 가장 실제에 가까운 인구분포 모형을 얻을 수 있었다.

6. 결 론

6.1 서울시 건물단위 인구밀도의 재현

본 연구는 서울시를 대상으로 인구센서스 데이터를 주 데이터로, 건축물 현황도를 보조 데이터로 활용하여 대시메트릭 매핑을 수행하였다. 기존에 제공되는 인구데이터의 가장 작은 공간단위인 집계구(서울시 16,232개)를 바탕으로 바이너리 기법, 연면적 적용기법을 적용하여 모든 인구는 주거용 건축물에만 거주한다는 가정을 모델화하고 건축물의 층수를 고려한 3차원적인 매핑모델을 제안하였다. 또한, 주거용 건축물이 다섯 가지의 주택유형(단독주택, 다가구주택, 다세대주택, 연립주택, 아파트)으로 나뉘며 이에 따라 각각 다른 주거특성을 반영하는 것을 고려하였다. 이는 회귀분석 기법를 통하여 각 주택유형별 인구밀도(β값)을 도출하였으며, 도출된 주택유형별 인구밀도(β값)와 각 주택유형의 연면적의 곱을 통하여 개별 건축물별로 거주하는 인구를 추정할 수 있는 모델을 완성하였다. 이러한 연구를 통하여 서울시 16,232개 집계구를 바탕으로 한 서울시 전체 건축물별 인구분포모델과 각 지역의 주거 특성을 반영하는 각 자치구의 건축물별 인구분포모델(25개 케이스)을 재현하였다.

본 연구를 통하여 기존에 특정 공간단위를 바탕으로 재현하거나 대지 중심의 2차원적인 대시메트릭 매

핑의 한계를 보완하고자 실제로 인구가 거주하는 건축물에 해당하는 인구분포를 주거특성과 층수를 반영한 3차원적인 인구밀도분포를 모델화하였다. 이러한 모델을 통하여 기존의 인구분포모델에서는 과소 또는 과잉 추정되던 문제를 보완할 수 있으며 건축물 단위의 더 정밀한 인구 추정이 가능하다. 본 연구 결과를 토대로 인구정보분석에 기반을 둔 재난관리, 응급상황 대비, 범죄분석 등의 다양한 연구와 정책적 기초자료의 기반이 될 수 있을 것이다.

6.2 연구의 한계 및 향후 연구

그러나 본 연구는 기존의 집계공간단위(행정구역)를 기준으로 분석을 진행하여 행정구역이 아닌 지역적 특성이 발현되는 공간단위를 고려하지 못했다는 한계를 갖는다. 예를 들어, 각 자치구의 경계부에 위치한 행정동들은 해당 자치구의 반대편에 위치한 행정동보다 인접한 다른 자치구의 행정동과 유사한 특성을 지니는 것이 일반적이다. 이러한 공간적 경계를 넘어 존재하는 지역특성의 연속성은 본 연구에서 도출한 모델에는 반영되지 못하여 한계로 남는다.

이러한 한계의 극복을 위해 향후 공간적 특성을 반영할 수 있는 공간계량모델(Geographical Weighted Regression)을 통한 추가적인 연구가 요구된다. 공간계량모델에서는 (회귀분석을 위한 사례수의 한계를 넘어) 지역적 특성이 발현되는 공간적 범위를 기존 행정경계가 아닌, 인접한 지역들 사이에서 자기상관성을 가장 잘 보이는 영역으로 활용하기 때문에 이러한 문제를 극복할 수 있는 유용한 방안으로 판단된다. 따라서 이후 보다 실제에 가까운 인구분포 재현을 위한 추가적인 연구가 진행되어야 할 것이다.

References

[1] Lee, B. K. 2006, An Evaluation of Spatial Interpolation of Statistical Information Using Dasymetric Mapping, Journal of the Korean Society of Surveying, Geodesy, Photogrammetry, and Cartography 24(4): 343-350.

[2] Lee, S. I; Kim, K. Y. 2007, Representing the Population Density Distribution of Seoul Using Dasymetric Mapping Techniques in a GIS Environment, Journal of the Korean Cartography 7(2): 53-67.

[3] Bajat, B; Krunić, N; Kilibarda, M. 2011, Dasymetric mapping of Spatial distrivution of Population in Timok region, International Science Conference and XXIV Meeting of Serbian Surveyors.

[4] Eicher, C ; Brewer, C. 2001, Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation, Cartography and Geographic Information Science, 28(2): 125-138.

[5] Fisher, P ; Langford, M. 1996, Modeling Sensitivity to Accuracy in Classified Imagery: A Study of Areal Interpolation by Dasymetric Mapping, Professional Geographer, 48(3): 299-309.

[6] Holloway, S. R ; Schumacher, J. ; Redmond, R. 1996, People and place: Dasymetric mapping using ArcInfo, Missoula: Wildlife Spatial Analysis Lab, University of Montana.

[7] Holt, J; Lo, C. P ; Hodler, T. W. 2004, Dasymetric Estimation of Population Density and Areal Interpolation of Census Data, Cartography and Geographic Information Science, 31(2): 103-121.

[8] Kim, H. H. 2007, Comparison of Three Dasymetric Methods for Population Density Mapping, The Geographical Journal of Korea, 41(4): 411-419.

[9] Kim, H. H ; Choi, J. M. 2011, A Hybrid Dasymetric Mapping for Population Density Surface using Remote Sensing Data, Journal of the Korean Geographical Society, 46(1): 67-80.

[10] Langford, M ; Unwin, D. J. 1994, Generating and mapping population density surfaces within a geographical information system, The Cartographic Journal 31: 21−26.

[11] Lwin, K. K ; Murayama, Y. 2010, Development of GIS Tool for Dasymetric Mapping, International Journal of Geoinfomatics, 6(1): 11-18.

[12] MacEachren, A. M. 1994, Some truth with maps: A primer on symbolization ans design. Washington DC: Association of American Geographers.

[13] McCleary, G. F. 1984, Cartography, geography, and the dasymetric method. In: Proceeding, 12th Conference of International Cartographic Association, August 6-13, Perth, Australia, 1: 599-610.

[14] Mennis, J ; Hultgren, T. 2006, Intelligent Dasymetric Mapping and Its Application to Areal Interpolation, Cartography and Geographic Information Science, 33(3): 179-194.

[15] Su, M. D ; Lin, M. C ; Hsieh, H. I. ; Tsai, B. W. ; Lin,

C. H. 2010, Multi-layer multi-class dasymetric mapping to estimate population distribution, Science of the Total Environmental, 408: 4807-4816.

[16] Wright, J. K. 1936, A method of mapping densities of population, The Geographical Review 26: 103-110.

논문접수 : 2014.5.9
수 정 일 : 2014.6.29
심사완료 : 2014.6.30