

DMSP-OLS 야간 영상을 이용한 인구 추정 모델 변수 선정 연구

유수홍 · 한수희 · 허 준 · 손홍규[†]

연세대학교 공과대학 사회환경시스템 공학부

Variable Selection for Estimating Population Using DMSP-OLS Night-time Image

Yoo, Su Hong, Han, Soo Hee, Heo, Joon, and Sohn, Hong Gyo[†]

Dep. of Civil Engineering, Yonsei University

Abstract : It may be an important issue to estimate population of a concerned country. In this study, an appropriate variable was selected to establish a model which fits best the relationship between the night time imagery of DMSP-OLS and population data. Exponential model was selected which was proposed by previous study. Accuracy validation was also performed for each variable extracted from the night time imagery of DMSP-OLS. Consequently, the model showed high accuracy when applied to the area of a certain amount of light was existed. However, further consideration should be necessary when to applied other country or other part of regions.

Key Words : DMSP-OLS, Night-time Image, Population.

요약 : 한 국가에 거주하고 있는 인구수를 파악하는 것은 매우 중요한 사안 중 하나이다. 본 연구에서는 DMSP-OLS 야간 영상과 인구 통계치를 이용한 인구 추정 모델을 제작함에 있어 적합한 변수를 선정하였다. 모델 제작에 이용된 기본식은 선행 연구에서 제안한 지수식을 바탕으로 제작하였으며 야간 영상으로부터 취득할 수 있는 다양한 변수를 이용하여 각각에 대해 정확도 평가를 수행하였다. 연구 결과, 일정 크기 이상의 빛이 존재하는 지역의 면적을 변수로 사용하여 인구를 추정하였을 때 가장 정확도가 높은 것으로 나타났다. 한편, 본 연구를 다른 국가나 지역에 적용시키기 위해서는 보다 추가적인 연구가 필요할 것으로 사료된다.

1. 서론

한 국가의 인구를 파악하는 것은 정책 수립 및 수행을 위해 매우 중요한 작업 중 하나이다. 이를 위해 통계청에서는 5년을 주기로 행정 구역별, 성별, 연령별로 인구 통계치를 작성하고 있다. 한편, 중국이나 미국의 경우 매우 많은 수의 인구가 거주하고 있으며 그 조사 면

적 또한 매우 넓다. 이에 DMSP-OLS (Defense Meteorological Satellite Program-Operational Linescan System) 위성으로부터 취득한 야간 영상을 이용하여 인구를 추정하기 위한 모델을 제작하는 연구가 수행된 바 있다(Amaral 2005; Lo, 2001; Tobler, 1969). DMSP는 미국의 공군에서 수행하고 있는 국방 프로그램 중 하나로서, 다양한 센서를 탑재한 위성을 이

접수일(2011년 2월 16일), 수정일(1차 : 2011년 2월 19일), 게재확정일(2011년 2월 19일).

[†] 교신저자: 손홍규(sohn1@yonsei.ac.kr)

용하여 기상이나 해양 등의 모니터링을 수행한다. 그 중 OLS라는 센서를 이용해 밤영상을 취득한다. Tobler (1969)는 건물이 밀집하여 분포하고 있는 지역과 동일한 면적을 가진 원의 반지름이 인구수와 높은 상관관계를 가지는 점에 착안하여, 밤영상을 이용해 인구 추정 모델식을 제작하였다. Lo (2001) 역시 Tobler가 제안한 지수함수를 이용하여 인구 추정 모델을 제작하였으며, Amaral(2005)은 회귀분석식을 통하여 인구 추정 모델을 제작하였다.

한편, 선행 연구에서는 빛이 분포하고 있는 지역의 면적과 인구수가 높은 상관관계를 가진다는 선행 연구에 기반을 두어, 빛이 존재하는 지역의 면적만을 독립 변수로 사용하였다. 그 결과, 다양한 모델 식을 사용하였음에도 불구하고 오차가 굉장히 큰 것으로 나타났다. 실제 Lo (2001)가 수행한 연구에서는 평균 100% 이상의 오차를 보여 인구 추정 모델로써 무리가 있는 것으로 판단된다. 따라서 신뢰성이 높은 모델을 제작하기 위해서는 다양한 모델이 아닌 인구와 상관관계가 높은 변수를 선정하는 연구가 필요하다.

본 연구에서는 DMSP-OLS의 야간 영상을 통해 취득할 수 있는 여러 가지 변수들을 이용하여 남한의 인구 추정 모델 제작하였다. 인구 추정 모델은 선행 연구를 통해 입증된 지수식을 사용하였으며, 변수는 각 행정구역당 면적, 밝기값의 총합, 일정 크기 이상의 빛이 존재하는 지역의 면적, 광원 분포비, 일정 크기 이상의 광원 분포비를 사용하여 정확도 평가를 수행하였으며, 이 후 도출된 결과를 비교 분석하였다.

2. 연구 대상지 및 데이터

1) 연구 대상지

본 연구의 대상지는 남한 전 영역으로서 북위 33° 07' 01" ~ 38° 36' 31", 동경 124° 36' 49" ~ 130° 55' 49"에 해당되며, 총면적은 99,693km²이다(Fig. 1). 많은 수의 인구가 특정 도시에 밀집하여 거주하고 있으며, 지역 간에 개발 격차가 심하게 나타나고 있음을 영상의 밝기 값으로부터 알 수 있다.

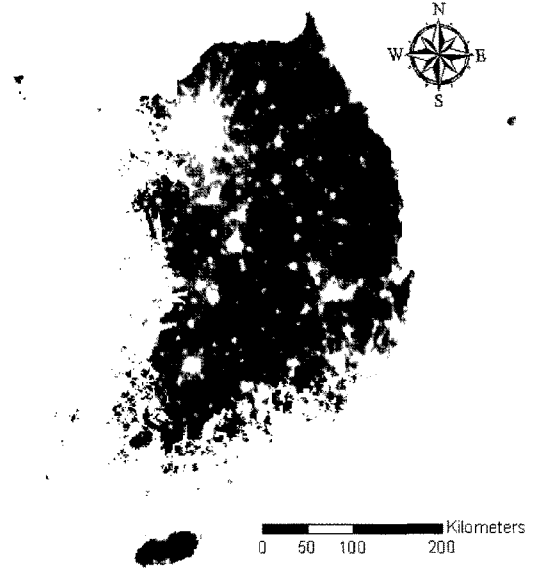


Fig. 1. DMSP imagery over Korean Peninsula.

2) 인구통계자료와 DMSP 영상

본 연구에서 이용한 자료는 총 두 가지로 인구수 통계 자료와 DMSP-OLS 야간 영상이다. 인구수 통계 자료는 통계청의 국가통계포털 시스템을 통하여 취득하였으며, 인구, 가구, 고용, 임금, 물가, 보건, 교통, 정보통신 등 다양한 통계치를 사용할 수 있다. 본 연구에서 사용한 인구수 통계 자료의 경우, 1925년부터 5년을 주기로 2005년도까지 자료가 구축되어 있으며, 행정구역별, 연령별, 성별로 정리되어 있다. Table 1은 통계청에서 제공받은 각 행정 구역별 인구수를 나타낸 것이다

DMSP는 미국의 국방부에서 운영하는 프로그램으로 미공군 미사일 센터에서 제어하고 있다. 해양과 기상, 태양에 대한 지면의 물리적 관계를 관측하기 위한 목적으로 발사되었으며, NOAA(National Oceanic And Atmospheric Administration)의 EOG(Earth Observation Group)에서 제공하는 웹서비스를 통해 영상을 취득할 수 있다(NOAA /NGDC). 야간 영상은 도시나 지속적으로 빛이 존재하는 지역에 대한 정보만을 가지고 있으며, 일시적으로 발생한 빛에 대해서는 보정되어 제공된다. 해당 영상은 1km × 1km의 해상도와 0에서 63 사이의 밝기값을 가지며 현재 1992년부터 2009년에 대한 영상을 제공한다.

본 연구를 수행하기 위해서는 인구수 통계 자료가 조

Table 1. Population of each administrative district

행정구역	인구수 (단위: 천명)	
	2000	2005
부산광역시	3,655	3,513
충청북도	1,463	1,454
충청남도	1,840	1,879
대구광역시	2,474	2,456
대전광역시	1,366	1,439
강원도	1,485	1,461
광주광역시	1,351	1,414
인천광역시	2,466	2,518
제주도	513	531
전라북도	1,887	1,779
전라남도	1,994	1,815
경상북도	2,716	2,595
경기도	8,938	10,341
경상남도	2,971	3,041
서울특별시	9,854	9,763
울산광역시	1,012	1,045



(a) year 2000



(b) year 2005

Fig. 2. Night Time Imagery of DMSP-OLS.

사된 연도와 야간 영상이 취득된 연도가 동일한 것이 좋으며 이 조건을 만족하기 위해서는 1995년, 2000년, 2005년 자료만을 사용하였다. 한편, 울산광역시의 경우 1998년도에 지정되었으므로 자료의 통일성을 위해 1995년도 자료를 제외하여 2000년과 2005년에 취득된 자료만을 사용하였다(Fig. 2).

3. 인구 추정 모델 제작

1) 인구 추정 모델 선정

이전의 연구자에 의해 수행된 연구 결과에 따르면, 건물이 밀집하여 분포하고 있는 지역과 동일한 면적을 가진 원의 반지름이 인구수와 높은 상관관계를 가지는 것으로 나타났다(Tobler, 1969). 이에 Tobler는 식 (1)을 사용하여 인구 추정 모델을 제작하였으며, Lo (2001)도 Tobler의 연구 결과를 바탕으로 식 (2)를 이용하여 인구 추정 모델을 제작한바 있다.

$$r = aP^b \quad (1)$$

$$P = aA_{\text{built-up}}^b \quad (2)$$

여기서, r : 건물이 밀집한 지역의 면적과 동일한 원의 반지름,

P : 인구,

a 와 b : 산출될 계수,

$A_{\text{built-up}}$: 건물이 밀집한 지역의 면적,

두 연구에서는 지수식을 공통적으로 사용하고 있으며, 채효석 등(2003) 또한 야간 영상을 소개하는 과정에서 지수식에 대해 언급하고 있다. 본 연구에서는 사전 연구에서도 보편적으로 사용되고 있는 지수식을 기본 인구 추정 모델로 선정하였다. 한편, 지수식의 계수를 산출하기 위해 수치 해석적으로 접근하기 용이하도록 이를 선형화할 필요가 있다. 지수식의 경우, 양변에 로그를 취하면 간단하게 식을 선형화할 수 있는데, 식 (2)에서 양변에 로그를 취한 뒤 정리하면 식 (3)을 얻을 수 있다. 식 (3)에서 계산의 편의성을 위해 $\log a$ 을 α 로 치환하여 최종적으로는 식 (4)를 인구 추정 모델로 이용하였다.

$$\log P = \log a + b \log A \quad (3)$$

$$\log P = \alpha + b \log A \quad (4)$$

여기서, P : 인구, A : 독립 변수, b 와 α : 산출될 계수

2) 변수 선정

야간 영상을 통해 생성 가능한 변수들로, 각 행정구역당 빛이 존재하는 지역의 면적(A_i), 행정구역내 존재하는 빛(밝기값)의 총합(A_s), 일정 크기 이상의 빛이 존재하는 지역의 면적(A_h), 행정 구역의 면적 대비 빛이 존재하는 지역의 면적을 비율로 나타낸 광원 분포비(A_r), 일정 크기 이상의 광원 분포비(A_{hr})로 총 5가지의 변수에 대해 고려하였다. 이 중 일정 크기 이상의 빛을 가지는 지역은 그 기준에 따라 결과가 달라지므로, 모든 범위에 대해 기준치를 정하여 각각에 대해 정확도 평가를 수행하였다.

한편, 본 연구에서 선정한 변수 중 일정 크기 이상의 빛이 존재하는 지역의 면적은 기준치에 따라 결과값이 달라지므로 이에 대한 분석이 필요하다. 이에 각각의 영상소 값을 기준치로 선정하여 기준치를 변경해가며 각각에 대해 정확도를 산출하였다. 이후, 도출된 각 모델 중 가장 작은 오차를 가지는 기준치를 최종 인구 추정 모델 제작에 사용하였다. 일정 크기 이상의 광원 분포비 변수는 앞서 언급한 방법에 의해 도출된 기준치를 사용하여 인구 추정 모델 제작에 사용되었다.

3) 정확도 평가

인구 추정 모델의 평가는 Lo (2001)가 사용한 평균 상대 오차법(MRE, Mean Relative Error)과 평균 절대 오차법(MAE, Mean Absolute Error)을 통해 수행하였으며, 각각 식 (5)와 식 (6)과 같다.

$$MRE = \frac{1}{n} \sum_{i=1}^n \left(\frac{A_i - F_i}{A_i} \right) \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (6)$$

단, n: 총 행정구역의 수(16개), A_i : 실제 값,
 F_i : 추정 값, i: 각 행정 구역

한편, MRE의 경우 MAE와 달리 음의 값을 보존하므로, 서로 비슷한 음의 값과 양의 값을 가지는 데이터가 존재할 경우 정확한 평가가 이뤄지지 않을 가능성이 있다.

4. 연구 결과

본 연구는 상용 소프트웨어인 Matlab을 이용하여 프로그래밍을 수행하였으며, 각 변수에 대한 인구 추정 모

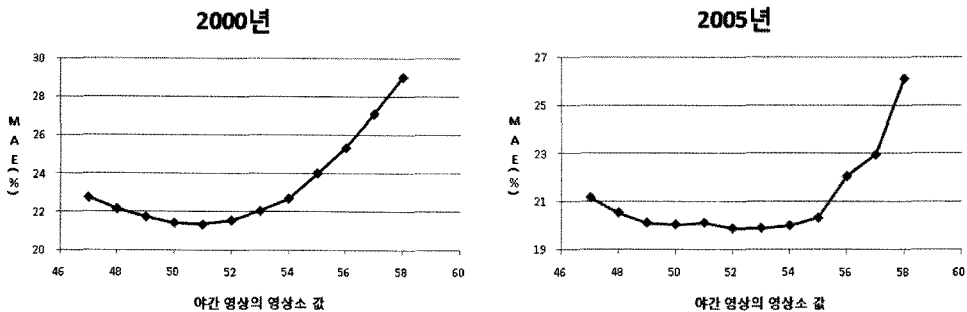


Fig. 3. Trend of accuracy within change criteria using MAE.

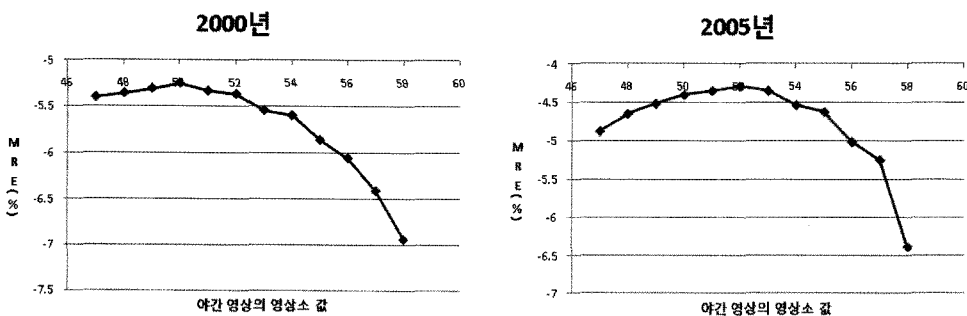


Fig. 4. Trend of accuracy within change criteria using MRE.

Table 2. Deducted model for population and accuracy

년도	인구 추정 모델	정확도
2000	$\log P = 1.647 + 0.340 \log A_s$	MRE= -18.64 %, MAE= 50.41 %
	$\log P = 1.671 + 0.661 \log A_h$	MRE= -5.33 %, MAE= 21.34 %
	$\log P = 2.543 + 0.562 \log A_r$	MRE= -20.97 %, MAE= 55.42 %
	$\log P = 3.565 + 0.240 \log A_{hr}$	MRE= -20.43 %, MAE= 55.34 %
	$\log P = 3.214 + 0.034 \log A_l$	MRE= -25.68 %, MAE= 58.86 %
2005	$\log P = 1.742 + 0.320 A_s$	MRE= -20.01 %, MAE= 52.57 %
	$\log P = 1.400 + 0.755 \log A_h$	MRE= -4.29 %, MAE= 19.86 %
	$\log P = 2.546 + 0.554 \log A_r$	MRE= -21.44 %, MAE= 54.30 %
	$\log P = 3.563 + 0.246 \log A_{hr}$	MRE= -20.70 %, MAE= 54.40 %
	$\log P = 3.253 + 0.024 \log A_l$	MRE= -25.79 %, MAE= 59.36 %

델의 계수는 최소 제곱법을 이용하여 산출하였다.

3.2절에서 언급한 바와 같이 본 연구에서 선정한 A_h 변수의 경우 기준치에 따라 인구 추정 모델이 변하므로 변수를 사용하기에 앞서 기준치 선정에 대한 연구가 수행되어야 한다. Fig. 3에서 4는 기준치가 변화함에 따라 제작된 모델의 정확도를 도시한 것으로, 기준치를 50~52 사이로 선정했을 때 오차가 가장 작은 것으로 확인되었다. Table 2는 본 연구에서 고려한 모든 변수에 대해 제작된 인구추정 모델로 오차가 가장 작은 경우에 대해 나열한 것으로, 보다 신뢰성이 높은 MAE평가 방법을 기준으로 선정하였다.

실험 결과, A_l 변수를 사용하였을 경우, 다른 변수를 사용했을 때보다 가장 큰 오차를 가지는 것으로 나타났다. 반면, 본 연구에서 선정한 A_h 변수를 사용하여 인구 추정 모델을 제작하였을 경우, 가장 작은 오차를 가지는 것으로 나타났다. 이는 A_h 변수가 다른 변수에 비해 인구수와 높은 상관관계를 가지기 때문인 것으로 사료된다. 각 변수에 대해 인구수와의 상관관계(Pearson's Correlation Coefficient)를 산출한 결과, A_s 는 0.351, A_h 는 0.826, A_l 은 0.505, A_r 은 0.505, A_{hr} 는 0.500으로 A_h 가 가장 높은 상관관계를 가지는 것으로 나타났다.

5. 결론

본 연구는 야간 영상을 이용하여 남한에 거주하고 있는 인구를 추정할 수 있는 모델을 제작하고 이에 대한 정확도 평가를 수행하였다. 인구 추정 모델의 선정은 선

행 연구를 바탕으로 지수식을 선정하였으며, 이에 대한 정확도 평가 방법으로 MRE와 MAE를 이용하였다.

현재까지 야간 영상을 이용한 인구 추정 모델 연구가 수행되었지만, 정확도가 현저히 낮은 것으로 나타났다. 한편, 본 연구에서는 다양한 변수를 사용하여 인구수를 추정한 결과, 기존에 제안된 변수(A_l)보다 특정 기준치를 기준으로 인구 추정 모델을 제작하였을 때 더 높은 정확도를 가지는 것으로 나타났다. 이는 선행 연구에서 수행한 방법처럼 다양한 추정 모델을 이용하여 정확도 향상을 유도하는 것보다 연구 대상지에 적합한 변수 선정이 선행되어야 한다는 점을 증명할 것이라 판단된다.

한편, 본 연구에서 사용한 영상의 수가 적다는 점과 연구 대상지에 따라 기준치가 변할 가능성이 높다는 한계점이 있다. 본 연구의 성과물을 다양한 대상 지역에 적용하기 위해서는 많은 수의 영상과 추정 모델에 대한 통계 분석이 필요할 것으로 사료된다.

참고문헌

- 강인제, 김태중, 2007. 한국의 지역내총생산 작성체계에 관한 연구, 한국지방재정학회, 12(2): 149-174
- 채효석, 김광은, 김성준, 김영섭, 이규성, 조기성, 조명희, 2003. "환경원격탐사", 시그마프레스, 451-452
- Amaral, S., Monteiro, A. M. V., Camara, G., Quintanilha, J. A., 2005. DMSP/OLS night-

time light imagery for urban population estimates in the Brazilian Amazon, *International Journal of Remote Sensing*, 27(5): 855-870.

Lo, C. P., 2001. Modelling the population of China using DMSP Operational Linescan System

nighttime data, *Photogrammetric engineering & Remote sensing*, 67(9): 1037~1047.

NOAA/NGDC - <http://www.ngdc.noaa.gov/dmsp/dmsp.html>

Tobler, W. R., 1969. Satellite confirmation of settlement size coefficients, *Area*, 1(3): 30-34.