

도시환경변수를 이용한 격자 인구추정에 있어서의 유전적 알고리즘기법 활용 연구

최내영^{1*}

An Application of the Genetic Algorithm on Population Estimation Using Urban Environmental Factors

Nae-Young CHOEI^{1*}

요 약

본 논문에서 활용한 유전적 알고리즘(GA)은 격자형 인구분포자료를 기반으로 인구추정 모형을 구축하는 데 적용되고 있는 일종의 인구표면모형 중 하나이다. 본 연구는 신도시 개발 등이 활발히 진행되고 있는 화성시 동부권을 대상지로 하여 격자형 인구분포자료를 구축하고 도시환경변수를 이용해 GA 인구설명 모형을 추정하였다. 분석결과 기반시설, 지역공공서비스시설 등은 이미 동부권 내에 공간적으로 균등 분포하는 관계로 설명력이 없는 것으로 나타난 반면 표고, 경사 등의 자연환경변수와 개발촉진 및 규제 관련 용도지역지구 등의 변수에 대해서는 매우 설득력 있는 계수 값들이 도출되었다. 결과적으로 본 연구에서 활용한 GA모형은 적자생존의 유전학적 원리를 기반으로 한 매우 직관적이고 학제적인 적정해 탐색 기법 중 하나임을 확인할 수 있었다.

주요어 : 유전적 알고리즘, 인구추정, 도시환경, 한국토지정보체계, 건축행정시스템

ABSTRACT

The Genetic Algorithm has been frequently applied by many researchers as one of the population surface modelling tool in estimating the regional population based on the gridded spatial system. Taking the East-Hwasung area as the case, this study first builds a gridded population data based on the KLIS and eAIS databases as well as municipal population survey data, and then constructs the attribute values of the explanatory variables by way of GIS tools. The GA model is run to maximize its fitness function measuring the correlation coefficient between the observed and predicted values of the 70 population cells. It is shown that the GA output predicted reasonably consistent and meaningful coefficient estimates for the explanatory variables of the model.

2010년 7월 20일 접수 Received on July 20, 2010 / 2010년 8월 30일 수정 Revised on August 30, 2010 / 2010년 9월 5일 심사완료 Accepted on September 5, 2010

1 홍익대학교 건설도시공학부 Department of Urban Engineering, Hongik University

* 연락처 E-mail : nychoei@hongik.ac.kr

KEYWORDS : Genetic Algorithm, Population Estimation, Urban Environment, KLIS, eAIS

서 론

지역 공간의 면적에 따른 실제 인구분포를 어떻게 가능한 한 정확히 예측, 추정하는가의 문제는 도시 및 지역정책 수립에 있어 매우 중요한 쟁점 중 하나이다. 현재 정례 센서스와 같은 인구 총 조사뿐만 아니라 다수 지자체의 부정기적 지역 인구조사와 같은 모든 주요 인구자료는 행정구역 혹은 별도의 인구조사 구역(census tract) 등 특정 공간영역을 기반으로 조사, 집계되고 있다. 따라서 이러한 대부분의 현행 인구자료들은 인구를 기초로 한 주요 사회경제적 정책수립이나 물리적 도시시설 계획 등에 있어 지역적 특수성과 다양성을 포함한 조사구역 간의 미시적 특이점들을 적절히 차별화하기 어려운 것이 현실적 제약임을 간과할 수 없다.

이러한 맥락에서 인구표면모형(population surface model)은 임의의 좌표체계를 설정하고 지역공간을 그에 따른 정규 격자형으로 분해하여 비정형적 조사구역 공간의 인구를 세부 격자에 수치적으로 공간 내삽(areal interpolation)하는 일련의 방식을 지칭하는 것으로 특정 위치에서의 인구상황을 보다 정확히 반영하게 되어 기존 인구자료의 공간적 제약을 극복하는 데 있어 매우 유용한 대안이라 할 수 있다(Yue *et al.*, 2003). 이와 같은 격자형 인구자료의 장점은 1) 필요시 어떠한 공간형태로도 재편이 가능하며, 2) 이질적(heterogeneous) 공간단위로 작성된 원천자료간의 호환성(compatibility)을 확보할 수 있을 뿐만 아니라, 3) 인위적인 행정경계 기준에서 발생하는 정치적 이해관계를 회피할 수 있다는 장점이 있다(Yue and Song, 2008). 본 연구에서도 같은 맥락에서 대상지 행정구역 단위로 작성된 인구자료를 정형적인 격자형태로 분해하여 본 분석을 위한 기초자료로 활용코자 한다.

본고의 본 분석에서 적용코자 하는 유전적

알고리즘(Genetic Algorithm: 이하 GA라 함)은 이러한 격자형 인구분포 자료(grided population)를 기반으로 제반 도시시설 적정입지 탐색 및 인구추정 모형을 구축하는 데 적용되고 있는 일종의 인구표면모형 중 하나이다.

GA는 Holland(1975)가 애초 진화적 알고리즘(Evolutionary Algorithm) 관련 접근방식 중 하나로 처음 원천 알고리즘을 개발하였고, 이후 Koza(1990)가 개발한 유전적 프로그래밍(Genetic Programming)기법과 함께 다양한 응용분야에서 최적해 탐색기법으로 활용되어 왔으며 앞서 언급한 도시 및 환경관련 분야에서도 많은 실험적 연구가 시도된 바 있다(Muttill and Lee, 2005). 이렇듯 GA기법이 공간탐색 연구에 다양하게 적용된 이유는 GA가 복잡한 비선형 탐색공간(complex non-linear search space)에서 경험적 자료에 기반(inductive data-driven)한 확률적 검색방법(stochastic search method)을 통해 적정해를 자동 탐색하는 알고리즘이기 때문에 별도의 사전 모형을 설정(specification)할 필요가 없다는 실용적 장점이 있기 때문이다(Li and Yeh, 2005).

본 연구는 최근 동탄 제1신도시 준공에 이어 현재 제2신도시가 계획되고 있는 화성시 동부권 일대를 대상지로 하여 읍·면·동의 행정구역 단위로 조사된 관내 인구데이터를 격자형 인구분포자료로 변형한 후 GA 기반 인구추정 모형을 도출해 봄으로써 인구추정에 있어 GA 기법의 유용성을 모의실험해 보고자 하는 데 그 연구 목적이 있다. 특히 기존의 인구격자를 사용한 연구에서는 정밀한 필지 및 건축정보가 부재한 조건 하에 공간가중치(areal weighting)를 적용하거나 크리깅(Kriging) 등의 방식으로 격자를 구축하는 데 비해 본 연구는 한국토지정보체계(KLIS : Korea Land Information System)와 건축행정시스템(eAIS : electronic

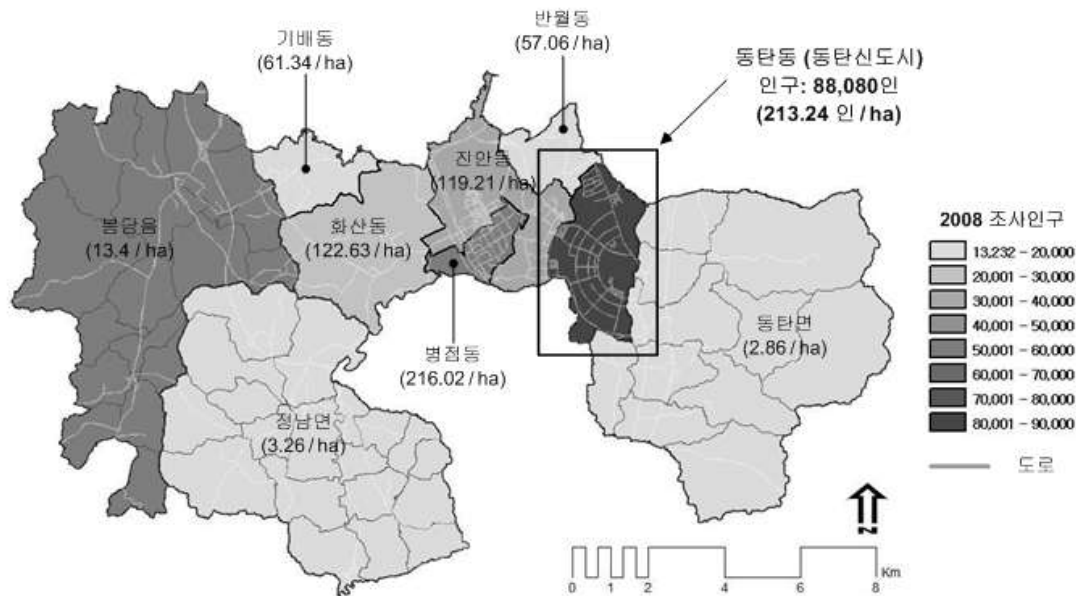


FIGURE 1. 화성시 동부권 Key Map

Architectural Information System) DB을 사용함으로써 실제 필지 및 건축물 정보를 미시적 단위에서 정밀하게 반영하여 구축하였으며, 격자형성 또한 GIS 내재 툴로 생성한 래스터(raster) 격자가 아닌 벡터(vector) 격자를 구축하였다는 특징이 있다.

연구의 공간적 범위 및 원천자료

1. 연구 대상지 개관

화성시는 전체면적이 688.3km²로 제부도를 위시한 서해안 집안지역으로부터 내륙 산악지대에 이르기까지 매우 광범위한 지역에 걸쳐있어 행정편의 상 동부권, 서부권, 남부권 등 크게 세 권역으로 나뉘어 관리되고 있다. 본 연구는 그중 최근 수도권 내 성공적인 2기 신도시로 평가되는 동탄신도시가 조성되면서 그에 따라 인접 택지개발사업지구 조성이 활발히 진행되어 인구증가가 크게 일어나고 있는 화성시 동부권을 사례지로 선택하였다. 위 그림 1은

화성시 동부권의 행정구역 및 인구현황을 보이는 키맵으로 도면의 행정구역별 그레이스케일은 인구규모에 따른 농도를 나타낸다. 그림에서 권역 서측의 봉담읍과 정남면, 그리고 권역 동측의 동탄면 등은 중앙부 6개 동(洞)에 비해 매우 넓은 면적임에도 불구하고 괄호 안 인구밀도(인/ha)에서 보듯이 인구분포가 상대적으로 희박함을 알 수 있다. 이는 봉담읍의 경우 상당부분이 산악지대이며 정남면의 경우 대부분 안양천으로 이어지는 황구지천 주변 평야를 중심으로 전답 위주의 생산녹지로 구성되어 있기 때문이다. 그러나 동탄면(面)의 경우 대부분이 산악지대로 현재 인구밀도는 낮으나 동탄동(洞) 연접부 경부고속도로변 저지대 인근에 동탄 제2신도시가 계획되어 향후에는 큰 인구증가가 예상되는 곳이다. 반면 중앙부 6개 동은 제1신도시가 조성된 동탄동을 비롯하여 최근 다수 택지개발사업들이 밀집한 도시지역으로 2006년도 이전에는 태안읍(浬)이란 일개 행정구역이었으나 제1신도시 조성시기에 맞춰

행정구역 세분화를 통해 현재와 같이 6개 동으로 승격, 세분화된 고밀도 시가화 지역이다.

2. 분석을 위한 원천자료 개관

KLIS는 격자분할에 필요한 필지정보뿐만 아니라 지형정보, 용도지역지구 정보 등 주요 도시계획 공간정보가 통합 구축된 종합 토지정보 체계로서 본 분석에서 사례지에 대한 자연환경, 시설입지, 접근거점 위치 및 용도지역지구 지정현황 등의 설명변수 속성값 구축을 위한 원천자료로도 활용하였다. 또한 eAIS는 기존의 개별 건축물대장에 나타나 있는 건축연면적, 건폐율, 용적률, 층수, 건축물용도 등 공부자료를 전산화하여 구축한 비공간 속성DB이다. 특히 KLIS의 필지 지번과 eAIS의 대지 지번이 동일한 PNU(Parcel Numbering Unit) 코드체계로 작성되어 있어 두 데이터베이스를 PNU 코드를 매개로 결합(join)시킴으로써 KLIS의 토지대장 정보와 eAIS의 건축대장 정보를 개별 필지단위로 통합 연동하여 사용하였다. 또한 사례지자체인 화성시로부터 민원협조를 얻어 2008년도 관내 읍·면·동·리 단위 인구조사자료를 확보하여 사용하였는데 DB는 현재 화성시가 관내 각 읍·면·동에 파견 운영 중인 지역단위 출장소에서 실사를 통해 구축한 각 지역별 인구자료로서 봉담읍, 정남면 및 동탄면의 경우 위 키맵에서 보이는 리(里) 단위까지 구분하여 조사된 매우 정밀한 화성시 자체조사 인구자료이다. 즉, 현재 화성시 인구주택총조사 조사구역(census tract)이 봉담읍, 정남면, 동탄면, 그리고 중앙 6개동 통합구역(이전 태안읍) 등 크게 4개 구역으로 일괄 조사되고 있는 점과 비교할 때 상대적으로 매우 정교한 자체 인구자료임을 짐작할 수 있다.

격자 인구자료 구축

1. KLIS와 eAIS를 활용한 인구자료의 구축

격자인구자료(grid population)를 구축하

기 위해서는 우선 격자의 규모를 설정할 필요가 있다. 격자 크기는 연구목적에 따라 임의로 설정하게 되는데 기존 격자공간 분석사례를 보면 인구추정의 경우 중국의 성(省 : 우리나라의 도 단위에 해당) 단위 인구추정의 경우 1×1km(Yue, 2003)로부터 우리나라 시·군 단위 연구의 경우 200×200m(국토부, 2008a, 2008b), 동(洞) 단위의 경우 50×50m(김형복 등, 2007; 최내영, 2009)까지 적용한 사례가 있고 기타 농경지 및 산림분포 연구를 위한 토지피복(land cover) 분석사례에서는 최소 17×17m(Manson, 2005)를 적용한 경우도 있다. 본 연구에서는 준 도시급 공간분석의 경우로 편의상 200×200m 규모의 격자를 선택하였으며 이 경우 화성시 동부권 전체 격자시스템은 총 4,207개 셀로 분할되게 된다.

설정된 격자 시스템에 대해 각 셀 당 인구를 추계하는 과정은 다음과 같다. 1) 개별 동(洞) 및 리(里) 각각의 1인당 평균 주거점유면적을 계산한다. 이는 eAIS로부터 각 행정구역별 주거건축물 연면적을 집계하여 인구조사자료의 각 해당 행정구역 인구수로 나누어 줌으로써 구할 수 있다. 2) 대상지 전체 KLIS 베이스맵을 200×200m 격자 셀로 분할한 다음 KLIS와 eAIS Join DB로부터 각 셀에 속한 주거건축물 연면적을 합산한다. 이때 필지가 격자선에 의해 분할된 경우는 필지 폴리곤 중심점(centroid)이 속한 셀에 필지를 귀속시켜 계산하였다. 다음 그림 2는 화산동 소재 1개 셀(Feature ID: 2956)을 확대하여 귀속 필지 및 건물이 선택(selection)된 결과를 예시하고 있다. 3) 이렇게 구한 셀별 주거건축물 연면적을 1)에서 구한 1인당 주거면적으로 나누어 해당 셀의 인구를 추계한다. 아래 그림 3은 이와 같은 과정에 의해 화성시 동부권 전체를 200×200m 격자공간으로 분할하여 도출한 격자인구분포현황을 플로팅한 결과이다. 앞의 그림 1에서 읍·면·동 단위로 표시된 행정구역별 인구분포도와 비교할 때 그림 3에서는 공간

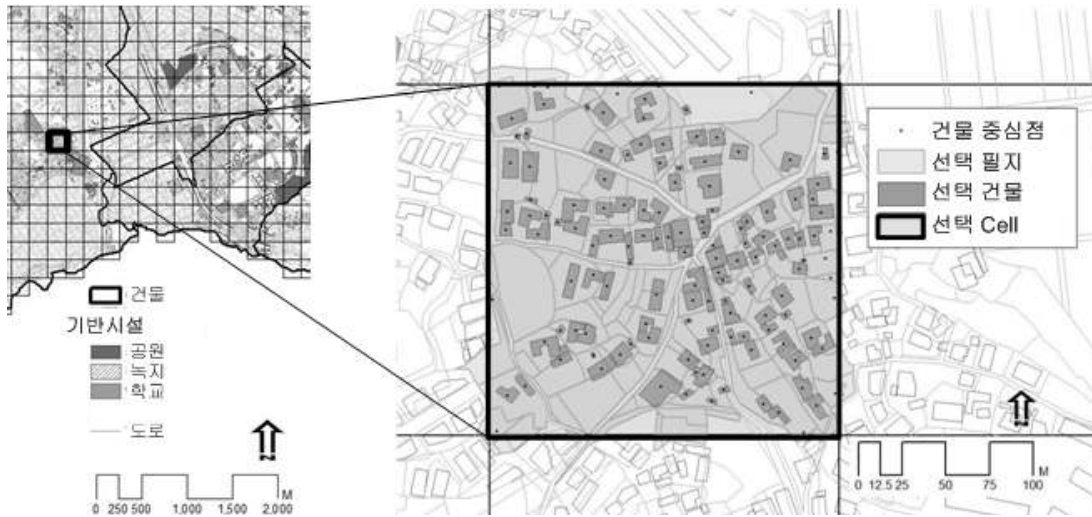


FIGURE 2. 특정 Cell(Feature ID: 2956)의 귀속 필지 선택결과 예시도

적으로 매우 정밀하게 인구밀집 구역의 위치와 인구수가 분해되어 재편되었음을 확연히 알 수 있다.

2. GA모형 추정을 위한 대표 셀의 선정

그림 3의 격자분석 결과는 인구밀도가 낮은

비도시지역에 거주인구가 전무한 셀들이 많을 뿐만 아니라 중앙부 6동의 경우도 신도시와 택지개발사업지구 등에 인구가 고도 밀집한 셀들이 다수 존재하고 있다. GA 인구추정모형의 신뢰도를 높이기 위해서는 이와 같은 돌출값(outliers)들을 제거할 필요가 있으며 본 연구

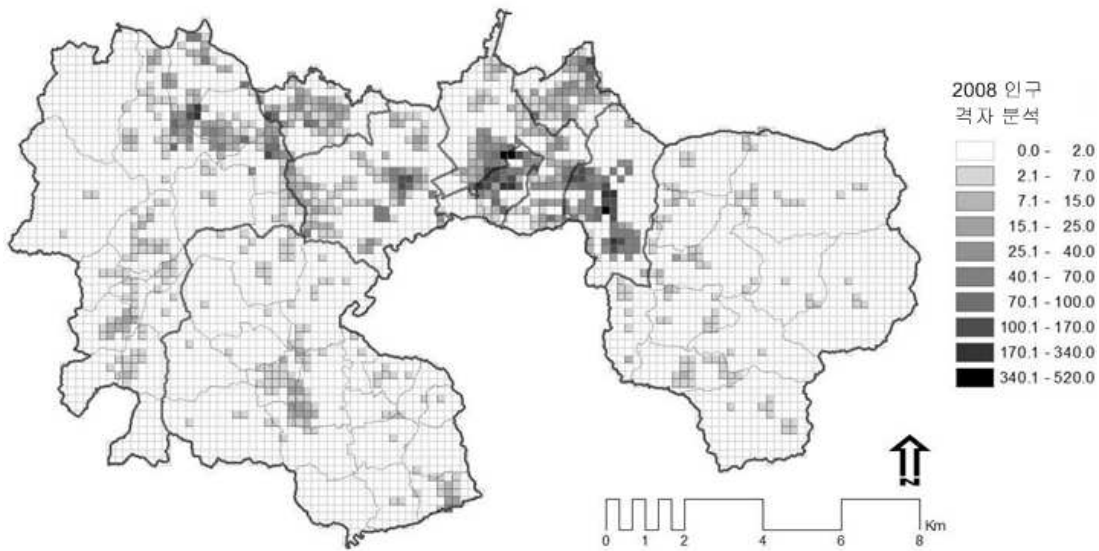


FIGURE 3. 화성시 동부권 200X200m 격자인구분포(grided population) 추정결과

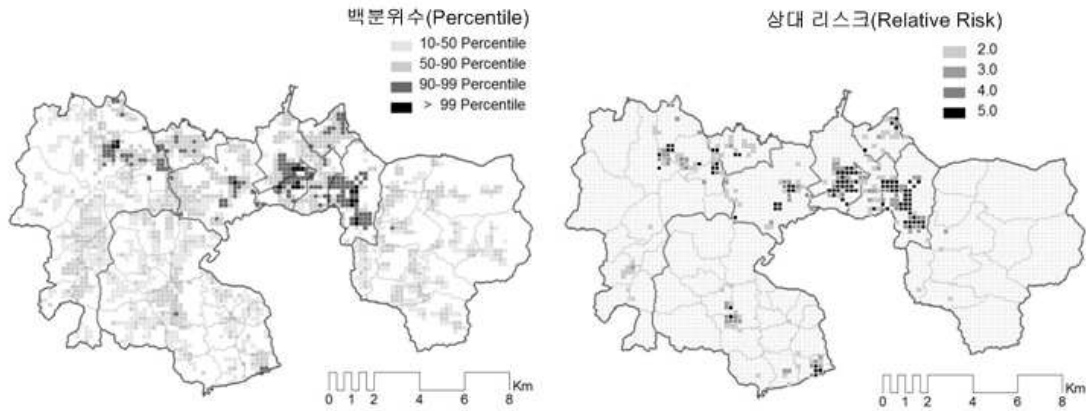


FIGURE 4. 인구격자의 백분위수(Percentile) 및 상대 리스크(Relative Risk) 분석

에서는 그림 4와 같이 GeoDa 0.9.8 프로그램으로 구현한 백분위 맵(Percentile Map)과 상대리스크 맵(Relative Risk Map) 분석을 통해 GA모형 피설명변수값으로 채택할 인구셀들을 선별하였다. 백분위 분석에서 99% 이상의 상위 18개 고밀 셀을 제거하고 산악지역과 농림지역 및 일부 도시지역 저밀 셀들을 90 Percentile까지 제거하여 유효 설명력을 가질 수 있는 셀들을 1차 선별하였다. 여기에 다시 상대리스크 분석을 통해 리스크 5.0 이상의 고리스크 셀들을 제외시켰다. 참고로 상대리스크란 전 구역(셀)에 대한 총 사상(event) 발생 빈도로부터 구한 평균 리스크(average risk)를 기초로 특정 구역(셀)의 사상발생 기대치(expected number of events)를 계산하고 이를 실제 발생 관측치와 비교하여 구한 상대비율(ratio)을 의미한다(Anselin *et al.*, 2006). 리스크 수치가 높을수록 돌출값 내재개연성이 높으며 이는 GeoDa에서 백분위 및 표준편차 choropleth map 분석과 병행하여 빈번히 사용되는 돌출값 판별기준 중 하나이다. 이상의 판별기준에 의하면 70개 셀이 선택되는데, 이는 choropleth map 사전분석과정에서 여타 대안기준을 적용한 다양한 셀 조합 중 가장 우수한 GA 추정결과를 보여 이하 본고에서 최종 분석결과를 정리한 것이다.

GA 인구추정 모형

1. 모형의 설명변수 구축

인구의 입지선택에 영향을 줄 것으로 예상되어 모형에 투입한 주요 설명변수는 다음 표 1에 열거한 바와 같다. 우선 변수특성에 따라 크게 6개 범주로 나누고 각 범주별 세부 요소들을 고려하였는데, 1) 자연환경변수로는 표고와 경사도를 고려하였고 특히 이 두 가지 변수는 표 2에 표시한 바와 같이 구간에 따른 점수(score)를 변수값으로 채택하였다. 수계는 물론 자연환경 변수이기도 하나 상하수도가 정비된 도시환경에서는 수원확보 목적보다는 수변공간의 쾌적성(amenity) 확보의 의미가 공원, 녹지 등과 유사한 것으로 판단하여 2)의 기반시설 범주에 포함시켰다. 2) 대표적인 기반시설로 도로, 학교, 공원, 녹지를 고려하였다. 3) 지역공공시설로 경찰서, 주민센터(동사무소), 우체국, 소방서를 채택하였고, 4) 주요지역거점으로는 서울로의 통근을 위한 집분산점으로서의 고속도로 IC를 고려하였는데, 동탄신도시는 IC를 통해 서울까지 40km 거리를 평균 18분에 도달할 수 있어 서울로의 출퇴근 인구가 많다는 점을 감안한 것이다. 그 외 지역중심 역할을 하는 동탄신도시 중심상업지역도 지역거점으로 고려하였다. 마지막으로 범주 5)와 6)

은 용도지역지구와 관련한 변수들로서 5)번은 개발을 진흥, 촉진하는 용도지구를 6)번은 역으로 개발을 억제하는 규제적 용도지역지구를 고려하였다.

TABLE 1. GA모형의 주요 설명변수

범 주	변 수
1 자연환경	1.1 표고
	1.2 경사
	2.1 수계
2 기반시설	2.2 도로
	2.3 학교
	2.4 공원
	2.5 녹지
3 지역공공 서비스	3.1 경찰서
	3.2 주민센터
	3.3 우체국
	3.4 소방서
4 주요지역거점	4.1 고속도로 IC
	4.2 신도시 중심상업지역
5 개발진흥 용도지역지구	5.1 택지개발예정지구
	5.2 개발진흥지구
	5.3 취락지구
6 개발규제 용도지역지구	6.1 개발제한구역
	6.2 보전녹지
	6.3 계획관리지역

표고와 경사 등 2개 자연환경 변수 외에 나머지 17개 변수 값으로는 모두 대상 요소까지의 직선거리(m)를 적용하였는데 이때 GIS의 NEAR Tool 기능을 이용하여 범주 2)부터 4)까지의 Point Feature 뿐 아니라 5), 6)의 Polygon Feature에 대해서도 각 인구 셀의 중심점으로부터 polygon boundary에 대해 직각으로 교차하는 수선(垂線)을 따라 최단거리를 측정함으로써 변수값을 구축하였다.

2. GA모형 작동기제 및 모형 설정

GA의 작동기제는 자연생태계에서의 종의 진화과정을 설명하는 다윈의 적자생존(survival of the fittest) 원리에 그 기초를 두고 있다 (Balakrishnan, 1996). 추정코자 하는 계수들의 십진법 수치는 전산처리 과정에서 보다 긴 자릿수를 갖는 0과 1의 2진법 숫자(binary string)로 변환되는데 이때 스트링을 생명체의 유전자 염색체(chromosome) 구조와 같이 해석하여 그림 5의 다이어그램과 같이 임의의 값을 갖는 초기 개체군(initial population)들을 생성시킨 후 한 쌍씩의 선대(parents) 개체로부터 2진법 스트링 일부를 1) 교차(Crossover)시켜 후대(offspring) 개체를 생성시킨다. 새로운 후대 개체 값들이 모형의 최대화(Max) 혹은 최소화(Min)를 달성하려는 목적함수(Fitness Function) 값을 향상시킬 경우 보다 적합한(fitter) 개체로 채택하여 다음 세대(generation) 개체군에 진입시키고 그렇지 않을 경우 기각시킨다. 이때 후대보다 여전히 우수한 선대 스트링을 우수개체(Elite)로 간주하여 후대 개체군에 재 진입시키는 방식을 Elitism이라 한다. 이와 함께 2) 변이(Mutation)도 발생시키는데 단순히 염색체 교차만으로는 때로 국지적 극대, 극소값(local maxima or minima)에서 진화를 멈출 수 있기 때문에 특정 염색체 스트링에 작은 돌연변이(mutation)을 발생시킴으로써 전구간 최대, 최소값(global maximum or minimum) 도달 가능성을 높이는 기제로 작용하게 된다(Li and Yeh, 2005).

본 연구에서는 GA 추정인구(P_{GA}) 모형을 다음과 같이 설정하였다.

TABLE 2. 자연환경 변수로서의 표고 및 경사도의 점수체계

기 호	요 소 (단위)	점 수 체 계					
		5	4	3	2	1	0
1.1	표고(m)	$h \leq 30$	$30 < h \leq 60$	$60 < h \leq 100$	$100 < h \leq 150$	$150 < h \leq 200$	$200 < h$
1.2	경사(%)	$d \leq 5$	$5 < d \leq 8$	$8 < d \leq 10$	$10 < d \leq 12$	$12 < d \leq 15$	$15 < d$

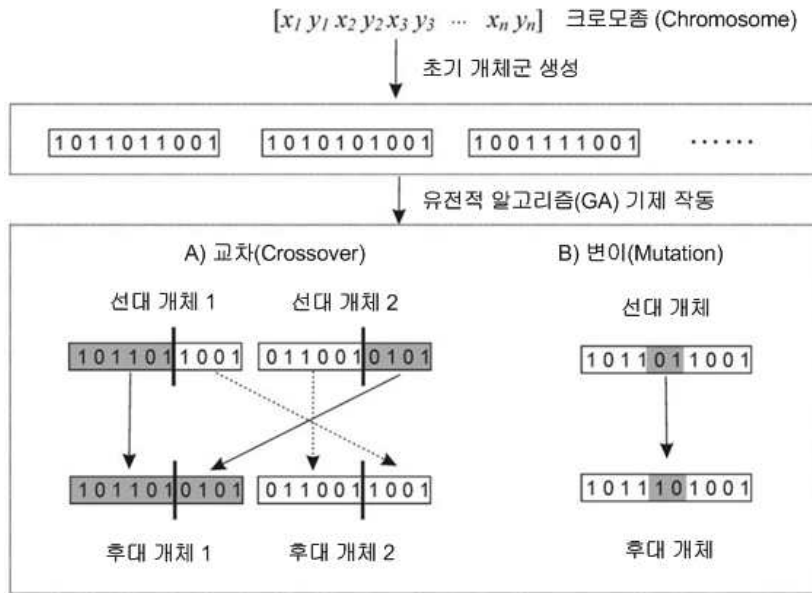


FIGURE 5. GA의 교차(Crossover) 및 변이(Mutation) 작동기제

$$P_{GA} = \alpha \sum_{i=1}^6 \left\{ Coeff_i \sum_{j=1}^{n_j} coeff_j (v_j) \right\} \quad (\text{식 1})$$

식에서 $Coeff_i$ 는 6개 설명변수 범주들의 계수, $coeff_j$ 는 각 범주의 설명변수들 각각의 계수이며 이때 각 범주 또는 계수들의 상대적인 기여도를 비교하기 위해 $Coeff_i$ 와 $coeff_j$ 모두 그 합이 각각 1이 되도록 모형 제약조건을 설정하였는데 따라서 이들은 개념상 일종의 가중치(weight)라고도 할 수 있다. v_j 는 앞 절에서 나열한 설명변수 값으로서 표고와 경사의 구간값(표 2)을 제외하고는 앞서 설명한 바와 같이 모두 대상요소까지의 최단거리를 변수값으로 갖는다. 각 범주는 n_j 개의 설명변수를 갖고 있고 α 는 Summation 수식을 관찰값에 normalizing시키는 임의의 비례규모 상수(random rescaling factor)이다. 따라서 모형의 개체(chromosome)들은 α , $Coeff_i$ 및 $coeff_j$ 의 세 가지가 되는데, GA 알고리즘을 통해 $Coeff_i$, $coeff_j$ 및 α 의 개체 스트링 구

조를 병행(parallel)하게 교차, 변이시키면서 목적함수 최적화를 달성하기 때문에 세 chromosome의 모형성능 향상 기여도는 균등하다고 할 수 있다.

한편 모형 적합성을 판단하는 목적함수로는 관측된(observed) 70개(k) 격자인구 P_{grid_k} 와 위 식 1에 의해 계산된 P_{GA_k} 값 간의 상관계수(Correlation Coefficient)인 C_c 값으로 정의하고 이를 최대화하는 것으로 모형의 성능을 판단하였다. C_c 는 다음 식과 같이 계산된다.

$$C_c = \left\{ \sum_k (P_{grid_k} - \mu_{grid})(P_{GA_k} - \mu_{GA}) \right\} / \left\{ \sum_k (P_{grid_k} - \mu_{grid})^2 \sum_k (P_{GA_k} - \mu_{GA})^2 \right\}^{\frac{1}{2}} \quad (\text{식 2})$$

식에서 μ_{grid} 와 μ_{GA} 는 격자인구와 GA 추정인구 각각의 평균을 나타낸다.

3. GA 패러미터 설정 및 추정결과

위 추정모형 및 목적함수와 함께 본 연구에

서는 모형 패러미터(parameter)를 표 3과 같이 설정하여 GeneHunter 2.1 프로그램으로 모형을 추정하였다. 표에서 보인 바와 같이 초기투입 개체군으로 각 chromosome 당 200 개체를 무작위 생성하였는데 유사 선행연구에서 200 개체를 투입할 때 일반적으로 가장 우수한 결과가 도출되는 것으로 보고된 바 있으며 본 연구의 사전분석(pilot test)에서도 200 개체까지 증가 투입했을 때 가장 우수한 결과에 도달하였고 그 이상을 투입할 경우 계산시간(computing time) 증가에 비해 성능향상이 더 이상 진전되지 않음을 알 수 있었다. 마찬가지로 비트 수의 경우도 현재 PC 운용체제(OS) 하에서는 8, 16, 32비트까지 적용이 가능하나 8비트 이상의 경우 기하급수적인 런타임 증가에 비해 모형성능에서 큰 차이가 나타나지 않아 8비트로 추정하였다. 또한 교차 및 변이확률의 경우도 현재의 0.9와 0.01 값이 런타임과 모형성능향상 면에서 가장 무난한 셋팅으로 판단되었으며, 선대의 우수개체를 후대개체 풀(pool)에 존속시키기 위한 Elitism 조건을 적용하여 진화기체의 효율을 높이고자 하였다. 마지막으로 알고리즘 종료기준은 200 세대(generation)가 진행되도록 더 이상의 성능향상이 발현되지 않을 때 진화과정을 종료하는 것으로 설정하여 최대한 적정값에 도달할 수 있는 충분한 여유를 확보토록 설정하였다.

TABLE 3. GA모형의 패러미터 설정조건

초기개체군 개체수(Population Size)	200
2진법 비트수(Binary Bits)	8-bit
교차 확률(Crossover Prob.)	0.9
변이 확률(Mutation Prob.)	0.01
우수개체 보존(Elitism)	채택
정지기준 세대수(Generation Elaps)	200세대

다수의 추정결과 중 가장 높은 향상성능을 보인 3개 모형에서 추적된 목적함수 향상례적은 그림 6과 같이 나타난다. 초기 개체군의 값은

임의로 설정하였기 때문에 그림에서 보듯 초기 목적함수 값은 매우 낮으나 세대를 거듭하면서 점차 우수 개체가 선택되어 목적함수 값이 빠르게 향상됨을 볼 수 있다. 약 1,000세대에 이르게 되면 최대값에 거의 근접하고 약 1,500 세대를 넘어서면 더 이상의 높은 해가 나타나지 않는 최적해(optimal solution)에 도달함을 알 수 있다. 진화과정 중 일정구간 상승이 멈추는 현상은 일시적인 국지적 극대값에 머무르는 현상으로 이후 일시에 대폭 성능이 도약(leap)하는 지점은 돌연변이(mutation) 기체를 통해 모형 성능이 급격히 향상된 결과로 이해된다.

4. 추정 결과치의 해석

표 4는 위와 같이 관측된 3개 최적 성능 GA 모형에서 도출된 결과치를 요약한 것이다. 우선 표 상단의 목적함수 값을 살펴보면 상관계수 최대값은 0.558을 넘지 않음을 알 수 있다. 최적해가 보다 더 높은 수준으로 향상되지 못하는 점은 앞서 격자인구분포 개관에서도 살펴보았듯이 화성시 동부권이 고도로 도시화된 지역 주변에 저밀인구의 산악지역, 농림지역이 함께 병존하고 있는 입지적 특수성에 기인하는 것으로 여겨진다. 즉 매우 보수적인 기준에서 압축한 70개 모형 추정용 셀 중에서도 인구분포의 변이가 여전히 적지 않음을 암시하는 것으로 판단된다.

표의 하단부는 좌측에 범주 계수 추정치, 우측에 각 설명변수 계수 추정치가 요약되어 있다. 우선 6개 범주 중에서는 1) 자연환경, 5) 개발진흥 용도지구 및 6) 개발규제 용도지구에서 세 모형 모두 일관되게 높은 계수값을 나타내고 있다. 즉, 추정치 그대로를 해석하면 인구밀집요인 중 가장 큰 48-49% 정도가 1) 자연환경 요소, 다음으로 34-44% 정도가 6) 개발규제 용도지역지구 지정현황, 마지막으로 5-15% 정도가 5) 개발진흥 용도지역지구 지정현황의 영향을 받는다고 해석된다. 그 중 먼

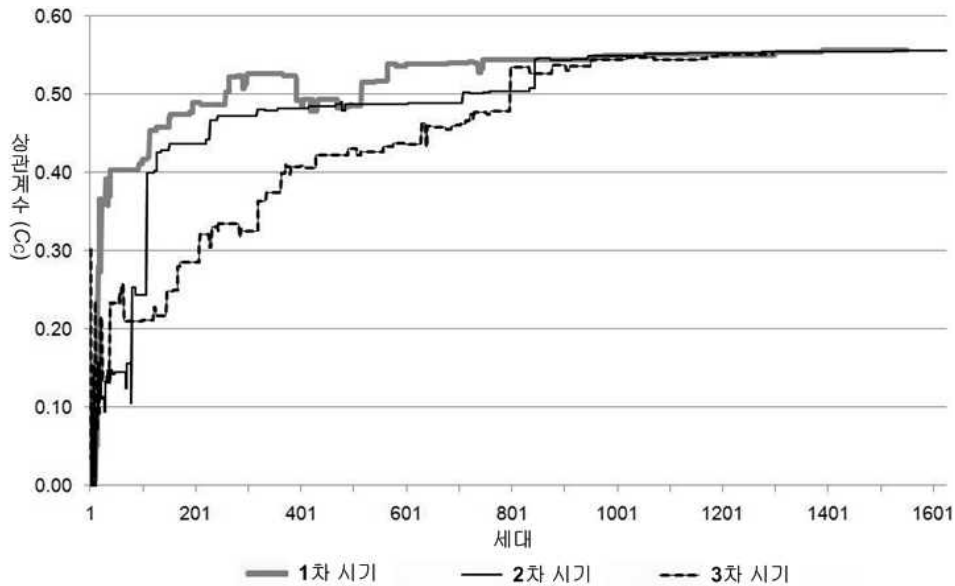


FIGURE 6. GA모형 추정에서 최적해에 도달한 3개 모형의 목적함수 향상계적

저 1) 자연환경 관련 변수에서 98-99%의 절대적인 비율로 경사도에 의해 인구규모가 결정되고 지역의 표고는 영향이 거의 없는 것으로 나타난다. 표고는 주변과 해발고도가 달라도 평탄면을 유지할 경우 토지이용에 지장이 없는 반면 경사도는 어느 입지상황에서나 토지이용에 결정적이라는 점에서 추정치가 의미있는 것으로 판단된다. 5) 개발진흥 용도지역의 경우, 역시 88-98%의 압도적인 부분이 택지개발예정지구와의 근접성이 좌우하는 것으로 나타난 반면 농림지역의 취락지구는 영향력이 전무함을 보이고 있다. 최근의 인구증가는 택지개발사업지구 인근으로 집중되고 있으며 기존의 영농취락지구 정비사업은 인구증가에 전혀 무관하다는 의미로 해석된다. 6) 개발규제 용도지역 중에서도 인구증가 억제요인의 60% 정도가 계획관리지역 근접성, 30-40%가 보전녹지 근접성의 영향을 받으나 그린벨트 지역은 그 영향력이 미미한 것으로 나타나고 있다. 이는 그린벨트의 경우 이미 많은 부분이 해제되었고 현재 남아있는 그린벨트는 개발에 부적합한 산지에 잔존하여 인근 개발잠재력이 희박한데 기

인하는 것으로 보인다.

같은 맥락에서 2), 3), 4) 등 나머지 3개 범주는 0~8% 이하의 미미한 영향력을 가지는 것으로 도출되었는데 일단 범주 계수의 영향력이 이처럼 낮은 경우 해당 범주 관련 설명변수들은 각각의 계수 값이 별도로 측정되었다라든 큰 의미가 없게 된다. 이 같은 해석 기준에서 볼 때 우선 2) 기반시설인 도로, 학교, 공원, 녹지 등의 요소들은 거의 인구집중에 영향을 주지 않는 것으로 도출되었는데 이는 도시화지역을 중심으로 현재 화성시 동부권 기반시설들이 어느 위치에서나 보행접근범위 내에 비교적 골고루 조성되어 있어 입지적 차별성이 없다는 것으로 해석된다. 이는 3) 지역공공서비스와 관련한 경찰서, 주민센터, 우체국, 소방서 등의 경우에도 마찬가지인데 이들의 경우 범주 계수는 0에 가까워 전혀 입주결정에 영향을 주지 못한다고 볼 수 있다. 이처럼 2)와 3)의 결과는 현실적으로 설득력이 있으나 4) 주요지역 거점으로서의 고속도로 IC나 특히 화성 동부권의 생활권 중심지 역할을 할 것으로 기대했던 동탄신도시 중심상업지역 접근성 등의 영향력

TABLE 4. GA 3개 최적모형의 목적함수 및 설명변수 계수 추정결과

목적함수 값		GA 최적해		
		1차	2차	3차
C_c		0.558	0.558	0.557

범 주	GA 최적해			요 소	GA 최적해			
	1차	2차	3차		1차	2차	3차	
1	자연환경	0.482	0.498	0.498	1.1 표고	0.012	0.008	0.008
					1.2 경사	0.988	0.992	0.992
					2.1 수계	0.435	0.435	0.125
2	기반시설	0.067	0.067	0.000	2.2 도로	0.067	0.067	0.208
					2.3 학교	0.000	0.000	0.133
					2.4 공원	0.463	0.467	0.243
					2.5 녹지	0.004	0.035	0.231
					3.1 경찰서	0.478	0.478	0.059
3	지역공공서비스	0.008	0.000	0.000	3.2 주민센터	0.220	0.196	0.059
					3.3 우체국	0.165	0.173	0.686
					3.4 소방서	0.173	0.173	0.220
4	주요지역거점	0.000	0.000	0.082	4.1 고속도로 IC	0.698	0.718	0.745
					4.2 신도시 중심상업지역	0.290	0.294	0.251
5	개발진흥 용도지역지구	0.137	0.149	0.043	5.1 택지개발예정지구	0.890	0.878	0.984
					5.2 개발진흥지구	0.008	0.008	0.141
					5.3 취락지구	0.000	0.000	0.000
6	개발규제 용도지역지구	0.373	0.349	0.443	6.1 개발제한구역	0.016	0.016	0.067
					6.2 보전녹지	0.416	0.416	0.341
					6.3 계획관리지역	0.569	0.569	0.592

이 매우 미미하다는 사실은 다소 의외의 결과로 받아들여진다. 다만 IC의 경우는 일단 차량을 이용한 접근거리이기 때문에 동부권 내에서는 IC까지의 주행 접근시간에 큰 차이가 없기 때문일 수도 있고 동탄신도시 상업지역의 경우도 현재 입주가 완료되지 않은 단계이어서 영향력이 형성되기엔 아직 완성도가 낮은 시점이기 때문일 가능성을 배제할 수 없다고 본다.

결 론

본 연구에서 격자인구기반 GA모형을 위해 채택, 실험한 설명변수들의 분석결과는 비교적 일관되고 설득력 있는 추정치가 도출되었다고 판단된다. 그러나 본 사례대상지는 도시와 산지, 농지가 혼재하며 급속한 개발단계에서 인

구동태와 도시현황이 아직 안정되지 못한 입지 특성상의 제약이 있어 GA모형의 목적함수 적정해는 다소 기대보다 높지 못하였다는 한계가 있었다. 이와 관련하여 향후 후속연구에서는 GA모형의 프로그램 코딩과정에서 본 연구에서 설정한 상관계수 외에 보다 다양한 다중 목적함수(multiple fitness function)를 채택함으로써 다면적 평가를 추가 모색해 볼 필요가 있다고 여겨진다. 또한 본 연구에서는 GA모형의 염색체 교차(crossover)와 변이(mutation) 패러미터 설정의 조합을 좀 더 다양화하여 미세한 변화를 보다 면밀하게 모의실험함으로써 모형성능향상의 정도를 상세하게 추가 비교해 볼 필요도 있을 것이다. 더욱 큰 관점에서는 본 연구에서 시험적으로 활용한 GA기법 외에 이와 상응하는 대안적 인구모형 추정방법들로서

Simulated Annealing, Area-to-point Kriging, OLS 및 Stepwise Regression, 기타 Gravity Model 등 여타 다양한 기법들과의 성능비교연구 등도 매우 의미 있는 확장연구 과제가 될 수 있다고 여겨진다.

결론적으로 본 연구를 통해 GA모형 운용을 위해서는 공간변수 구축과정에서 지리정보체계의 처리가 불가분하다는 점에서 GA기법과 GIS의 연계활용이 필수적이라는 점과 함께 GA모형이 인구추정이라든가 도시현상 탐색 등의 문제(problem)들에 있어 적정해(optimal solution)를 찾을 수 있는 매우 강력하고 학제적인 분석방법 중 하나라는 점을 확인할 수 있었다. **KAGIS**

참고 문헌

- 국토해양부. 2008a. 기반시설부담구역제도 시행 및 운영활성화를 위한 연구.
- 국토해양부. 2008b. 기반시설부담구역제도 운영 편람.
- 김형복, 전병은, 최내영. 2007. 기반시설부담금에 관한 법률 및 그 운영에 대한 특강자료. 대한국토·도시계획학회 국토도시아카데미.
- 최내영. 2009. 인구증가 분석격자의 공간정보를 이용한 기반시설 부담구역 설정방안. 한국지리정보학회지 12(4): 74-83.
- 화성시. 2004. 화성시 기반시설부담구역 지정 및 부담계획 기준수립 연구.
- Anselin, L., N. Lozano and J. Koschinsky. 2006. Rate transformations and smoothing. A discussion paper of the Spatial Analysis Laboratory, University of Illinois at Urbana-Champaign.
- Balakrishnan, P.V. and V.S. Jacob. 1996. Genetic algorithms for product design, Management Science 42(8):1105-1117.
- Holland, H.H. 1975. Adaptation in Natural and Artificial Systems. The University of Michigan Press: Ann Arbor.
- Koza, J.R. 1990. Genetic Programming: A Paradigm for Genetically Breeding Populations of Computer Programs to Solve Problems, Stanford University Report, Report No. STAN-CS-90-1394.
- Li, X. and A.G. Yeh. 2005. Integration of genetic algorithms and GIS for optimal location search. International Journal of Geographical Information Science 19(5):581-601.
- Manson, S.M. 2005. Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. Agriculture, Ecosystems and Environment 111:47-62.
- Muttill, N and J. Lee. 2005. Genetic programming for analysis and real-time prediction of coastal algal blooms. Ecological Modelling 189:363-376.
- Yue, T.X., Y.A. Wang, S.P. Chen, J.Y. Liu, D.S. Qiu, X.Z. Deng, M.L. Liu and Y.Z. Tian. 2003. Numerical simulation of population distribution in China. Population and Environment 25(2):141-163.
- Yue, T.X., Z.P. Du and Y.J. Song. 2008. Spatial models and geographic information systems. Encyclopedia of Ecology, 3315-3325. **KAGIS**