

유전알고리즘을 이용한 동태적 공간모델의 보정

Calibration of Dynamic Spatial Model Using Genetic Algorithms

김복환 국토해양부 도시재생과 서기관(제1저자)
Kim Bokhwan Senior Deputy Director, Ministry of Land, Transport,
Maritime Affairs, Urban Regeneration Division.
(Primary Author)

양광식 순천향대학교 조교수
Yang Kwangsik Assistant Professor, SoonChunHyang Univ.

목 차

- I. 서론
- II. 모델보정과 유전알고리즘
 - 1. 모델보정
 - 2. 유전알고리즘
 - 3. CA모델의 설정과 GA를 이용한 모델보정
- III. GA를 이용한 모델보정의 결과분석
 - 1. 민감도분석(정태적 분석)
 - 2. 가상데이터와 GA를 이용한 모델보정의 정확성 분석
 - 3. 실제데이터와 GA를 이용한 모델보정의 일관성 분석
 - 4. GA를 이용한 모델보정결과의 해석
- IV. 결론

I. 서론

모델개발은 문제의 정의, 모델설계, 모델보정, 모델검증, 예측 등의 절차를 거치는 일련의 반복적인 과정이다(Waddell, 2003). 이중 모델보정(Model Calibration)은 모델의 수치값에 대한 조정을 통해 모델의 산출물과 실제데이터를 일치시키는 최적의 수치를 찾아내려는 노력이고 이를 통해 결국 모델에 미래예측능력을 부여하는 단계로서 실증적인 모델의 개발을 위해서는 필수적인 과정이다. 현실에서의 모델보정은 모델의 산출물과 실제데이터를 비교한 후, 일반적으로 두 비교대상의 차이로 표현되는 목적함수를 최적화하는 반복적인 과정으로 진행된다. 따라서 모델보정은 최적화하려는 목적함수의 형태에 많은 영향을 받는데, 목적함수가 비선형이고 수학적으로 잘 정의되지 않거나, 다수의 해가 존재하는 경우, 미분 불가능한 경우 등은 모델보정에 어려움이 따른다.

또한, 도시가 공간에서의 복잡한 동태적 시스템이라는 인식이 확산되면서(Batty, 1996; Wilson, 2000), 1980년대 후반부터 일련의 학자들은 시·공간에서 동태적 모델링이 가능한 셀룰라오토마타(CA : Cellular Automata) 등 복잡성이론을 이용하여 도시성장을 모델링하기 시작하였다(Batty et al. 1989; White and Engelen, 1993; Batty and Xie, 1994; Clarke et al. 1997; Wu and Webster, 1998; Benenson et al. 2002). 그러나 도시모델링의 도구가 상향적·동태적 접근방법으로 더욱 복잡해질수록 모델보정과정도 이에 비례하여 더욱 난해해졌다. 즉, CA 등 복잡성이론을 이용한 도시성장모델은 동태적 도시성장의 모델링에 적합한 유연성, 투명성, 단순성 등 여러 장점을 지녔음에도 불구하고 ① CA가 많은 모수를 포함하고 있고 이들 모수 간 상호작용이 복잡하며, ② 이러한 모

델내부의 불확실성으로 이들 모수의 수치값들이 연구자에 따라 자의적으로 결정되었고, ③ 특히, 실증적·동태적 CA도시성장모델은 수많은 반복적 비교과정과 확률적인 요소들을 포함하고 있는 등의 이유로 이러한 모델의 보정은 더욱 어려워졌다.

결국 CA 등 복잡성이론을 이용한 모델의 보정은 모델의 성패를 좌우하는 중요한 요소로서(Torrens, 2000; Torrens and O' Sullivan, 2001; Li and Yeh, 2001, 2002; Waddell, 2003), 이러한 모델이 동태적 도시성장을 성공적으로 모델링할 수 있는 도구로 자리매김 되기 위해서는 보다 더 견고한 모델보정방법의 개발과 아울러 모델보정결과의 정확성과 일관성 검증 등에 대한 체계적인 연구가 절실히 필요하다고 할 수 있다. 그럼에도 불구하고 아직까지도 이에 대한 연구는 체계적으로 진행되지 않았다(White and Engelen, 1993; Li and Yeh, 2001; Wu, 2002).

이러한 상황에서 인공지능과 휴리스틱을 이용한 모델보정기법은 하나의 대안이 될 수 있다. 이 기법은 목적함수가 복잡하고, 다수의 해가 존재하거나, 모수들 간에 복잡한 상호관계가 존재하는 상황에서도 적용이 가능하여 이미 몇몇의 연구에서 복잡성이론을 이용한 동태적 모델의 보정에 적용된 사례가 있다(Li and Yeh, 2001, 2002; Yeh and Li, 2003; Straatman et al. 2004; Heppenstall, 2005). 다만, 인공지능과 휴리스틱을 이용한 모델보정은 검은상자(Black Box) 기법이고 모델보정결과의 일관성과 정확성 등에 대한 신뢰성문제 등이 비판받고 있다.

본 연구는 이러한 연구경향 및 비판에 대한 논의를 배경으로 한다. 본 연구는 CA를 이용한 동태적 도시성장모델을 인공지능의 일종인 유전알고리즘(Genetic Algorithms : GA)을 이용하여 보정하고 그 보정결과의 정확성과 일관성에 대한 통계적 검증

을 시도하였다. 또한, 이러한 연구목적을 달성하기 위하여 다음과 같은 4가지의 세부적인 방법론을 제시하였다. 첫째, 다른 모수값들이 불변인 상황에서 목적함수를 최적화시키는 개개의 모수값을 찾아내는 정태적분석(민감도분석)을 시도하였다. 둘째, 모든 모수값이 변화하는 상황에서 GA와 가상데이터를 이용하여 최적모수값을 찾는 동태적 분석을 하였고 모델보정결과의 정확성(Accuracy)을 검증하였다. 또한, 앞의 정태적 분석과 동태적 분석의 결과를 비교하였다. 셋째, 실제데이터를 이용하여 GA를 이용한 모델보정결과의 일관성(Consistency)을 검증하였다. 넷째, GA를 이용하여 찾아진 보정값들의 실질적 의미에 대한 해석을 시도하였다.

이러한 방법론에 근거하여 II장에서는 우선 모델보정 및 GA에 대한 이론적 고찰과 CA모델의 설정 및 GA를 이용한 모델의 보정과정 등 실증적인 연구환경에 대해 살펴보고, III장에서 위에서 제시한 4가지의 연구결과를 논의한 후, IV장에서 결론을 도출하기로 한다.

II. 모델보정과 유전알고리즘

1. 모델보정

Rykiel(1996)은 모델보정을 “모델의 산출물과 실제데이터의 일치도를 향상시키기 위한 모델의 모수 및 상수값들에 대한 예측과 조정”이라고 정의하였고, Birkin et al.(1996)은 모델의 보정을 “수치값(Numerical Values)들이 미지의 모델의 모수에 할당되면서 모델이 정확하게 실제의 패턴을 재생산하는 과정”이라고 정의하였다. 또한, Clarke et al.(1997)은 모델보정을 모델의 산출물이 “과거자료와 일치하는 결과를 산출하고, 모델에 같은 규칙을 반복 적용하여 미래를 예측하기 위한 것”이라고

주장한다. 이러한 의미에서 볼 때, 모델보정은 모델을 이용한 미래예측의 전제가 되는 과정으로 모델의 산출물이 실제데이터와 일치하는 모델의 모수값 등을 찾아가는 과정이라고 정의할 수 있다.

사실, 한두 개의 모수를 가진 모델이거나 모델보정의 목적함수가 수학적으로 잘 정의된 경우라면 엑셀이나 최적화 프로그래밍 또는 라그랑지 함수 등을 이용하여 모델을 쉽게 보정할 수 있다. 그러나 ① 모델이 다수의 모수를 포함하고 이러한 모수들 간에 복잡한 상호관련이 있으며, ② 외관원의 문제나 체스게임과 같이 목적함수의 해결공간이 방대하거나, ③ 목적함수의 형태가 비선형 또는 수학적으로 잘 정의되지 않는 복잡한 형태라면 모델보정은 보다 심각한 문제에 직면한다(Openshaw and Openshaw, 1997).

더욱이, 최근 셀룰라오토마타 등 복잡성이론을 이용한 ‘실증적’·‘동태적’ 도시성장모델의 보정은 더욱 어려움이 가중된다. 즉, 가상의 도시성장을 대상으로 했던 모델의 보정은 찾아진 모수값 등에 대한 통계적 유의성이나 일관성 등을 검증하는 이론적·학술적 의미에 치중했다고 할 수 있다. 반면에, ‘실증적’ 도시성장모델의 보정은 실제의 도시성장패턴을 반영하는 모수값 등을 찾아내는 과정이어서 모델이 지속적으로 도시영상을 시뮬레이션하고 이를 비교시점의 실제 도시영상과 비교하면서 목적함수를 최적화하는 반복적이고 지루한 과정을 거쳐야 하므로 방대한 컴퓨팅 비용과 시간을 필요로 한다.

또한, 복잡성이론(특히 CA)을 이용한 ‘동태적’ 도시성장모델은 이산적 시간의 흐름에 따라 점진적으로 성장하는 도시영상이 다시 모델내부에 반복적으로 재투입되면서 진화하고, 이 과정에서 확률적인 요인이 결합되므로 종국적인 모델의 산출물은 초기값 및 확률적인 과정에 민감하게 반응하

게 된다. 따라서 초기조건의 미세한 차이는 최종산출물에 커다란 변화를 야기할 수 있고, 모델내부의 확률적 요인에 의한 불확실성은 모든 조건이 동일한 상태에서 동일한 모수값을 적용한 시뮬레이션이라도 최종산출물을 달라지게 할 수 있다. 결국 모델의 최종산출물이 초기값이나 확률적인 불확실성에 민감하다는 것은 일반적으로 모델의 산출물과 실제 현상과의 비교를 통해 얻어지는 목적함수의 최적화를 어렵게 한다.

실증적인 도시성장모델을 구현하려는 목적이 실제의 도시성장패턴을 반영하는 모수값 등을 찾아내고, 이로부터 과거 도시성장패턴의 문제점과 동인을 이해하며, 다양한 What-If형태의 시뮬레이션을 통해 미래의 지속가능한 도시성장전략 및 정책수단을 제시하는 것이라면 이러한 모델에서의 모델보정은 비록 어렵더라도, 반드시 거쳐야 하는 필수적인 과정일 것이다. 그럼에도 불구하고, CA 등 복잡성이론을 이용한 실증적·동태적 도시성장모델의 보정에 대한 학계의 연구는 부족한 것이 현실이다(White and Engelen, 1993; Li and Yeh, 2001; Wu, 2002).

이러한 상황에서 인공지능기법을 이용한 휴리스틱(heuristic) 탐색방법을 이용한 시·공간 동태 모델의 보정은 하나의 대안으로 제시되고 있다. 휴리스틱 탐색은 전체 문제해결공간에서 최적의 해법이 아닌 대안들을 제거해 나가면서 문제의 최적대안을 찾아가는 방법이다. 비록 휴리스틱 탐색이 알고리즘(Algorithms)과 달리 지침적 특성 때문에 항상 최고의 대안을 찾아내지 못한다 하더라도 방대한 탐색공간에서 알고리즘을 이용한 순차적 최적해의 탐색이 시간과 비용적인 측면에서 곤란할 때, 효과적으로 적용될 수 있는 기법이다(오세진 외 14인, 2005). 또한, 최근 휴리스틱 탐색방법은 인공지능(Artificial Intelligence: AI)기법들과 접목되어, AI가 동태적 공간모델에 지능을 제공함에 따라 단

지 공간모델만 사용한 경우와 비교할 때, 보다 현실적인 시뮬레이션 결과를 제공하게 된다(Openshaw and Openshaw, 1997). 이러한 장점으로 동태적 공간모델과 AI를 접목하려는 시도를 최근 몇몇 연구에서 찾아볼 수 있다.

우선, Richards et al.(1990)은 액체상태의 수정이 나무와 같은 패턴을 보이면서 응고되어 가는 과정을 CA를 이용하여 모델링하고 GA를 이용하여 이 모델을 보정하였다. 즉, 액체 또는 고체(0 또는 1)인 중심셀의 상태변화가 28개의 주변이웃에 의하여 결정되는 과정을 시뮬레이션 하였는데, 그들은 결국 모든 가능한 이웃의 상태(228개)에서 실제의 응고과정을 설명할 수 있는 최적의 이웃의 상태를 GA를 이용하여 탐색하였다. Clarke and Gaydos(1998)는 '자기수정적(Self-Modified)'인 과정을 통해 CA도시성장모델을 보정하였다. 그들의 모델에서 5개의 모수값은 시뮬레이션된 결과치가 실제치를 과도 또는 과소하게 예측하는 경우, 모델내부에서 자체적/점증적으로 수정되는 과정을 거쳐 보정되었다. Li and Yeh(2001, 2002)와 Yeh and Li(2003)는 확률적 요소가 가미된 CA도시성장모형을 신경망(ANN: Artificial Neural Networks)을 이용하여 보정하였다. 그들은 몇 개의 같은 지점에서 도시성장요소들의 샘플을 채취하여, 신경망을 이용하여 이들의 가중치(Weights)를 구하였고, 구해진 가중치를 이용하여 도시성장을 시뮬레이션 하였다. Straatman et al.(2004)은 White and Engelen(1993)의 CA 도시성장모델을 '깊이와 너비 탐색방법(Length and Width Searching Methods)'을 사용하여 보정하였다. Heppenstall(2005)은 주유소들이 이익극대화 목표하에 경쟁하는 과정을 Multi-Agents 모델로 설명하고, 모수값들을 진화프로그래밍(Evolutionary Programming)을 이용하여 보정하였다.

2. 유전알고리즘

유전알고리즘(Genetic Algorithms : GA)은 준확률적인 탐색알고리즘으로 “부모유전자의 상속과 다윈의 생존경쟁이라는 자연적인 현상을 모델링하려는 시도다”(Michalewicz. 1994). GA는 거칠고 변화하는 환경에서 적자생존의 원칙하에 살아남은 최적의 부모세대의 능력이 교차와 변이를 통해 다음세대에 전해짐으로써 적응과 진화를 통해 살아남기 위한 생명체들의 능력을 반영하려고 노력한다.

GA는 일반적으로 난수를 이용하여 초기 해집단을 작성하면서 시작되는데, 개개의 초기 해집단은 염색체라 불리며 이는 주어진 문제의 하나의 잠정 해결책으로 간주된다. 또한, 하나의 염색체는 몇 개의 유전자로 구성되어 있는데 각각의 유전자는 그 염색체의 하나 또는 그 이상의 특징을 표현한다(Michalewicz. 1994). 초기 해집단의 염색체는 평가(Evaluation), 재생산(Reproduction), 그리고 유전자 조작과정인 교차(Cross-Over)와 변이(Mutation)과정을 거쳐 새로운 자식단계의 해집단으로 진화한다. 이러한 진화는 반복적인 과정을 거쳐 주어진 최적화문제에 더욱 바람직한 자식들을 생산하면서 보다 적합한 문제해결방안을 제시한다.

GA는 우선 모수 간의 관계가 복잡하고 비선형이며 거대한 탐색공간에서도 적용될 수 있는 장점이 있다(Ghanea-Hercocck. 2003). GA는 전통적인 경사도 탐색기법 등과 달리 초기점의 영향을 덜 받으며, 탐색공간이 다차원이거나, 목적함수의 미분이 불가능한 경우에도 유용하게 적용할 수 있다(Wang. 1997; Wong et al. 2001).

또한, GA는 CA를 이용한 동태적 모델을 보정하는데 적합한 기법이라고 할 수 있다. 종전의 CA를 이용한 동태적 도시성장모델을 보정하는 데 사용되었던 다기준평가방법(Wu and Webster. 1998)

이나, 계층적분석방법, Logistic Regression(Wu. 1998, 2002) 등의 방법은 정태적인 기법으로 다양한 입력변수들 간에 다중공선성이 존재하고, 변수들의 공간적 형태를 결정하는 모수들이 한꺼번에 변화하는 복잡하고 동태적인 환경에서는 적용하기 어려운 기법들이다(Li and Yeh. 2001, 2002). GA는 이러한 여건에서도 모수의 변화와 이와 연계된 변수들의 공간적 분포형태의 변화를 동시에 보정하는 것이 가능하여 동태적모델을 보정하는 데 적합한 기법이라고 할 수 있다.

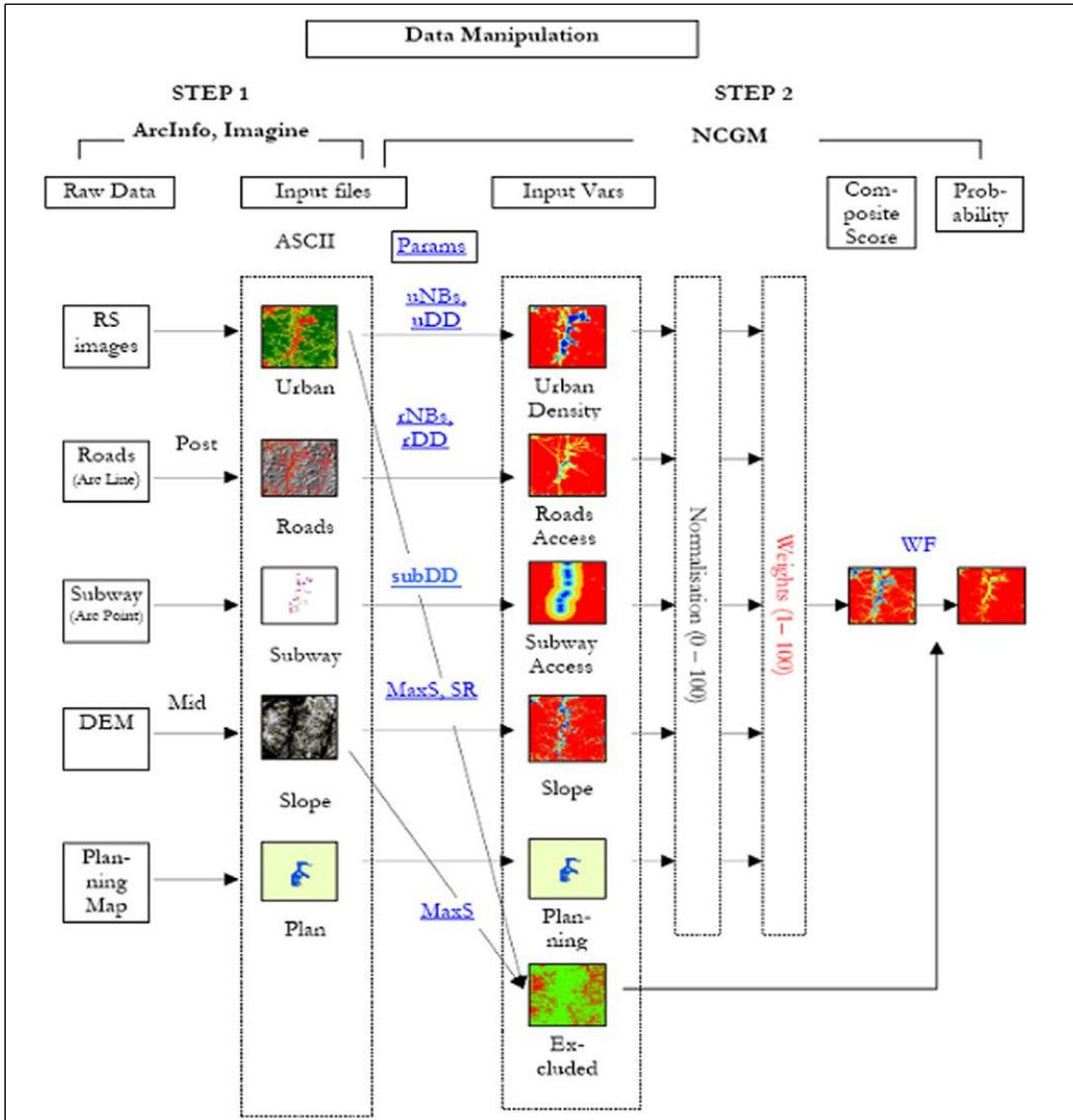
GA는 교차와 변이과정을 거침으로써 종종 지역적 최소점(Local Minima)에 고착되는 현상으로부터 벗어날 수 있다. 이는 단순히 GA의 교차와 유사한 방법만을 이용하여 수백 시간의 모델보정과정이 요구되었던 Clarke and Gaydos(1998)의 모델보정보다 효율적인 방법이라고 할 수 있다. 또한 Li and Yeh(2001, 2002)는 신경망을 이용하여 그들 모델의 입력변수에 대한 가중치를 보정하는 과정에서 입력변수들의 샘플자료를 이용하였으나, GA는 입력변수들의 전체자료를 이용하여 적정모수와 가중치를 동시에 찾으므로 보다 신뢰성 있는 결과를 얻을 수 있다.

3. CA모델의 설정 및 GA를 이용한 모델보정

본 연구에서 GA는 CA를 이용한 신도시성장모델(NCGM : New City Growth Model)¹⁾을 보정하는데 적용되었다. NCGM은 Wu and Webster(1998)의 초기모델과 후속모델(Li and Yeh. 2000, 2001; Ward et al. 2000; Wu and Martin. 2002)을 기반으로 하여 신도시성장을 모델링할 수 있도록 수정된 모델이다.

<그림 1>은 NCGM의 구동과정과 모델에 포함된 가중치와 모수를 보여준다. 우선 5개의 입력과일

그림 1_NCGM의 구동과정과 가중치 및 모수



주: 입력변수 (6개): 도시밀도(Urban Density), 도로접근성(Roads Accessibility), 지하철접근성(Subway Accessibility), 경사도 (Slope), 계획지역(Planning Areas), 제외지역(Excluded Areas)

Weights (5개): 도시밀도, 도로접근성, 지하철접근성, 경사도, 계획지역

Parameters (8개)

uNBs, rNBs: 도시밀도변수와 도로접근성변수의 이웃(Neighbourhoods)의 범위를 결정

uDD, rDD, subDD: 도시밀도, 도로접근성, 지하철접근성변수의 거리조락모수(Distance Decay)

MaxS (Maximum Slope): 허용최대경사도(최대경사도 이상의 경사에서는 도시성장불가)

SR (Slope Resistance): 경사저항모수

WF(Weight Factor): 가중요소모수

- 1) NCGM의 주요한 특징은 첫째, 입력변수에 계획변수와 용도지역·지구 등 외생적인 정책수단을 포함하여 계획적인 도시성장을 시뮬레이션할 수 있고, 정책변화에 따른 What-If 형태의 도시성장 시뮬레이션이 가능하다. 둘째, 신도시성장의 시기별 시뮬레이션이 가능하다. 즉, '신도시계획' 변수를 추가하여 신도시 건설 중의 도시성장을 모델링하고, 이 변수를 제거하여 신도시건설 이전 또는 이후의 도시성장을 모델링할 수 있다. 또한, 도로 등의 인프라건설의 예측을 추가하여 도시성장을 시뮬레이션 하는 것도 가능하다. 셋째, 분산모수를 가중요소(Weight Factor)모수로 수정하여 보다 일반적인 인간의 행동원리를 표현하고자 하였다(김복환·양광식, 2007).

(Input File)은 5개의 원시데이터(Raw Data)에서 추출되며, 모델에 투입되어 6개의 입력변수(Input Variables)로 변환된다. 이 과정에서 입력변수들의 공간적 분포형태를 특징짓는 7개의 모수와, 전환적합도(Suitability of Conversion 또는 Composite Score)를 개발확률(Development Probability)로 변환하는 과정에서 가중요소모수가 적용되어 총 8개의 모수가 모델내부에서 작용한다. 또한 제외지역(Excluded Areas)을 제외한 5개의 변수에는 가중치가 적용되어 각 입력변수들의 중요도를 결정한다.

NCGM의 입력변수와 각 입력변수와 연관된 모수들은 신도시 계획 전/계획 중/계획 후로 구분되는 성장시기와 교통인프라시설의 예측유무에 따라 달라진다(<표 1> 참조). 따라서 시기별/예측유무별 입력변수와 모수의 변화에 따라 이 둘의 조합이 가질 수 있는 전체범위도 달라진다. 교통시설의 건설을 예측한 상황에서 신도시계획 중의 시기가 가장 많은 조합의 범위를 가지는데(5×10^{20}) 이는 가중치와 모수들의 결합으로 생성가능한 문제해결공간이 방대함을 의미하며, 결국 NCGM의 보정 시 최적화해야 하는 다차원 목적함수의 비선형의 탐

색공간이 방대함을 의미한다.

GA를 이용한 NCGM의 보정과정을 살펴보면, 우선 난수를 발생시켜 50개의 염색체를 가지는 임의의 초기해집단을 설정한다. 하나의 염색체는 시기별/예측의 유무별로 가중치와 모수의 개수에 따라 10~13개의 유전자로 구성되어 이는 문제해결공간에서 하나의 대안으로 인식될 수 있다. 초기해집단의 각 염색체는 평가과정을 거쳐 모델의 산출물과 실제 도시영상 간의 비교를 통해 평균제곱오차값(Mean Square Error : MSE)이 계산되고 각 염색체에 꼬리표처럼 부착된다. GA는 이러한 꼬리표를 인식하여 재생산-교차와 변이과정-평가의 순환과정을 50번 반복하여 MSE를 가장 최소화하는 염색체를 찾아낸다.

III. GA를 이용한 모델보정의 결과분석

1. 민감도분석 (정태적 분석)

민감도분석은 종속변수와 설명변수 간의 관계의 강도가 모호한 경우 설명변수의 변화에 따른 종속변수의 변화값을 측정함으로써 각기 다른 설명변수들이

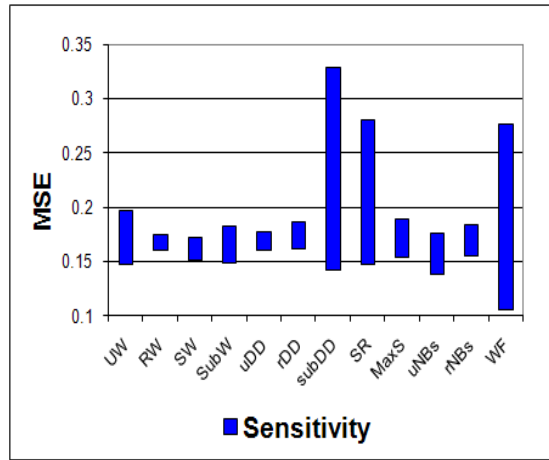
표 1_NCGM의 시기별/예측유무별 가중치와 모수의 개수 및 이들의 조합의 수

| 구분 | 가중치 [개수] | 모수 [개수] | 개수 | 조합의 수 |
|-------|---|---|---------|--|
| 계획 전 | 도시밀도, 도로접근성, 경사도 [3] | uDD, rDD, SR, MaxS, uNBs, rNBs, WF [7] | 10 | 5×10^{15} |
| 계획 중 | 도시밀도, 도로접근성, 경사도, 계획지역, (지하철접근도) [4(5)] | uDD, rDD, SR, MaxS, uNBs, rNBs, WF, (sDD) [7(8)] | 11 (13) | 5×10^{17} (5×10^{20}) |
| 계획 후 | 도시밀도, 도로접근성, 경사도, 지하철 접근도 [4] | uDD, rDD, SR, MaxS, uNBs, rNBs, WF, sDD [8] | 12 | 5×10^{18} |
| 값의 범위 | 1~100 | 1~10 (uDD, rDD, sDD, SR) 10~30 (MaxS), 1~100 (WF) 1~50 (uNBs, rNBs) | | |

주: ()는 교통인프라시설의 건설을 예측한 경우

중속변수에 미치는 영향을 파악하는 데 유용한 분석 방법이다. 본 연구에서는 NCGM의 가중치와 모수 값을 한 단위씩 증가시키면서 시뮬레이션된 도시영상을 얻고 이를 실제의 도시영상과 비교하여 평균제곱오차를 계산하였고 이러한 MSE의 변화를 측정하여 각 가중치와 모수값의 영향력을 측정하였다. 하나의 가중치 또는 모수값의 민감도를 분석할 때 나머지 가중치와 모수값들은 각각의 중간값에 불변인 상태로 고정되었다. NCGM의 민감도분석은 분당신도시의 신도시 건설 후(1994~2000)와 건설 중(1990~1994)의 시기에 적용되었다.

그림 2_가중치와 모수의 민감도크기



1) 신도시 건설 후 시기(1994~2000)

<그림 2>는 분당신도시 건설 후 시기의 모든 가중치와 모수에 대한 최대·최소 MSE값과 두 값의 차이인 민감도의 크기(막대그래프의 길이)를 보여준다. 우선 민감도의 크기는 subDD(지하철접근성의 거리저항모수), WF(가중요소), SR(경사저항모수)의 순으로 크다. 또한, 이 3가지 모수는 다른 변수들과 비교하여 최대의 MSE값(막대그래프의 상단의 값)을 생성하며, 그 값은 대략 subDD(0.330), WF(0.281), SR(0.277)이다. 즉, 3가지 모수의 변화에 모델이 민감하게 반응하며 그 원인의 상당부분은 이 3가지 모수의 특정 모수값이 실제의 도시 성장과 상당부분 불일치하는(따라서, MSE값이 높은) 도시영상을 생성하기 때문이다. 그러나 이 3가지 모수가 생성하는 최소 MSE값(막대그래프의 하단의 값)을 고려할 때, WF 모수가 보다 중요한 의미를 가진다. 그 이유는 subDD와 SR의 최소 MSE는 다른 가중치와 모수들의 최소 MSE값과 거의 유사한 반면, WF는 가장 낮은 최소 MSE값을 보여주기 때문이다. 즉, 이 시기에 대한 민감도분석 결과는 WF 모수의 변화에 따라 실제영상과 가장 일치

하거나 또는 상당부분 불일치하는 도시영상이 시뮬레이션 됨을 암시한다.

<그림 3>의 6개의 그래프는 모든 가중치와 모수 값이 순차적으로 증가할 때 MSE값의 변화를 연결한 MSE곡선을 보여준다. 모든 그래프의 Y축의 값은 0.1~0.3으로 정하여 각 그래프 간의 비교가 가능하도록 하였다. 민감도분석에서 MSE곡선의 형태 역시 주의 깊게 고려되어야 하는데, 이 곡선의 기울기가 급하면 가중치나 모수의 약간의 변화도 MSE의 큰 변화를 야기하기 때문이다. 이런 의미에서 그래프 (f)의 WF모수의 변화는 역시 유의하다. WF가 0~100까지 순차적으로 증가함에 따라 MSE 곡선은 우하향으로 감소하다가 최소 MSE점을 지나 급격하게 증가한다. 결국, 민감도의 크기와 MSE곡선의 형태를 동시에 고려할 때, 이 시기에서는 WF모수의 변화가 모델의 산출물에 미치는 영향이 가장 크다고 할 수 있으며, 이 모수의 변화가 주의 깊게 관찰되어야 함을 암시한다.

또한, 그래프 (c)와 (d)는 uDD~uNBs와 rDD~rNBs 간의 상쇄관계(Trade-Off)를 암시한다. uNBs(rNBs)가 증가하면 MSE가 증가하나,

uDD(rDD)값이 증가하면 MSE가 낮아진다. 이는 모델보정과정에서 두 모수 간의 상호연관의 문제를 야기할 수 있다(후술 참조).

2) 신도시 건설 중 시기(1990~1994)

분당신도시 건설 중 시기의 민감도분석도 신도시 건설 후 시기와 유사하여 <그림 2>의 신도시 건설 후 시기와 유사한 그래프가 도출되었고, subDD, SR, WF의 민감도가 크게 나왔다. 다만, 가중치에 대한 민감도는 약간의 차이를 보였다. 즉, ‘계획지역’(<표 1>참조) 변수가 추가되었으며, 이 변수의 가중치(PW)에 대한 민감도가 상대적으로 크게 나왔고, ‘도시밀도’ 변수의 가중치에 대한 민감도는 상대적으로 작아졌다. 또한 MSE곡선의 형태에 대한 분석결과도 분당신도시의 건설 후 시기와 유사하였다. 이 시기에 대한 민감도 분석결과도 역시 민감도의 크기, 최소 MSE값과 MSE곡선의 형태를 고려할 때 WF 모수의 변화가 모델의 산출물에 가장 민감한 것으로 분석되었다.

2. 가상데이터와 GA를 이용한 모델보정의 정확성(accuracy) 분석

일반적인 상황에서 실제의 도시성장과 일치하는 시뮬레이션 영상을 생성하는 모델의 가중치와 모수값은 알 수 없다. 그러나 가상의 가중치와 모수값을 이용하여 GA를 이용한 NCGM의 보정결과에 대한 정확성(Accuracy)을 검증하였다.²⁾ 이를 위하여 우선, 분당신도시의 건설 중 기간(1990~1994)을 대상으로 GA를 이용하여 NCGM을 보정

하고 그 결과로 ① 1994년의 시뮬레이션된 도시영상과 ② 해당 도시영상을 생성하는 가중치와 모수값을 얻는다. 다음으로 이 두 가지 가상의 산출물을 실제로 간주하고, GA를 이용하여 NCGM을 재보정하고 그 결과 생성되는 도시영상 및 가중치와 모수값이 앞의 두 가지 산출물과 얼마나 정확하게 일치하는지 여부를 측정하였다. 재보정은 재보정 결과의 정확성에 대한 통계적 유의성 검증을 위하여 20번 반복되었고 초기보정과 재보정 시 산출되는 도시영상 간 비교는 MSE를 사용하였다.

<표 2>는 이들 20번의 모델보정과정에서 도출된 최저 MSE를 생성하는 가중치³⁾와 모수값들의 값을 보여준다. 표의 하단에는 이들 값들의 평균값과 함께 가상의 실제값, 그리고 각 보정값들과 가상의 실제값들의 차이를 통해 보정결과의 정확성을 측정하는 통계량인 ‘표준화된 평균제곱오차근(Standardized Root Mean Square Error : SRMSE)’도 표기하였다.

SRMSE는 20번의 모델보정 결과값들과 실제값 간의 평균제곱오차근(Root Mean Square Error : RMSE)을 구한 후, 이 값들을 <표 1>의 각 가중치와 모수값의 범위로 나누어 표준화함으로써 각 가중치와 모수들의 모델보정결과의 정확도를 비교할 수 있도록 하였다.

$$SRMSE_i = \sqrt{\frac{\sum_{r=1}^n (C_{ir} - R_i)^2}{n}} / Range_i$$

여기서, SRMSE_i 는 i 가중치 또는 모수의 표준화된 평균제곱오차근, C_{ir} 은 i 번째 가중치 또는 모수의 r 번째 보정값, R_i는 i 번째 가중치 또는 모수

2) Wang (1997)은 가상의 데이터를 이용한 강우유출량모델(rainfall-runoff model)에서 GA가 최적의 모수값을 찾는지 여부에 대한 분석을 시도하였다.

3) UW, RW, SW, SubW, PW 등 5개 가중치는 상대값으로 각 입력변수의 상대적 중요도를 나타낸다.

표 2_ 가상데이터와 GA를 이용한 NCGM의 보정결과

| RUN | UW | RW | SW | SubW | PW | uDD | rDD | subDD | SR | MaxS | uNBs | rNBs | WF | MSE |
|-------|-------|-------|------|-------|-------|------|------|-------|------|-------|------|-------|-------|----------|
| 1 | 38.64 | 18.18 | 3.79 | 8.33 | 31.06 | 9 | 5 | 9 | 7 | 27 | 5 | 23 | 32.49 | 0.018115 |
| 2 | 34.22 | 17.78 | 4.89 | 4.89 | 38.22 | 4 | 8 | 8 | 8 | 25 | 3 | 50 | 32.49 | 0.018637 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 19 | 30.26 | 28.29 | 9.54 | 0.66 | 31.25 | 4 | 5 | 8 | 9 | 19 | 5 | 33 | 15.21 | 0.024473 |
| 20 | 42.34 | 6.76 | 4.05 | 25.23 | 21.62 | 8 | 6 | 9 | 8 | 23 | 7 | 29 | 16.00 | 0.025358 |
| Mean | 34.54 | 24.03 | 6.41 | 8.85 | 26.03 | 6.24 | 5.57 | 7.00 | 7.62 | 24.86 | 6.33 | 31.71 | 16.75 | |
| Real | 41.08 | 21.62 | 4.32 | 5.41 | 27.57 | 2 | 3 | 6 | 5 | 22 | 2 | 25 | 12.96 | |
| SRMSE | 0.03 | 0.08 | 0.03 | 0.05 | 0.13 | 0.31 | 0.33 | 0.22 | 0.17 | 0.16 | 0.06 | 0.42 | 0.16 | |

의 실제값, n 은 모델보정 횟수, Range_i는 i 번째 가중치 또는 모수값의 범위 모델보정결과의 전반적인 정확도를 보기 위하여 <표 2>의 하단 3개의 행을 취하여 <그림 4>에 표현하였다. 각 가중치 및 모수의 평균보정값은 실제값과 상당히 일치하고 있으며, 이중 5개의 입력변수에 대한 가중치와 uNBs, WF의 SRMSE값들은 낮은 수준으로 실제값에 대한 예측정확성이 높은 반면, uDD, rDD, rNBs, subDD, SR, MaxS는 높은 SRMSE값을 보임으로써 실제값을 상대적으로 부정확하게 예측하였음을 보여준다.

다만, 이들 모수들이 상대적으로 높은 SRMSE값들을 보이는 원인은 다양하게 해석할 수 있는데, 우선, subDD, SR, MaxS은 아마도 이들 모수들과 관련된 입력변수인 ‘지하철접근도’와 ‘경사도’의 가중치가 낮기 때문으로 판단된다⁴⁾. 즉, 각 모수값의 변화는 입력변수의 공간분포형태를 결정하고, 이러한 입력변수의 공간분포는 가중치와 결합하여

최종 산출물의 공간분포형태를 결정하게 되는데, 결국 낮은 가중치를 갖는 입력변수와 연관된 모수값의 변화는 최종 시뮬레이션된 도시영상의 공간분포형태에 미치는 영향이 감소될 것이다. 이러한 원인은 NCGM의 모델보정과정에서 보다 덜 중요한 입력변수와 관련된 모수값들이 높은 SRMSE를 가지는 것을 용인할 가능성이 높다고 할 수 있다.

또한, uDD, rDD, rNBs가 높은 SRMSE를 가지는 원인은 이들 모수 간의 상호 상쇄관계(trade-off)가 있기 때문으로 추측된다. 즉, rDD와 rNBs는 ‘도로접근성’ 변수와 관련된 모수로서 도로접근성 변수는 rDD가 커지면 도로를 중심으로 조밀하게 분포하게 되고 rNBs가 커지면 반대로 도로로부터 넓게 분포하게 된다. 따라서 rDD-rNBs를 각각 X-Y축으로 하는 공간에서는 원점으로부터 우상향하는 대각의 방향으로 일정한 공간분포를 유지하는 ‘도로접근성’ 변수가 생성된다. 이는 같은 도로접근성의 공간분포를 유지하기 위해서는

4) <표 2>에서 지하철접근도(SubW)와 경사도(SW)에 대한 가중치는 10보다 낮아 가중치의 값이 20을 초과하는 기타 입력변수들에 비해 상당히 낮은 수준에 있다.

그림 4_ 각 가중치 및 모수의 평균보정값, 가상의 실제값, SRMSE값

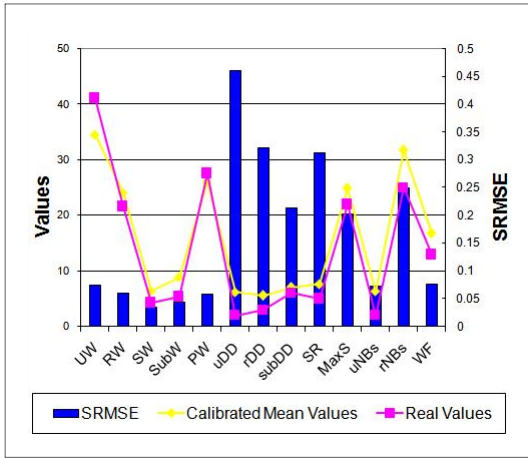
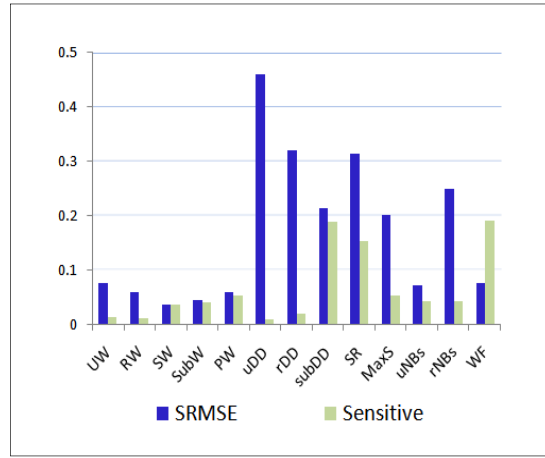


그림 5_ 민감도와 모델보정의 SRMSE (분당신도시 건설 중 시기)



두 모수가 같은 방향으로 동시에 움직여야 함을 의미하고⁵⁾, 이것은 또한 rDD-rNBs의 최적해의 조합이 다수일 수 있음을 암시한다. 결국 주어진 탐색공간에서 유사한 적합도(Goodness of Fits)를 가지는 다수해(multi-solution)가 존재하는 경우라면 이들 모수의 SRMSE값이 커질 것을 예측할 수 있다. 또한 ‘도시밀도’ 변수와 연관되어 있는 uDD-uNBs 모수들도 ‘도로접근성’ 변수와 동일한 상쇄관계가 있으므로 동일한 논리로 높은 SRMSE 값을 보일 수 있다. 이상의 분석을 요약하면, GA를 이용한 모델보정결과의 해석에는 일정부분 다양성이 존재한다. 일부는 실제값을 정확하게 예측한 반면, 일부에 대해서는 높은 오차값을, 또 다른 일부 모수에 대해서는 모델이 유사한 결과를 산출하는 다수의 해를 가지고 있는 것으로 해석된다.

또한, 일반적으로 높은 민감도를 가지는 가중치나 모수에 대해서는 모델보정과정에서 낮은 오차,

즉 높은 정확성을 보이는 것이 바람직하다. 이러한 논리에서 위의 민감도분석과 모델보정결과인 MSE와 SRMSE를 동시에 보임으로써 두 지표 간의 관련성을 분석해 보았으나, 이 두 지표의 움직임에는 어떠한 정 또는 역의 방향의 관련성이 있다고 간주하기 어려운 것으로 판단된다⁶⁾(<그림 5> 참조).

그 원인은 두 분석의 분석환경과 목적이 다르기 때문일 것이다. 우선 분석환경의 차이를 보면, 민감도분석은 다른 모수들이 일정한 값에 고정되어 있고 하나의 모수만이 변화하는 정태적인 환경을 배경으로 하는 반면, GA를 이용한 모델보정은 모든 가중치와 모수가 변화하고 서로 간에 연관되어 있는 동태적인 상황을 배경으로 한다. 흥미로운 것은 이러한 동태적 환경에서 GA를 이용한 모델보정의 경우가 민감도분석보다 낮은 MSE값을 생성한다는 것이다. 신도시 건설 중 시기의 MSE는 민감도분석 시 대략 0.1~0.3의 수치를 보이는데 비해

5) <표 2>를 이용한 rDD-rNBs 간의 상관계수는 0.807(신뢰구간 99%, 양측검정)로 강한 정의 상관을 나타냈으며, uDD-uNBs 간에는 0.648이다.

6) 다만, 민감도분석에서 가장 주의를 요했던 WF모수는 모델보정과정에서 낮은 SRMSE를 보여 실제값을 정확하게 예측하고 있다.

GA를 이용한 모델보정 시에는 <표 2>와 같이 0.018~0.025 수준의 낮은 수치에 머물고 있다. 이는 모델내부의 보정과정에서 서로 상쇄 또는 연관된 변수와 가중치 간의 상호작용이 모델의 오차를 줄이는 역할을 한 것으로 판단된다. 또한, 두 분석의 목적을 고려할 때, 민감도분석은 모델의 입력변수의 변화가 출력에 얼마나 큰 영향을 미치느냐에 대한 사전적·양적인 분석이라면⁷⁾, 모델보정의 SRMSE는 모델의 성과와 관련하여 좋고 나쁨을 측정하는 질적인 분석이라 할 수 있다. 따라서 민감도가 높다는 것이 반드시 모델의 성과와 직접적인 관련이 있는 것은 아니다. 앞의 민감도분석에서 subDD와 SR모수의 민감도가 큰 이유가 높은 오차값(MSE)에 기인하였기에 이것이 모델의 성과와 직접적으로 연계되지 않을 수 있기 때문이다.

3. 실제데이터와 GA를 이용한 모델보정의 일관성(consistency) 분석

GA를 이용한 모델보정의 일관성분석은 분당신도시의 실제 성장과정을 대상으로 하여 이를 건설 후(1994~2000), 건설 중(1990~1994), 건설 전(1985~1990)으로 구분하고, 각각의 시기를 다시 교통인프라시설(도로 및 지하철)의 건설에 대한 예측유무에 따라 추가 구분하여 각 시기별/ 예측유무별로 20번씩 보정하여 그 결과에 대한 일관성을 검증하는 방법으로 시행되었다⁸⁾. 각 시기별 입력 변수와 가중치, 모수는 <표 1>에 정의한 것과 같다.

1) 신도시 건설 후 시기(1994 ~ 2000)

<표 3>은 실제데이터를 이용하여 모델을 보정한 결과 도출된 최저 MSE를 산출하는 최우량 염색체의 유전자(즉, 가중치와 모수값)를 보여준다. 가중치의 값들은 다른 가중치의 값들과 비교한 상대값이다. 다만, 실제 도시영상을 이용한 모델보정에서는 이러한 영상을 발생시키는 가중치와 모수값은 알 수 없으므로 표의 하단은 앞의 모델보정의 정확성 검증 때와는 달리 평균, 중앙값, 각 보정값들의 표준편차와 표준화된 표준편차(Standardized Standard Division : SSD)⁹⁾를 추가하였다. 20번 보정된 가중치와 모수들의 평균과 중앙값들의 값은 거의 일치하는 수준이어서 전반적으로 보정값들의 이상치(Outlier)는 크지 않다고 할 수 있으나, 일부 보정값들은 평균값과의 차이를 보이기도 한다. 따라서 보다 자세한 모델보정의 일관성분석을 위하여 <표 3>의 하단의 SDD와 평균값들을 <그림 6>에 표현하였다.

<그림 6>에서 가중치의 평균값은 '도시밀도(UW)', '도로접근성(RW)', '경사도(SW)', '지하철 접근성(SubW)' 변수의 순서로 높고 이들 4개 가중치에 대한 SDD는 uNBs와 WF모수를 제외한 다른 모수들에 비해 상대적으로 낮은 값을 보인다. 따라서, GA는 이 기간의 도시성장예에 작용한 주요변수를 도시밀도와 도로접근성 변수라고 일관되게 예측하였다고 할 수 있다.

uNBs 모수는 최저수준의 SDD값을 보인다. 또한, 이 모수의 평균보정값도 낮은 수치를 보인다

7) 이러한 의미에서 Kirkby et al (1992)는 민감도분석은 모델개발의 초기단계에서 어떠한 모수값이 모델에 얼마나 큰 영향을 미치는지를 파악하고 따라서 더욱 큰 영향을 주는 모수에 더욱 많은 주의를 기울이기 위한 사전적 분석방법임을 강조하였다.

8) NCGM은 교통시설의 건설에 대한 예측을 가정한 경우에 보다 좋은 성과를 보였고, 본 논문에서는 이 경우에 한정하여 논의를 진행하기로 한다.

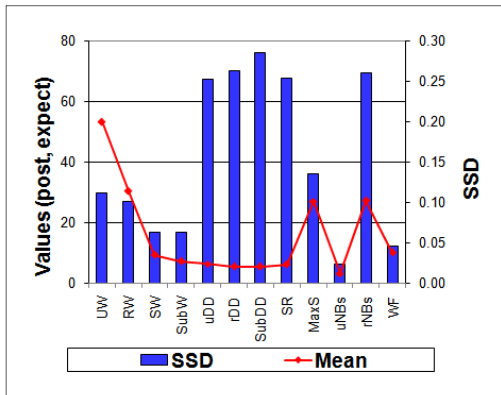
9) 앞의 SRMSE의 경우와 마찬가지로 표준편차를 각각의 범위로 나누어 비교가 가능하도록 표준화하였다.

표 3_ 실제데이터와 GA를 이용한 NCGM의 보정결과 (건설후 기간)

| RUN | UW | RW | SW | SubW | uDD | rDD | SubDD | SR | MaxS | uNBs | rNBs | WF | MSE |
|--------|-------|-------|-------|-------|------|------|-------|------|------|------|-------|-------|----------|
| 1 | 66.23 | 27.81 | 2.65 | 3.31 | 4 | 6 | 2 | 6 | 29 | 2 | 27 | 19.36 | 0.056134 |
| 2 | 67.61 | 25.35 | 5.63 | 1.41 | 4 | 8 | 5 | 3 | 28 | 2 | 49 | 17.64 | 0.056865 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 19 | 54.70 | 20.99 | 1.66 | 22.65 | 3 | 7 | 7 | 9 | 27 | 2 | 45 | 8.41 | 0.06336 |
| 20 | 55.11 | 31.25 | 12.50 | 1.14 | 9 | 3 | 7 | 7 | 20 | 4 | 30 | 7.84 | 0.064603 |
| Mean | 53.28 | 30.31 | 9.28 | 7.12 | 6.35 | 5.45 | 5.45 | 6.15 | 27 | 3.15 | 27.15 | 9.94 | |
| Median | 54.52 | 31.39 | 8.88 | 6.07 | 7 | 6.5 | 6 | 6.5 | 28 | 3 | 28.5 | 9 | |
| SD | 11.16 | 10.13 | 6.25 | 6.24 | 2.52 | 2.63 | 2.86 | 2.54 | 2.83 | 1.14 | 13.02 | 4.53 | |
| SSD | 0.11 | 0.10 | 0.06 | 0.06 | 0.25 | 0.26 | 0.29 | 0.25 | 0.13 | 0.02 | 0.26 | 0.05 | |

김복환·양광식, 2007, “동태적 신도시성장모델의 개발과 검증”. 국토계획 vol. 42, no. 5. p 23.

그림 6_ 각 가중치와 모수의 평균 및 SDD (건설 후 기간)



(3.15). 이는 uNBs값이 20번의 모델보정과정에서 일관성 있게 낮게 예측되었음을 의미하고, 그 공간적 범위는 본 연구에 사용된 격자의 크기가 30m임을 감안할 때 ‘도시밀도’ 변수의 공간분포형태는 기존도시와 100m 이내의 범위에서만 양의 값을 가지는 매우 조밀한 형태일 것으로 예측된다.

WF모수의 평균보정값은 9.94이며 SDD는 0.05로 낮은 값을 유지하고 있다. 이는 역시 GA가 모델

보정과정에서 일관된 WF값을 찾았음을 의미한다. WF모수는 종합점수지도(Composite Score Map)를 개발확률지도(Development Probability Map)로 변환하는 과정에 적용되며(<그림 1> 참조), 종합점수와 개발확률을 X-Y로 하는 평면에서 그 값이 0~1 사이에서는 우상향하는 오목한 지수형태를, 1 이상의 값에서는 S 형태(Sigmoid)의 변환이 일어난다(김복환·양광식, 2007). 즉, WF가 1 이하면 높은 종합점수만이 높은 개발확률을 가지게 되나, WF가 1보다 커질수록 종합점수지도의 낮은 수치만이 개발확률지도에서 낮은 수치로, 종합점수지도의 높은 수치는 개발확률지도에서 더욱 높은 수치로 변환되어 종합점수지도의 낮은 수치와 높은 수치의 차이가 개발확률지도에서는 더욱 증폭되는 결과를 야기한다. 이러한 변환과정은 그림으로 표현하면 더욱 명확해지는데, <그림 7>의 (b), (c), (d)는 (a)의 종합점수지도를 WF가 각각 1, 10, 25로 변환한 경우의 개발확률지도를 보여준다. 즉, (a)지도에서 높은 수치를 보였던 노란색부분은 WF=1인 (b)에서는 최상의 수치들만을 남기

그림 7_ 종합점수지도(Composite Score Map) (a)와 가중요소모수(WF)가 1, 10, 25 인 경우(b), (c), (d)

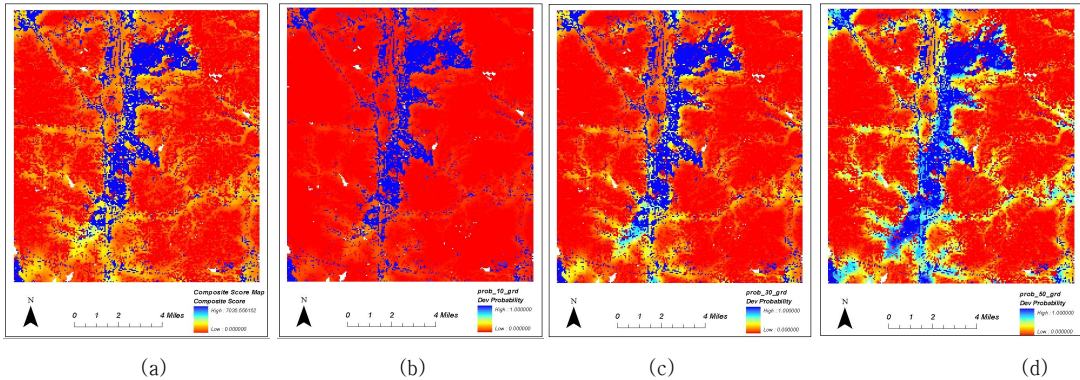
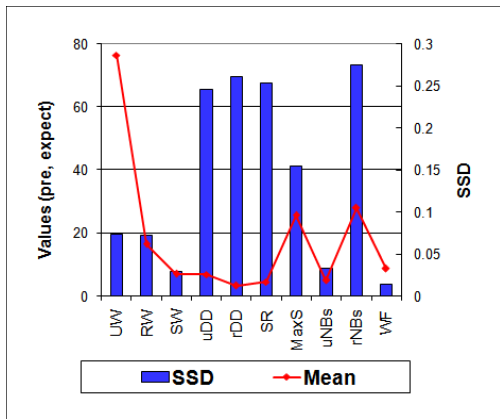
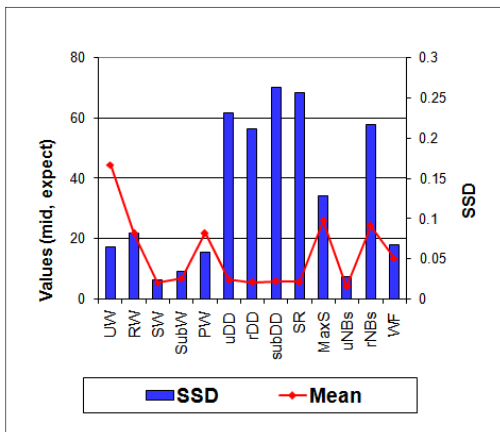


그림 8_ 각 가중치와 모수의 평균 및 SDD



(a) 건설 중



(b) 건설 전

고 거의 사라지나, WF=25인 (d)에서는 높은 종합 점수값을 가진 셀들이 더욱 두드러지게 되어 노란 색부분이 확대된다. 지도 (c)는 WF=10의 경우로서, (b)와 (d)의 중간 정도로 보인다. 즉, GA가 모델보정과정에서 WF가 10 정도의 값을 지속적으로 찾았다면, 분당신도시 건설 후 시기의 도시성장은 지도 (c)에서와 같이 어느 정도 기존도시 주변에서 조밀하게 발생하였다고 할 수 있으나, 그 정도가 (b)와 같이 너무 조밀한 것도, (d)와 같이 너무 넓게 퍼진 형태도 아닐 것으로 예측가능하다.

uDD, rDD, rNBs의 높은 SDD값은 역시 uDD-uNBs, rDD-rNBs 간의 상호 관련성에 기인한 것으로 예측되며, 또한, SR, MaxS와 SubDD의 SDD값이 높은 것 역시 이들의 낮은 가중치와 연계되어 이해될 수 있다.

2) 신도시 건설 중 시기(1990~1994)

신도시 건설 중 시기의 20번의 보정결과에 대한 각 가중치의 평균값과 SDD값에 대한 분석결과 이 시기의 SDD값 역시 전반적으로 가중치는 낮게, 모수는 높게 나와 건설 후 시기와 유사한 결과를 도출하였다(<그림 8-(a)> 참조). uNBs와 WF에 대한 SDD

는 낮은 값을 유지하여 역시 이 두 변수의 보정값은 일관되게 낮은 상태로 예측되었음을 암시한다. 또한, 입력변수에 대한 가중치는 ‘도시밀도’ 변수가 가장 높았고, ‘계획구역’ 변수와 ‘도로접근성’ 변수가 그 다음이었다. 이 기간 중 ‘계획구역’ 변수가 높은 가중치를 보인 것은 의미가 있다. ‘계획구역’ 변수는 더미변수와 같은 역할을 수행하여 신도시계획구역 내는 1의 값을, 외부지역은 0의 값을 가지며, 이 변수의 가중치는 이 변수의 중요도를 의미한다. 따라서 이 변수의 값이 1인 신도시계획구역 내에서 도시성장이 활발히 일어났다면 GA는 이 변수의 가중치값을 높게 예측할 것인데, 보정결과와 같이 신도시건설 중 기간에 ‘계획구역’ 변수의 높은 가중치는 이러한 측면에서 현실적이라고 이해될 수 있다. 분당신도시 건설전 기간의 각 가중치와 모수에 대한 평균값과 SDD값 역시 타 기간과 유사한 SDD의 변화를 보이며, ‘도시밀도’ 변수에 대한 가중치 값이 압도적으로 높다(<그림 8-(b)> 참조).

4. GA를 이용한 모델보정결과의 해석

자연적인 도시의 성장과 달리 신도시의 성장은 신도시건설 중 기간에 계획에 따른 도시성장이 일시적으로 집중되므로 사전적으로 신도시 건설 전, 건설 중, 건설 후의 도시성장의 동인과 패턴은 다를 것으로 예상할 수 있다. 또한, 만일 한 시기에 대해 모델을 반복적으로 보정했는데 그 결과가 일관적이고 다른 시기와는 차별적이라면, 이는 합리적인 범위 내에서 다른 시기와는 차별화된 그 시기만의 독특한 도시성장의 특성을 보여준다고 할 수 있을 것이다. 그러나 일반적으로 GA를 이용하여 찾은 수치값은 적극적으로 해석하지 않는다. 도시성장의 경우만 보더라도, 도시성장은 그 근원인 사람과 사람, 사람과 환경, 사람과 정책 등 다양한 인과관

계에 근거한 것인데, GA는 이러한 도시성장의 원인을 분석하는 과정이 아니라, 실제 도시성장과 가장 유사한 시뮬레이션 영상을 찾아내는 과정에서 도시성장모델의 수치값을 찾아내는 비행태적인 과정이기 때문이다. 그럼에도 불구하고, 어떤 값이 GA를 통하여 일관되게 찾아지고 이러한 수치값을 이용하여 시뮬레이션된 도시영상이 실제의 도시영상과 통계적으로 유의하고, 시각적으로 유사한 결과를 도출한다면 이는 시뮬레이션의 대상이 된 도시시스템 또는 도시성장현상에 대한 특징을 상당 부분 반영한다고 주장할 수 있을 것이다. 이러한 논리와 각 시기별로 GA가 찾은 수치값을 배경으로 신도시의 각 시기별 성장과정에 내재되어 있는 공간적 패턴에 대한 특성을 분석해 보기로 한다.

우선, 도시성장요소에 관해서는 ‘도시밀도’ 변수가 모든 기간 동안 지배적인 영향을 미쳤다고 할 수 있다. 다만 이 변수의 영향력은 각 시기별로 차이가 있는데, 우선, 건설 전 기간에는 가중치가 80에 달할 정도로 높아 가장 중요한 변수인데 반해, 건설 중 기간에는 50 정도로 낮아지고, 대신 ‘계획구역’ 변수에 대한 가중치가 높아진다. 반면, 건설 후 기간 중에는 가중치가 다시 60 정도 상승하며 ‘도로접근성’ 변수에 대한 가중치가 높아진다. 따라서 각 시기별 주요 도시성장요소는 건설 전 기간에는 ‘도시밀도’ 변수, 건설 중 기간에는 ‘도시밀도’와 ‘계획구역’ 변수, 건설 후 기간 중에는 ‘도시밀도’와 ‘도로접근성’ 변수라고 할 수 있다. 그렇다면, 각 시기별 도시성장패턴은 건설 전 기간 중에는 기존도시의 주변에서, 건설 중 기간 중에는 기존도시와 신도시계획구역 내에서 중점적인 도시성장이 이루어졌다고 해석할 수 있다. 또한 신도시는 건설 중 기간에 크게 성장하지만 이후에도 계속해서 도시 성장에 영향을 미치게 되어, 건설 후 기간 중의 도시성장은 신도시지역주변과 도로를 따라 성장하였

다고 해석이 가능하다.

모수값에 관한 보정결과 중에서는 WF가 가장 독특하다. GA는 WF 모수값에 대해 모든 기간 내내 일관된 값을 찾았으며, 이는 모든 기간의 도시성장이 공통적으로 기존도시의 주변지역에서 다소 조밀하게 이루어졌음을 암시한다. 아마도 도시기반 시설이나 교통여건이 상대적으로 좋은 기존 도시 지역의 주변이 최적의 개발입지로 선호되었던 것으로 짐작된다.

GA는 '도시밀도에 대한 이웃의 범위(uNBs)' 모수 역시 모든 기간 내내 일관되게 낮은 값을 예측하였다. 이는 대부분의 도시성장이 다소 기존의 도시 주변에서 발생하였음을 의미한다. 반면, '도로접근성변수에 대한 이웃의 범위(rNBs)'는 보다 높은 값을 찾았는데, 높은 값의 rNBs가 도로를 중심으로 넓은 '도로접근성' 변수의 공간분포를 생성함을 감안할 때, 도로로부터는 다소 멀리에서도 도시성장이 발생하였던 것으로 해석된다. 다만, rNBs는 높은 SDD값을 보이는데 이는 전술한 바와 같이 rNBs-rDD의 상호 관련성에 기인한 것으로 판단되며, uDD의 높은 SDD값도 역시 같은 원인으로 설명된다. SR과 MaxS, subDD도 모든 기간 동안 높은 SDD값을 보이는데, 이것도 역시 전술한 바와 같이 낮은 수준의 가중치와 관련이 있는 것으로 판단된다.

IV. 결론

CA 등 복잡성이론을 이용한 도시성장모델은 최근 시·공간에서 동태적으로 성장하는 도시를 시뮬레이션 하는데 적합한 도구로 각광받고 있다. 그러나 다양한 도시성장의 동인과 이들 간의 상호 관련성, 복잡성이론 내부에 존재하는 불확실성 등은 이러한 이론을 이용한 도시성장모델링의 보정을 어렵

게 한다. 동태적·실증적인 도시성장모델에서 모델의 보정과 검증은 모델의 성과를 좌우하는 필수적인 요소임에도 불구하고, 이에 대한 체계적인 연구는 아직 본격적으로 진행되지 못하고 있다.

본 연구는 CA를 이용한 동태적·실증적 도시성장모델을 인공지능과 휴리스틱 탐색방법의 일종인 GA를 이용하여 보정하였고 그 보정결과의 정확성과 일관성에 대한 분석을 시도하였다. 이에 앞서 정태적 분석이라고 할 수 있는 민감도분석을 실시하였고 이를 GA를 이용한 모델보정과 비교하였다.

가상의 데이터와 GA를 이용한 모델보정의 정확도를 분석한 결과 ① 20번의 모델보정의 평균값들은 가상의 실제값에 상당히 근접하였으며, ② SRMSE값들에 대한 분석에서 가중치의 값들과 주요 모수값들은 상당부분 정확하게 예측되어, GA를 이용한 모델보정의 정확도는 상당히 높음이 확인되었다. 또한, GA를 이용한 모델보정결과가 정태적분석인 민감도분석보다 낮은 MSE를 산출하여 보다 좋은 성과를 보였으나, 두 분석은 그 분석환경과 목적이 다름을 강조하였다. 다만, GA의 정확도분석의 해석에는 다양성이 존재하여, 일부는 실제값을 정확하게 예측한 반면, 일부에 대해서는 높은 오차값을, 또 다른 일부 모수에 대해서는 모델이 유사한 결과를 산출하는 다수해를 가질 수 있음을 보여주었다.

실제데이터와 GA를 이용한 모델보정의 일관성 분석결과 ① 평균보정값과 중앙값이 근접하고, ② 가중치와 주요 모수에 대해서는 낮은 SDD를 보임으로 상당부분 일관된 모델보정값을 찾았다고 할 수 있으나, 일부 모수에 대한 낮은 일관성은 모수 간의 상호 관련성이나 낮은 가중치에서 그 원인을 찾았다. 또한, 가중치에 대한 분석을 통해서 신도시의 성장시기별 다른 도시성장의 동인이 작용했음을 추측할 수 있었고, 이에 대한 해석을 시도하

었다.

본 연구결과, GA는 동태적·실증적 공간모델의 견고한 보정기법으로 사용될 수 있음을 확인하였고, 이는 결과적으로 CA도시성장모델을 성공적으로 보정함으로써 이러한 모델이 도시성장의 동태적 모델링 도구로서 보다 튼튼한 자리매김이 가능함을 확인하였다.

그럼에도 불구하고, CA를 이용한 모델링기법이나 GA를 이용한 최적화기법 등은 현상의 동인(動因)에 대한 고민보다는 시뮬레이션 결과와 실제 현상 간의 결과론적 유사함이 중요한 판단기준이 되는 비행태론에 기초한 다소 검은색에 가까운 회색 상자기법이라고 할 수 있다. 그러나 도시성장이 다양한 원인들이 내부 또는 주변환경 및 정책에 반응하고 적응하면서 진화하는 복잡한 상호식의 순환적 인과관계(Causality)의 결과로서 형성되는 창발(Emergence)과정이라고 한다면 CA 등 복잡성 이론을 이용한 도시모델링은 정당성을 찾는다. 더욱이 최근의 Multi-Agents(MA)는 스스로 행동하는(Autonomous) 개체(Agents)들을 모델링의 기본단위로 사용하여 인간의 선택행위와 유사하게 지능(Intelligent)을 가지고 환경(Environment)과 반응하면서 적응·진화하여 종국적으로 거시적인 형상을 창발한다(Benenson and Torrens, 2004). 이는 CA보다는 더욱 행태론에 근접한 기법으로 도시성장의 행태론적 모델링에 새로운 가능성을 열고 있다. 다만, 이 경우에도 개체들의 행위를 규정하는 다수의 규칙과 이러한 규칙에 포함된 모수들이 어떻게 인간의 행위를 적절하게 반영할 수 있는가의 문제와 그러한 규칙과 모수를 찾아내는 과정은 지속적으로 연구되어야 할 부분이다.

참고문헌

- 김복환·양광식. 2007. "동태적 신도시성장모델의 개발과 검증". 국토계획 제42권 제5호. 서울: 대한국토·도시계획학회. pp7-31.
- 오세진 외. 2005. 인간행동과 심리학. 서울: 학지사.
- Batty, M. 1996. "Visualizing urban dynamics". (ed.) Longley, P. and Batty, M. *Spatial Analysis: Modelling in a GIS Environment*. Cambridge: Geoinformation International. pp297-320.
- Batty, M. and Xie, Y. 1994. "From cells to cities". *Environment and Planning B* vol. 21. pp31-48.
- Batty, M., Longley, P. and Fotheringham, S. 1989. "Urban growth and form: scaling, fractal geometry, and diffusion-limited aggregation". *Environment and Planning A* vol. 21, no. 11. pp1447-1472.
- Benenson, I., Omer, I. and Hatna, E. 2002. "Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv". *Environment and Planning B* vol. 29. pp491-512.
- Benenson, I. and Torrens, P. M. 2004. "Geosimulation: object-based modelling of urban phenomena". *Computers & Environment and Urban Systems* vol. 28, no. 1. pp1-8.
- Birkin, M., Clarke, G., Clarke, M. and Wilson, A. 1996. *Intelligent GIS*. Geoinformation International. Cambridge.
- Clarke, K. C., Hoppen, S. and Gaydos, L. 1997. "A self-modifying cellular automata model of historical urbanization in the San Francisco Bay area". *Environment and Planning B* vol. 24, no. 2. pp247-261.
- Clarke, K. C. and Gaydos, L. 1998. "Loose Coupling a Cellular Automaton Model and GIS: Long-Term Growth Prediction for San Francisco and Washington/Baltimore". *International Journal of Geographical Information Science* vol. 12, no. 7. pp699-714.
- Ghanea-Hercock, R. 2003. *Applied Evolutionary Algorithms in Java*. NY: Springer.
- Heppenstall, A. J. 2005. "Application of Hybrid Intelligent Agents to modelling a Dynamic, Locally Interacting Retail Market". Ph. D. thesis. University of Leeds.
- Kirkby, M. J., Naden, P. S., Burt, T. P. and Butcher, D.

- P. 1992. *Computer Simulation in Physical Geography*. Chichester : John Wiley and Sons.
- Li, X. and Yeh, A. G. 2000. "Modelling sustainable urban development by the integration of constrained cellular automata and GIS". *International Journal of Geographical Information Science* vol. 14, no. 2. pp131-152.
- Li, X. and Yeh, A. G. 2001. "Calibration of cellular automata by using neural networks for the simulation of complex urban systems". *Environment and Planning A* vol. 33, no. 8. pp1445-1462.
- Li, X. and Yeh, A. G. 2002. "Neural-network-based cellular automata for simulating multiple land use change using GIS". *International Journal of Geographical Information Science* vol. 16, no. 4. pp323-343.
- Michalewicz, Z. 1994. *Genetic Algorithms + Data Structure = Evolution Programs*. Berlin: Springer-Verlag.
- Openshaw, S. and Openshaw, C. 1997. *Artificial Intelligence in Geography*. NY: John Wiley and Son.
- Richards, F. C., Meyer, T. P. and Packard, N. H. 1990. "Extracting cellular automaton rules directly from experimental data". *Physica D* vol. 45, no. 1-3: pp189-202.
- Rykiel, E., J. 1996. "Testing ecological models: the meaning of validation". *Ecological Modelling* vol. 90. pp229-244.
- Straatman, B., White, R. and Engelen, G. 2004. "Towards an automatic calibration procedure for constrained cellular automata". *Computers, Environment and Urban Systems* vol. 28, no. 1-2. pp149-170.
- Torrens, P. M. 2000. How cellular models of urban systems work. (1. theory). *CASA working paper* vol. 28. URL http://www.casa.ucl.ac.uk/publications/working_papers.asp.
- Torrens, P. M. and O'Sullivan, D. 2001. "Cellular automata and urban simulation: where do we go from here?". *Environment and Planning B* vol. 28, no. 2. pp163-168.
- Waddell, P. 2003. "Introduction to Urban Simulation: Design and Development of Operational Models". (<http://www.urbansim.org/papers/waddell-ulfarsson-ht-IntroUrbanSimul.pdf>). [2006.12.7]
- Wang, Q. J. 1997. "Using genetic algorithms to optimise model parameters". *Environmental Modelling and Software* vol. 12, no. 1. pp27-34.
- Ward, D. P., Murray, A. T. and Phinn, S. R. 2000. "A stochastically constrained cellular model of urban growth". *Computers, Environment and Urban Systems* vol. 24, no. 6. pp539-558.
- White, R. and Engelen, G. 1993. "Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns". *Environment and Planning A* vol. 25, no. 8. pp1175-1199.
- Wilson, A. G. 2000. *Complex Spatial Systems: The modelling foundations of urban and regional analysis*. Singapore: Pearson Education.
- Wolfram, S. 1984. "Universality and complexity in cellular automata". *Physica D* vol. 10, no. 1-2. pp1-35.
- Wong, S. C., Wong, C. K. and Tong, C. O. 2001. "A parallelized genetic algorithm for the calibration of Lowry model". *Parallel Computing* vol. 27, no. 12. pp1523-1536.
- Wu, F. 1998. "SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules". *International Journal of Geographical Information Science* vol. 12, no. 1. pp63-82.
- Wu, F. 2002. "Calibration of stochastic cellular automata: the application to rural-urban land conversions". *International Journal of Geographical Information Science* vol. 16, no. 8. pp795-818.
- Wu, F. and Martin, D. 2002. "Urban expansion simulation of Southeast England using population surface modelling and cellular automata". *Environment and Planning A* vol. 34, no. 10. pp1855-1876.
- Wu, F. and Webster, C. J. 1998. "Simulation of land development through the integration of cellular automata and multicriteria evaluation". *Environment and Planning B* vol. 25, no. 1. pp103-126.
- Yeh, A. G. and Li, X. 2003. "Simulation of Development Alternatives Using Neural Networks, Cellular Automata, and GIS for Urban Planning". *Photogrammetric Engineering and Remote Sensing*. vol. 69, no. 9. pp1043-1052.

- 논문 접수일: 2007.12. 5
- 심사 시작일: 2008. 1.12
- 심사 완료일: 2008. 2. 4

ABSTRACT

Calibration of Dynamic Spatial Model Using Genetic Algorithms

Keywords: Model Calibration, Genetic Algorithms, Dynamic Spatial Model, Urban Growth Model, Cellular Automata

The attraction of Cellular Automata(CA) urban growth models resides in their simplicity, flexibility, and transparency as a modelling framework, and they have been cited in much of the literature as a promising tool to model realistic urban growth. However, the lack of a rigorous calibration process for CA urban growth models is certainly one potential obstacle to their use. Researchers have found the solution for the calibration problems of CA urban growth model from the field of Artificial Intelligence(AI) equipped with heuristic search routines. AI technique of Genetic Algorithms(GA) can be one of the promising alternatives.

A model for new city growth (the so-called NCGM: New City Growth Model) was developed using stochastically constrained CA. A GA was designed to calibrate the NCGM. Accuracy and consistency of the calibration results through the GA were sought using hypothetical data and real data and meanings for the calibrated values are sought. It was revealed that NCGM produced very reliable results based on the experiments using both hypothetical and real data. Underpinned by GA's accurate and consistent calibration results CA urban growth models, such as NCGM, can be strengthened their position in simulating realistic urban growth.

유전알고리즘을 이용한 동태적 공간모델의 보정

주요단어: 모델보정, 유전알고리즘, 동태적 공간모델, 도시성장모델, 셀룰라오토마타

셀룰라오토마타(CA : Cellular Automata)를 이용한 도시성장모델은 복잡한 도시의 성장을 현실적으로 모델링할 수 있는 도구로서 각광받고 있음에도 불구하고, 복잡한 모델보정과정과 모델보정결과에 대한 신뢰도의 문제 등으로 그 활용이 제약되고 있다. 본 연구는 CA를 이용한 동태적 도시성장모델을 인공지능과 휴리스틱기법을 이용하여 보정하고 그 보정결과의 일관성과 정확성에 대한 검증을 통해 이러한 모델이 도시성장의 동태적 모델링 도구로서 보다 확고한 지위를 확보하는 데 그 목적이 있다.

이를 위하여 확률적 제한요건이 가미된 CA 신도시성장모델(NCGM : New City Growth Model)을 유전알고리즘(GA : Genetic Algorithms)을 이용하여 보정하였고 가상의 자료와 실제 자료를 이용하여 모델보정결과에 대한 정확성과 일관성에 대한 분석을 시도하였다.

가상의 자료를 이용한 NCGM의 보정결과는 매우 신뢰할 수 있는 정도의 정확한 결과를 산출하였으며, 일부 낮은 수준의 정확성을 보이는 원인은 모수 간의 상호 관련성과 각각의 모수가 연계되어 있는 입력변수들의 낮은 중요도 등에서 그 이유를 찾았다. 실제자료를 이용한 모델보정결과도 유사한 결과를 보였다.

연구결과는 유전알고리즘을 이용한 CA 도시성장모델의 보정결과는 정확성과 일관성이 있음이 통계적으로 검증되었고, 이는 NCGM과 같은 CA를 이용한 동태적 도시성장모델이 도시의 성장을 현실적으로 모델링할 수 있는 도구로서 그 지위를 더욱 강화시키는 데 기여할 것이다.