

# 機能地域의 設定과 '空間單位 修正可能性의 問題(MAUP)'

이 상 일\*

## I. 序 論

정보통신기술의 급격한 진보는 현 사회를 총체적으로 특징짓는데 있어 '정보사회(information society)'라는 개념이 가지는 위상을 증대시키고 있다. 이러한 정보사회의 도래는 학문사회에도 지대한 영향을 끼치고 있다. 지리학적 관점에서 볼 때, 연구자들은 이제 인터넷을 통하여 자국이나 세계의 여타 지역에 대한 정보나 자료에 손쉽게 접근할 수 있게 되었다. 또한 세계 각국은 센서스 자료를 비롯한 각종 공간자료가 전자적으로 전달될 수 있는 시스템을 구축하기 위해 노력하고 있다. 여기에 덧붙여 특정한 지리적 정보기술(geographical information technology)로서의 지리정보체계(GIS)의 발달은 지리학의 연구 환경을 급진적으로 변화시키고 있다.

공간자료가 일반자료와 구분되는 본질적인 특징은, 그것이 특정한 공간단위로 구성되며 그 공간단위들은 서로서로 특정한 위상학적 관계를 맺고 있다는 사실이다. 이는 공간단위간의 위상학적인 관계나 더 나아가 공간분석의 결과 그 자체가 공간단위의 선택에 의존적이라는 점을 함축하고 있다. 이러한 측면에서 볼 때, 기본적으로 특정한 공간단위에 의거해 수집, 정리, 유통되는 공간자료는 그것을 사용하는 연구자에게 또 다른 과제를 던져준다. 연구자는 공간단위를 선택해야 할뿐만 아니라 주어진 공간단위를 조작해 새로운 공간단위를 창출해야 하기 때문이다. 특히 한 국가 전체를 대상으로 한

\* 미국 오하이오 주립대 지리학과 박사과정

지역분석 연구에서 이 과정은 필수적이다.

특정한 공간단위를 구축하는 문제는 지리학의 전통적인 연구과제 중의 하나인 '지역구분(regionalization)'과 밀접히 관련되어 있다. 지역구분은 이론적인 의미에서는 큰 스케일의 지역을 작은 스케일의 지역으로 나누는 것이지만, 실질적인 의미에서는 작은 스케일의 지역을 큰 스케일의 지역으로 '合域(spatial aggregation)'하는 것을 의미한다. 특히 '機能地域(functional region)'의 설정은 후자의 경우에 주로 해당된다. 하나의 기능지역은 선정된 변수를 가지고 次下 스케일의 공간단위를 기능지역 구분의 일반 원칙(예를 들어, 지역내 기능적 응집성의 최대화와 지역간 배타성의 최대화)에 따라 합역함으로써 생성된다<sup>1)</sup>. 그런데, 여기서 중요한 것은, 선택된 '合域樣式(aggregation scheme)'에 따라 동일한 자료에 대해 다양한 기능지역이 생성될 수 있다는 점이다. 즉 상이한 합역양식은 상이한 스케일의 기능지역을 생성하기도 하고, 유사한 스케일에서 상이한 구획방식을 가진 기능지역을 생성하기도 한다.<sup>2)</sup>

바로 이 지점에서, 기능지역 구성의 문제는 지리학자들에게 전통적이긴 하지만 무시되어 온 질문을 새롭게 제기하고 있다. 그것은 바로 合域 過程에 본질적으로 내포되어 있는 '空間單位 修正可能性의 問題(MAUP; Modifiable Areal Unit Problem)'이다.<sup>3)</sup> MAUP이란, 연구에 이용되는 공간단위의 선택은 기본적으로 작위적(공간단위의 수정가능성)이며, 연구의 결과는 이용된

---

1) 기능지역 개념의 비교적 최근의 개요로 다음의 논문을 참고할 수 있다. V. T. Noronha and M. F. Goodchild, 1992, "Modeling interregional interaction: implications for defining functional regions", *Annals of the Association of American Geographers*, vol.82, no.1, pp.86~102; 다음의 논문은 기능지역내의 기능적 응집성을 측정하는 문제를 다루고 있다. D. L. Barkley, M. S. Henry, S. Bao, K. R. Brooks, 1995, "How functional are economic areas? Tests for intra-regional spatial association using spatial data analysis", *Papers in Regional Science*, vol.74, no.4, pp.279~316.

2) 혹은 이것이 동시에 발생할 수도 있다. 동일한 수의 공간단위가 상이한 수의 기능지역으로 합역되었다면 이것은 스케일과 구획에서의 차이가 동시에 발생한 것이다.

3) MAUP에 대한 번역으로 이 용어를 사용하였지만, 이하에서는 MAUP을 그대로 쓰기로 한다. MAUP이라는 용어는 1979년 Openshaw와 Taylor에 의해 생성된 용어이지만, 이 문제를 최초로 제기한 것은 1934년 Gehlke와 Biehl에 의한 것으로 알려져 있다. S. Openshaw and P. Taylor, 1979, "A million or so correlation coefficients: three experiments on the modifiable areal unit problem", in N. Wrigley, ed., *Statistical Applications in the Spatial Science*, Pion : London, pp.127~144; C. E. Gehlke and K. Biehl, 1934, "Certain effects of grouping upon the size of the correlation coefficient in census tract material", *Journal of the American Statistical Association*, vol.29, Supplement, pp.169~170.

공간단위에 의존적임을 지적하는 것이다. 즉, 분석의 결과로 도출된 통계치(예를 들어 변수간 상관관계나 회귀분석에서 회귀계수나 결정계수의 값)들이 선택된 공간단위가 달라짐에 따라 변화한다는 것이다. 따라서, 극단적인 경우, 현상간 결합관계가 공간단위의 선택에 따라 正的일 수도 있고 的일 수도 있다. 만일 변수들의 공간관계에 대한 통계적 해석이 채택된 공간단위에 대해 '상대적'이라면, Openshaw가 적절히 지적한 것처럼, "MAUP은 현재 공간분석이 직면하고 있는 가장 중요한 未題 중 하나"<sup>4)</sup>임에 틀림없다. 더 나아가 MAUP은 왜 우리가 '일반' 통계학이 아닌 '공간' 통계학을 필요로 하는지에 대한 궁극적인 질문을 지향하고 있다. '지리적 자료 (geographically referenced data)'는 '일반' 자료와 달리 분석단위가 서로 서로에게서 독립적이지 않다. 예컨대, 자연과학에서의 일회의 실험은 다른 실험에 대해 대체적으로 독립적이지만, 한 단위공간은 또 다른 단위공간에 대해 독립적이지 않다. 이러한 자료의 '空間 依存性 (spatial dependency)' 혹은 '空間的 自己相關 (spatial autocorrelation)'은 MAUP과 밀접히 관련되어 있으며, 공간통계학의 가장 중요한 존재이유가 되고 있다.

이러한 관점에서, 본 연구는 다음의 두가지 연구 목적을 가진다. 첫째, 미국 지역구분의 현황을 기능지역의 설정을 중심으로 살펴보는 것이다. 여기에서 기능지역이란 카운티(county)를 기반으로 한 지역구분으로 한정되는 데, 다양한 함의양식의 예들이 제시될 것이다. 둘째, MAUP을 미국 기능지역을 중심으로 그 실재를 탐구하는 것이다. 제시된 기능지역 중 몇 개가 MAUP의 영향에 대한 실증적인 연구를 위해 선택될 것이다. 이를 통해 MAUP 연구가 앞으로 행해야 할 과제에 대한 제안을 하고자 한다.

## II. 미국에서의 機能地域의 設定

### 1. 미국 지역구분의 기초

여느 국가의 센서스와 마찬가지로 미국 센서스의 기본 단위는 개인이다. 개인 정보는 극히 예외적인 경우를 제외하고는<sup>5)</sup>, 특정한 공간단위에 담겨

4) S. Openshaw, 1984, *The Modifiable Areal Unit Problem, Concepts and Techniques in Modern Geography* 38, Norwich : Geo Books, p.6.

5) PUMS (Public Use Microdata Samples)는 센서스 응답의 1%, 3%, 그리고 5%의 표본에

일반인에게 제공된다. 이런 의미에서 미국에서의 공간정보의 최소단위는 센서스 블락(census block)이다. 1990년 센서스 당시 총 7,017,425개의 센서스 블락이 존재했다. 센서스 블락은 카운티(county)로 합역되고, 카운티는 주(state)로 합역된다. 이런 의미에서 미국 지역구분의 기초는 세 층위로 구분되어 있다고 보아야 할 것이다.

#### 1) 센서스 블락-기반 공간단위

표 1에는 센서스 블락에 기반하고 있는 공간단위가 열거되어 있다. 센서스 블락그룹(block group)과 센서스 트랙트(census tract)는 그 중에서도 가장 기초적인 공간단위에 속한다. 몇 개의 센서스 블락이 한 개의 블락그룹을 구성하고, 몇 개의 블락그룹이 하나의 트랙트를 형성한다. 그러므로 이 세 가지 공간단위는 완전한 포섭관계에 있다. 이중 센서스 트랙트는 도시내부구조에 대한 공간분석단위로 가장 널리 사용되어 왔다. 그러나 이외에도 센서스 블락에 기반하고 있는 공간단위는 의외로 많이 있다. 미국 교통부(U.S. Department of Transportation)는 TAZ(transportation analysis zone) 단위로 센서스 자료를 포함하여 TAZ간 통근자료를 제공한다. TAZ 단위로 구성된 통근 자료는 CTPP(Census Transportation Planning Package)라는 이름으로 CD에 담겨 무료로 배포되었다.<sup>6)</sup> 선거구(Voting District)는 또 다른 센서스 블락 기반 공간단위이다. 이것은 주지사나 지방 행정직을 선출하기 위해 구성되어 있는 공간단위이다. 그림 1의 (c)는 센서스 블락-기반 공간단위의 예를 오하이오주 Franklin 카운티를 대상으로 보여주고 있다.

센서스 블락-기반 공간단위이면서도 카운티의 경계에 의지하지 않는 공간 단위가 있는데 그것이 PUMA(Public Use Microdata Area)와 Zip Code

---

대해 개인단위로 제공되는 자료이다. PUMS에 대해서는 미네소타대 역사학과에서 운영하고 있는 Integrated Public Use Microdata Series Project의 홈페이지(<http://www.ipums.umn.edu/>)를 참고할 수 있다.

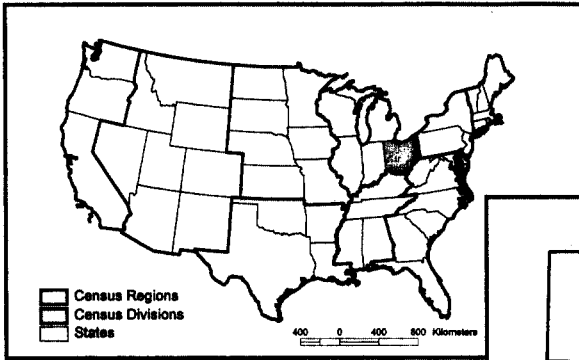
6) JHK & Associates, 1995, *CTPP Handbook: An Instructional Guide to the 1990 Census Transportation Planning Package*, Federal Highway Administration, U.S. Department of Transportation; TAZ를 이용한 연구로 다음 논문을 참고할 수 있다. E. K. Wyly, 1999, "Local labor markets and occupational sex segregation in an America metropolis", *Journal of Urban Affairs*, vol.21, no.1, pp.1~33; TAZ의 설정과 그것과 관련된 합역의 문제에 대해 다음 논문을 참고 할 수 있다. C. Ding, 1998, "The GIS-based human-interactive TAZ design algorithm : examining the impacts of data aggregation on transportation-planning analysis", *Environment and Planning B: Planning & Design*, vol.25, no.4, pp.601~616.

표 1. 미국 지역구분의 기초와 기능지역

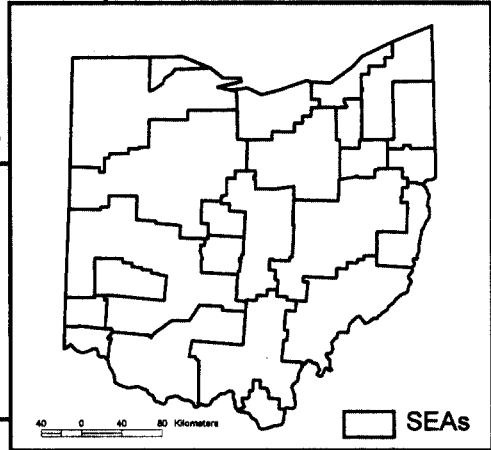
구 분	약 어	설 명	개수	설정기관 혹은 회사
블락-기반 공간단위	CBG	Census Block Group	229,192*	센서스국 (BOC)
	TAZ	Transportation Analysis Zone	200,000*	센서스국 (BOC)
	VD	Voting District	148,872*	센서스국 (BOC)
	CT	Census Tract	50,690*	교통부 교통통계국 (BTS)
	ZIP	Zip Code Area	29,469*	센서스국 (BOC)
	SD	School District	14,422*	센서스국 (BOC)
	PUMA	Public Use Microdata Area	1,726	센서스국 (BOC)
카운티-기반 공간단위 (기능지역)	CZ	Commuting Zone	741	농업부 (Tolbert & Sizer)
	CMA	Celluar Market Area	734	연방통신위원회 (FCC)
	SEA	State Economic Area	506	센서스국 (BOC)
	BTA	Basic Trading Area	493	Rand McNally
	CD	Congressional District	435*	센서스국 (BOC)
	LMA	Labor Market Area	394	농업부 (Tolbert & Sizer)
	DMA	Designated Market Area	211	Nielsen Media Research
	ADI	Area of Dominant Influence	211	Arbitron Company
	BEA	BEA Economic Area	172	경제분석국 (BEA)
	NTAR	National Transportation Analysis Region	89	교통부 교통통계국 (BTS)
	PSR	Public Safety Planning Region	55	연방통신위원회 (FCC)
	MEA	Major Economic Area	52	연방통신위원회 (FCC)
	MTA	Major Trading Area	51	Rand McNally
	VPC	VHF Public Coast Station Area	42	연방통신위원회 (FCC)
	REA	Regional Economic Area	12	연방통신위원회 (FCC)
	EAG	Economic Area Grouping	6	연방통신위원회 (FCC)
RPC	Regional PCS Area	5	연방통신위원회 (FCC)	
주-기반 공간단위	CR	Census Division	9	센서스국 (BOC)
	CD	Census Region	4	센서스국 (BOC)

\* Bureau of Census, 1993, A Guide to State and Local Census Geography, U.S. Department of Commerce, pp.14~15

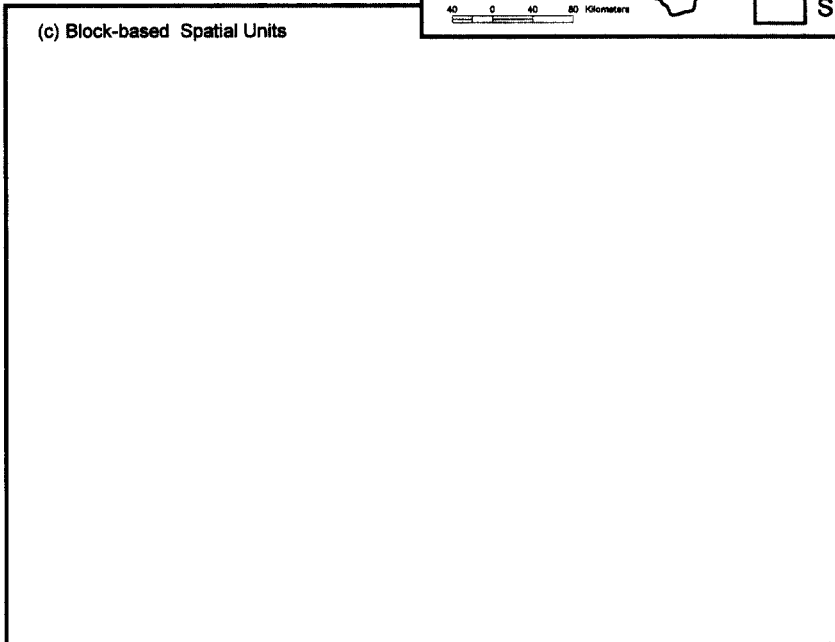
Area이다. PUMA는 PUMS가 제공하는 고유의 지역단위이다(각주 5 참조). 인구가 적은 지역에서는 두 개 이상의 카운티가 합쳐져서 하나의 PUMA를 구성하기도 하지만, 인구가 밀집한 도시 지역에서는 센서스 블락이 합쳐진 카운티의 일부가 하나의 PUMA를 구성한다. 그리고 각 표본(1%, 3%, 5%)



(a) State-based Spatial Units



(b) County-based Spatial Units



(c) Block-based Spatial Units

그림 1. 미국 지역 구분의 세가지 층위

에 대해 다른 PUMA의 구성을 보여주고 있다. PUMS 자료는 주로 개인별 자료를 많이 다루는 사회학 분야에서 널리 사용되었고 지리학 분야에서도 특히 인구가동과 통근 관련 연구에 많이 이용되었지만<sup>7)</sup>, 연구의 공간단위로서 PUMA가 사용된 논문의 예는 드물다.<sup>8)</sup>

미국의 우편구역이라고 할 수 있는 Zip Code Area는 카운티의 경계를 넘어 설정될 수 있지만 주 경계를 넘어설 수는 없다. 다시 말하면, Zip Code Area는 카운티보다 훨씬 작지만 센서스 블록에 기반하고 있기 때문에 인접 카운티의 다른 블록과 결합하여 하나의 Zip Code Area를 구성할 수 있다. 하지만 zip area는 기본적으로 주의 하위 공간단위이다.<sup>9)</sup> (그림 1의 (c)와 (b)를 비교해 보라)

표 1에서는 제외되어 있지만 미국의 市域 역시 블록을 기반으로 한 공간단위로 볼 수 있다. 그림 1의 (c)에서 볼 수 있듯이 미국 대부분의 도시는 시역 내부의 군데군데 아직 시 지역으로 포섭되지 않은 지역(unincorporated land)을 가지고 있다. 현 미국 도시의 행정 경계는 주변지역에 대한 지속적인 행정포섭(annexation)을 통해서 성장해 온 것인데, 未포섭지역은 도시내부에 마치 섬처럼 존재하고 있는 것이다.<sup>10)</sup>

---

7) 인구가동 연구와 통근 연구 각각에 대해 다음 논문을 참고할 수 있다. K. B. Newbold, 1998, "Outmigration from California : The role of migrant selectivity", *Geographical Analysis*, vol.30, no.2, pp.138~152; S. McLafferty and V. Preston, 1997, "Gender, race, and the determinants of commuting : New York in 1990", *Urban Geography*, vol.18, no.3, pp.192~212.

8) PUMA를 지역단위로 공간패턴을 연구한 논문으로 다음의 논문을 참고 할 수 있다. D. W. S. Wong, 1998, "Spatial patterns of ethnic integration in the United States", *Professional Geographer*, vol.50, no.1, pp.13~30; PUMA의 구획에 대한 논의는 다음의 논문을 참고할 수 있다. M. Ellis, M. Reibel, and R. Wright, 1999, "Comparative metropolitan area analysis : Matching the 1980 and 1990 Census Public Use Microdata Samples for metropolitan areas", *Urban Geography*, vol.20, no.1, pp.75~92.

9) zip code area는 특히 도시사회학 연구에서 널리 이용되고 있는데, 그것은 zip code area가 community의 공간 범위를 가장 잘 대변하고 있다고 믿기 때문이다. zip code area에 기반한 연구의 예로 다음의 논문을 참고할 수 있다. D. O. Lee, 1995, "Koreatown and Korean small firms in Los Angeles locating in the ethnic neighborhoods", *Professional Geographer*, vol.47, no.2, pp.184~195; 또한 zip code area는 시장조사를 위해 가장 애용되는 공간단위 중 하나인데, 센서스국에서 zip code area 별로 산업구조자료(Zip Code Business Patterns)를 발간할 뿐만 아니라, 상업적으로도 다양한 자료집들이 발간되고 있다. (<http://tier2.census.gov/zbp/index.html-ssi>)

10) M. I. Glassner, 1996, *Political Geography*, New York : John Wiley & Sons, pp.183~190.

## 2) 카운티-기반 공간단위

카운티-기반 공간단위는 지역분석에서 가장 일반적으로 사용되는 데, 특히 지방노동시장 (local labor market), 시장지역 혹은 상권(market area), 혹은 도시세력권 등을 구축하는데 사용된다. 그러므로 일반적인 기능지역 개념에 가장 잘 부합하는 공간단위라 볼 수 있으며, 본 연구에서 핵심적인 위치를 차지할 것이다. 미국의 메트로폴리탄 지역(metropolitan area)을 구성하는 것도 카운티이지만 미국 전체를 포괄하지 않는다는 단점 때문에 여기서는 논외로 한다.<sup>11)</sup>

미국의 카운티 개수는 준카운티(county-equivalent)를 포함하여 1990년 현재 3,141개이다. 이 카운티를 합역하는 다양한 방식에 따라 다양한 기능지역이 존재한다.(표 1 참조) 이 중에서 몇 개는 다음 장에서 자세하게 다루어질 것이다.

그림 1의 (b)는 카운티-기반 공간단위의 예로서 SEA(State Economic Area)를 오하이오 주를 대상으로 보여주고 있다. 대도시 지역에서는 중심카운티(central county)<sup>12)</sup> 하나가 하나의 SEA를 구성하고 있는 반면, 농촌지역에서는 여러개의 카운티가 하나의 SEA를 구성하고 있음을 볼 수 있다.

## 3) 주-기반 공간단위

주-기반 지역구분의 대표적인 것으로 미국 센서스국이 규정하는 센서스地域(Census Region)과 센서스 分域(census division)이 있다. 미국 전체는 4개의 센서스 지역(북동부, 중서부, 남부, 서부)으로 나누어지며, 각 센서스 지역은 그 아래에 2-3개의 센서스 分역을 가지고 있다 (그림 1의 (a) 참조). 이 대 분류법은 그것이 가지는 명백한 한계에도 불구하고 특히 지역경제학자

11) 메트로폴리탄 지역은 가장 일반적으로 알려진 기능지역이며, 도시 체계연구의 기본 단위로 이용될 뿐만 아니라, 메트로폴리탄 지역을 중심카운티(central county)와 근교카운티(suburban county)로 나눔으로써 도시 내부구조 연구에서도 널리 쓰여왔다. 메트로폴리탄의 카운티 구성은 지속적으로 변해 왔으며 1999년 6월 30일 마지막으로 개정되었다. 개정된 안에 따르면 현재 미국에는 푸에르토리코의 7개 메트로폴리탄 지역을 포함하여 모두 356개의 MAs(Metropolitan Areas)가 있는데, MAs는 다시 261개의 MSAs(Metropolitan Statistical Areas), 19개의 CMSAs(Consolidated Metropolitan Statistical Areas), 그리고 76개의 PMSAs(Primary Metropolitan Statistical Area)로 구분된다. 메트로폴리탄 지역의 개념과 구성방식에 대해서는 다음의 웹사이트를 참고할 수 있다.

<http://www.census.gov/population/www/estimates/aboutmetro.html>.

12) 메트로폴리탄 지역에서 중심도시(central city)가 위치해 있는 카운티를 의미한다. 그림 1의 예에서 Columbus시는 중심도시이고, Franklin카운티는 중심카운티이다.



및 경제지리학자들에 의해 널리 이용되어 왔다. 이 밖에도 일반 기업체들은 독자적인 주-기반 지역구분을 이용하여 효율적인 공간경영을 추진하고 있다.<sup>13)</sup>

## 2. 미국에서의 機能地域의 設定

이 장에서는 카운티-기반 지역구분 중 중요한 기능지역과 관련하여, 그것의 설정방식에 대해 간략히 소개하고자 한다. 표 1에서 보는 다양한 기능지역은 두 가지 범주로 구분가능한데, 지방노동시장의 개념에 기반을 둔 '경제지역 (economic area)'과 도시세력권 개념에 기반을 둔 '시장지역 (market area)'이다. 하지만 실질적으로 기능지역은 이 두 가지 성격을 어느 정도 공유하고 있다.

### 1) 經濟地域 (Economic Area)

'지역노동시장(local labor market)' 개념은 사람들의 주거지와 직장이 통근을 통해 통합되어 있는 하나의 공간단위로 정의할 수 있다. 따라서 이론적 의미에서는, 정의된 지역노동시장 내의 모든 거주자는 그 지역 내에서 직장을 가지고 있어야 한다. 지역노동시장은 동시에 지역경제의 공간단위가 되기 때문에 지역간 경제활동 비교를 위해 필수적인 공간구획이라 할 수 있다.

지역노동시장 개념에 기반을 두고 있는 대표적인 경제지역은 Killian과 Tolbert에 의한 LMA(Labor Market Area)이다.<sup>14)</sup>(그림 2의 (a) 참조) LMA의 구획은 기본적으로 카운티 간 통근에 대한 표본자료에 기반을 두고 있다. 이 통근자료에 의거해 구성된 통근행렬(flow matrix)에 계층적 군집 분석(hierarchical cluster analysis)을 적용하여 일차적으로 741개의 CZ(Commuting Zone)을 얻고, 인구 100,000명 이상이라는 기준과 통근권간의 통근량을 적용하면 최종적으로 394개의 노동시장지역이 도출된다.<sup>15)</sup> 이

13) 하나의 예로서, 미국의 최대 인터넷 대학교재 서점의 하나인 Varsity Books는 미국을 6개의 지역으로 나누어 관리하고 있다. <http://www.varsitybooks.com/>에서 지역구분의 지도를 살펴볼 수 있다.

14) M. S. Killian and C. M. Tolbert, 1993, "Mapping social and economic space : the delineation of local labor markets in the United States", in J. Singelmann and F. Deseran, Boulder, eds., *Inequalities in Labor Market Areas*, Westview Press, pp.69~79.

15) 노동시장지역 설정에 대한 보다 기술적인 절차에 대해서는 다음의 글을 참고할 수 있다. C. M. Tolbert and M. S. Killian, 1996, *U.S. Commuting Zones and Labor Market*

## 그림 2. 기능지역의 예(경제지역과 시장지역)

---

*Areas : A 1990 Update*, Rural Economy Division, Economic Research Service, U.S. Department of Agriculture, Staff Paper 9614, Washington, D.C. : U.S. Government Printing Office. 이 보고서는 다음의 웹 페이지에서 이용가능하다. [http://lapop.lsu.edu/czlma\\_docs.html](http://lapop.lsu.edu/czlma_docs.html).

지역구분은 미국 전체를 대상으로 한 사회경제학적 연구에서 비교적 널리 이용되고 있다.<sup>16)</sup>

경제지역의 범주에 묶일 수 있는 두 번째의 기능지역은 미 경제분석국(Bureau of Economic Analysis)에서 설정한 BEA(BEA Economic Area)이다. BEA를 설정하는 방식은 노동시장지역 과 통근자료를 이용한다는 측면에서 유사하지만, 기존에 규정된 메트로폴리탄 지역에 일차적인 기반을 둔다는 데 그 차이점이 있다. 3,141개의 카운티 중 836개가 메트로폴리탄 지역을 구성하고, 각 메트로폴리탄 지역을 구성하는 카운티 중 중심이 되는 카운티를 결절지(node)로 선정한 후, 2,305개의 비메트로폴리탄 카운티 중 통근자료에 의거해 결절지로 선정된 카운티를 합하여 348개의 결절지를 결정한다. 그 결절지와 주변의 카운티를 합하여 일차적으로 CEA(Component economic area)가 설정되고, 인구규모와 통근자료를 이용하여 최종적으로 172개의 경제지역이 도출된다.<sup>17)</sup> 경제분석국은 이 BEA를 기반으로 다양한 연구보고서를 발표하고 있으며,<sup>18)</sup> 지역경제승수 계산을 위한 투입-산출 행렬에 대한 기초 자료를 경제지역에 기반하여 제공하고 있다.<sup>19)</sup> 미 연방통신위원회(FCC; Federal Communications Commission)가 규정하고 있는 MEA(Major

---

16) Haynie의 연구는 통근권을 이용했다. 나머지는 노동시장지역에 기반한 연구 중 중요한 것들이다. D. L. Haynie, 1996, "A spatial analysis of occupational and industrial sex segregation across labor markets in 1990", Paper presented at the annual meeting of the American Sociological Association in New York (August); J. E. Kodras and I. Padavic, 1993, "Economic restructuring and women's sectoral employment in the 1970s: a spatial investigation across 380 united states labor market areas", *Social Science Quarterly*, vol.74, no.1, pp.1~27; J. Singelmann, F. A. Deseran, F. C. Mencken, and J. H. Li., 1993, "What drives labor market growth?: economic performance of labor market areas, 1980-86", in Singelmann and Deseran, *op. cit.*, pp.125~142.

17) K. Johnson, 1995, "Redefinition of the BEA Economic Areas", *Survey of Current Business*, February, pp.75~81.

18) Survey of Current Business는 경제분석국의 기관지인데, 매년 개인 임금과 각종 경제 통계를 경제지역별로 발표하고 있다. 최근 보고서 중의 한 예로 다음의 논문을 참고할 수 있다. W. K. Bailey, 1999, "Local area personal income, 1982-97", *Survey of Current Business*, May, pp.50-77 (<http://www.bea.doc.gov/bea/regional/articles.htm>); 지리학에 BEA 경제지역이 사용된 예는 드물지만, 다음의 논문이 그 중 하나이다. W. B. Beyers, 1985, "Trends in the producer services in the USA : the last decade", in P. W. Daniels, ed., *Services and Metropolitan Development: International Perspectives*, New York : Routledge, pp.146~172.

19) 경제분석국은 Regional Input-Output Modeling System (RIMS II)을 운영하고 있는데, 이것에 대한 개요는 다음의 웹페이지에서 찾아 볼 수 있다. <http://www.bea.doc.gov/bea/regional/rims/>.

Economic Areas)는 172개의 BEA 경제지역을 통신서비스의 권역개념에 기반하여 52개 기능지역으로 재구성한 것이다.<sup>20)</sup> 역시 연방통신위원회에서 규정하고 있는 EAG (Economic Area Grouping)는 더 나아가 MEA를 6개의 권역으로 합역한 것이다. 이와 유사하게 미 교통부(U.S. Department of Transportation)의 교통통계국(BTS; Bureau of Transportation Statistics)은 BEA를 89개의 NTAR(National Transportation Analysis Region)로 재구성하여 지역간 인적, 물적 유동량에 대한 정보를 정리 발표하고 있다.<sup>21)</sup>

경제지역의 범주에 포함될 수 있는 세 번째 기능지역은 앞서도 언급한 SEA(State Economic Area)이다. SEA는 경제활동에 있어서의 유사성에 기반하여 몇 개의 카운티를 합한다는 점에서 위에서 논의한 여타 경제지역과 비슷하지만, 주 경계를 넘어서서 설정될 수 없다는 점에서 현실적인 적용성은 비교적 낮다고 볼 수 있다. (그림 1 참조)

## 2) 市場地域 (Market Area)

시장지역의 범주에 속할 수 있는 첫 번째 예는 세계 최대 지도제작사 중의 하나인 Rand McNally사에서 규정한 BTA(Basic Trade Area)와 MTA (Major Trade Area)이다. Rand McNally는 1992년 미국전역을 493개의 BTA로 나눈 후, 그것을 다시 합역하여 51개의 MTA를 구성하였다. 따라서 BTA와 MTA는, CZ와 LMA처럼, 완전한 계층적 포섭관계에 있다. 이 지역 구분을 미연방통신위원회에서 1994년 수용하였다. BTA와 MTA의 이용은 Rand McNally사와 미연방통신위원회 간에 큰 차이가 있는데, Rand McNally는 시장수요분석을 위한 기초자료를 제공할 목적으로 사용하는 반면,<sup>22)</sup> 미연방통신국은 PCS(Personal Communications Services)와 같은 통신

20) 미 연방통신위원회에서는 MEA외에도 표 1에서 보는 것과 같이 방송이나 통신서비스 (TV와 휴대전화)의 도달 범위를 중심으로 미국전체를 다양한 '통신지역(communication area)'으로 구분하고 있다 (<http://www.fcc.gov/oet/info/maps/areas/>).

21) 권역설정에 대한 계략적인 설명은 <http://www.bts.gov/programs/cfs/ntar/ntar-defined.html>에서 찾아 볼 수 있다. 권역간 유동량에 대한 통계와 분석은 다음책의 7장을 참고할 수 있다. Bureau of Transportation Statistics, 1998, *Transportation Statistics Annual Report, 1998 : Long-Distance Travel and Freight*, U.S. Department of Transportation (다음의 웹페이지에서 책 전체에 대한 내용을 살펴볼 수 있다. <http://www.bts.gov/programs/transtu/tsar98/tsar98pt.html>).

22) Rand McNally, 1999, *Commercial Atlas & Marketing Guide*, 130th Edition, Chicago : Rand McNally.

서비스의 지역적 허가와 관리를 위해 사용하고 있다.<sup>23)</sup>

두 번째 시장지역의 예는 Nielsen Media Research사가 규정한 DMA (Designated Market Area)와 Arbitron Company가 규정한 ADI(Area of Dominant Influence)이다. DMA와 ADI는 모두 자체 조사한 지방 TV방송의 세력권에 대한 자료에 기반하고 있으며 미국 전체를 211개의 소지역으로 구분하고 있다. 두 회사는 각 DMA나 ADI별로 인구나 경제활동에 대한 기본 자료뿐만 아니라 각 지역이 가지고 있는 생활양식이나 구매력에 대한 정보를 제공함으로써 특히 상당한 범위의 지역에 대해 사업을 하고 있는 기업들에게 큰 영향력을 행사하고 있다.<sup>24)</sup>

DMA와 ADI와 같은 시장지역은 Reilly가 제시하였던 '상권분할의 원칙 (law of retail gravitation)'에 가장 부합된다는 점에서 사업지리학(business geography)의 분석에서 이용될 여지가 많다고 볼 수 있는데, 놀랍게도 그간 이 지역구분을 이용한 지리학관련 논문은 거의 없다.

### III. MAUP의 영향과 機能地域

#### 1. MAUP의 原因과 영향

최근 MAUP에 대한 연구가 활발하게 진행되고 있는 것은 Openshaw를 비롯한 몇몇 학자들의 꾸준한 연구성과에 힘입은 바 크지만, 실질적으로는 1980년대 후반 NCGIA(National Center for Geographic Informatin and Analysis)가 MAUP을 하나의 연구과제(research initiative)로 선정하였기 때문이다.<sup>25)</sup> MAUP에 대한 최근 일련의 연구는, 수많은 지리학적 연구가 얼마나 빈번하게 그 자료 자체에 대한 신중한 검토없이 공간적 패턴에 대한

---

23) <http://www.fcc.gov/wtb/pcs/bbftsh.html>.

24) DMA에 대해서는 다음의 자료집을 참고할 수 있다. Market Statistics, 1995, *Demographics USA 1995 : City Edition*, New York : Market Statistics; SRDA, 1998, *The Lifestyle Market Analysis, Des Plaines, IL* : SRDS; 다음의 책은 ADI를 바탕으로 미국의 지역별 생활양식을 심도있게 연구하였다. M. J. Weiss, 1988, *The Clustering of America*, New York : Harper & Row.

25) D. W. S. Wong and C. Amrhein, 1996, "Research on the MAUP: old wind in a new bottle or real breakthrough?", *Geographical Systems*, vol.3, p.73; 연구과제의 성과물은 다음의 책으로 발간되었다. M. F. Goodchild and S. Gopal, eds., 1989, *Accuracy of Spatial Databases*, London : Taylor and Francis.

결론을 내려왔는가를 지적하고 있으며, 더 나아가 타학문으로부터 여과없이 받아들여 사용하고 있는 모든 통계방법들에 대해 그것이 지리적 자료에도 과연 타당한 것인지에 대한 궁극적인 질문을 던지고 있다.<sup>26)</sup> 본질적으로 비공간적인 통계기법들을 공간화시키는 작업은 실로 1990년대 최소한 계량지리학계의 최대 과제였고, 당분간 이러한 흐름은 지속될 것이다.<sup>27)</sup> MAUP은 명백히 공간통계학을 정립하기 위한 과정에서 발생할 수밖에 없었던 필연적인 문제제기이다.

### 1) MAUP의 定意와 原因

MAUP은 상이한 단위공간의 이용이 동일한 자료에 대해 상이한 통계적 결과를 낳는다는 것을 지적한다. 상이한 단위공간의 효과는 두 가지로 나누어지는데, '스케일의 效果(scale effect)'와 '區劃의 效果(zoning effect)'이다. 위에서 살펴 본 미국의 기능지역 구성의 예를 보면, 카운티를 합하여 훨씬 적은 수의 기능지역을 구성할 때, 스케일의 효과는 보다 많은 공간단위를 가진 기능지역과 보다 적은 수의 공간단위를 가진 기능지역간에서 발생하며(예를 들어 CZ와 LMA), 구획의 효과는 비슷한 수의 지역단위를 가진 상이하계 정의된 기능지역간에서 발생한다(예를 들어, DMA와 ADI). MAUP에서 '수정가능한(modifiable)'이라는 용어는 한 연구에서 채택된 합역양식은 본질적으로 도출가능한 수많은 합역양식들 중 하나이므로 그 합역양식의 선택은 임의적이며 다른 합역양식으로 대체될 수도 있음을 의미하는 것이다.<sup>28)</sup> N개의 공간단위를 보다 적은 수의 M개의 공간단위로 합역하는 모든 가능한 방법의 개수는 얼마나 될까? 그리고 그 각각의 지역구분 형태에 대해 두 변수간의 상관관계가 도출된다면, 우리는 그것들 중 무엇을 두 변수간의 연관성에 대한 최적값으로 인정해야 할 것인가? Openshaw와 Taylor(1979)의 논문 제목이 함축하고 있는 것처럼,<sup>29)</sup> 우리는 백만 가지 이상의 상관계수 앞에서

26) Openshaw, *op. cit.*, p.8.

27) 공간통계학의 최근 발달에 대한 내용은 이 논문의 범위를 넘어서는 것이다. 개요를 파악하는데 다음의 논문들이 도움이 될 것이다. A. Getis, 1999, "Spatial statistics", in P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, eds., *Geographical Information Systems*, vol.1 : *Principles and Technical Issues*, 2nd Edition, New York: John Wiley & Sons, pp.239~251; D. J. Unwin, 1996, "GIS, spatial analysis and spatial statistics", *Progress in Human Geography*, vol.20, no.4, pp.540~551.

28) D. Holt, D. G. Steel, and M. Tranmer, 1996, "Area homogeneity and the modifiable areal unit problem", *Geographical Systems*, vol.3, p.181.

단지 하나를 선택하고 있는 것이다.

MAUP의 발생에 대한 가장 일반적인 설명은 합역이 발생할 때 스케일의 효과와 구획의 효과로 인하여 변수들의 변량과 변수간의 공변량이 변화한다는 것이다. Fotheringham과 Wong(1991) 그리고 Wong(1994)이 설명하고 있는 것처럼, 합역은 일반적으로 변수의 변량을 감소시키는 효과를 가져오는데, 이것은 단위공간이 합쳐지면서 일종의 '均質化 效果(smoothing effect)'가 발생하기 때문이다.<sup>30)</sup> 예를 들어, 합역 전의 이상치들(outliers)은 다른 지역의 값들과 합산되면서 평균값을 향해 수렴되는 경향이 기대되기 때문이다. 변량의 감소는 자연적으로 상관계수나 회귀계수의 값을 증가시키는데, 그것은 상관계수의 계산식에서 분모는 두 변수의 표준편차의 곱으로 계산되기 때문이고, 회귀계수는(단순회귀분석일 경우 특히) 기본적으로 상관계수의 함수로 표현될 수 있기 때문이다.

그러나 Amrhein(1995)이 밝힌 것처럼<sup>31)</sup>, 합역과 분산의 감소에 대해 결정적인 일반화를 하기는 어렵다. 보다 궁극적인 MAUP의 원인은 사물의 공간적 분포의 본질적 특성 그 자체에 있다. 이것은 공간자료의 공간 의존성, 혹은 공간적 자기상관, 혹은 Tobler의 지리학의 '제1법칙(First Law)'<sup>32)</sup>이 공통적으로 의미하는 바이다. 즉, 인접하고 있는 사물들의 속성은 떨어져 있는 사물들의 속성에 비해 그 유사성이 높다는 것이다. 이러한 공리가 MAUP에서 가지는 의미는, 만일 어떤 변수의 공간적 분포가 임의적(random) 이라면, 공간단위가 합역되더라도 그 변수의 공간적 분포는 여전히 임의적일 것이기 때문에 합역의 효과는 발생하지 않을지도 모른다.<sup>33)</sup> 그러나 변수 값의 분포가 공간의존적이고 공간적 자기상관을 보인다면, 합역의 과정은 유사한 값끼

29) Openshaw and Taylor, *op. cit.*

30) A. S. Fotheringham and D. W. S. Wong, 1991, "The modifiable areal unit problem in multivariate statistical analysis", *Environment and Planning A*, vol.23, no.7, pp.1025~1044; D. W. S. Wong, 1996, "Aggregation effects in geo-referenced data", in S. L. Arlinghaus, D. A. Griffith, W. C. Arlinghaus, W. D. Drake, and J. D. Nystuen, eds., *Practical Handbook of Spatial Statistics*, Boca Raton: CRC Press, pp.83~106.

31) C. G. Amrhein, 1995, "Searching for the elusive aggregation effect : evidence from statistical simulations", *Environment and Planning A*, vol.27, no.1, pp.105~119.

32) "모든 사물은 다른 사물들과 연관되어 있다. 그러나 인접한 사물들은 떨어져 있는 사물들보다 더 연관되어 있다." W. R. Tobler, 1970, "A computer movie simulating urban growth in the Detroit region", *Economic Geography*, vol.46, no.2 (Supplement), p.236.

33) G. Arbia, 1989, *Spatial Data Configuration in Statistical Analysis*, Norwell, MA : Kluwer Academic Publishers; Wang, *op. cit.*

리 연합하는 형태를 띠기 때문에 그 효과는 통계치에 반영될 수밖에 없다. 따라서 한 합역양식이 다른 합역양식에 비해 각 공간단위의 동질성을 보다 증대시켰다면, 합역의 효과는 그 합역양식에서 더 크게 드러날 것이다.<sup>34)</sup>

## 2) MAUP의 영향

MAUP의 영향을 확인하기 위해 연구자들은 다양한 모델들에 대해 실험했는데, 공간적 상호작용 모델(Spatial Interaction Model)<sup>35)</sup>, 포아송 회귀분석 모델(Poisson Regression Model)<sup>36)</sup>, 경기 예측 모델 (Economic Forecasting Model)<sup>37)</sup>, 입지-할당 모델(Location-Allocation Model)<sup>38)</sup>, 격리지수 (Segregation Index)<sup>39)</sup>, 그리고 카이제곱(chi-squared) 검증<sup>40)</sup> 등이 있다. 하지만 여기서는 보편적으로 논의되는 상관분석과 회귀분석에 초점을 맞추고자 한다. 연구자들은 주로 시뮬레이션<sup>41)</sup>을 통해 조건에 부합하는 가능한 모든 합역양식을 생성하고 선정된 통계치들을 합역양식간에 비교함으로써 다음의 몇가지 MAUP의 효과에 대한 일반화에 도달했다.

첫째, 분산이 큰 변수가 합역의 효과를 가장 크게 받는다.<sup>42)</sup>

둘째, 합역의 수준이 높아질수록, 즉 보다 적은 수의 단위 공간이 생성될

---

34) Holt, Steel, and Tranmer, *op. cit.*, p.184.

35) S. Openshaw, 1977, "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling", *Transactions of the Institute of British Geographers*, vol.2, pp.459~472; S. H. Putman and S.-H. Chung, 1989, "Effects of spatial system design on spatial interaction models. 1 : The spatial systems definition problem", *Environment and Planning A*, vol.21, pp.27~46.

36) C. G. Amrhein and R. Flowerdew, 1989, "The effect of data aggregation on a Poisson regression model of Canadian migration", in Goodchild and Gopa, *op. cit.*, pp.229~238.

37) J. R. Miller, 1998, "Spatial aggregation and regional economic forecasting", *Annals of Regional Science*, vol.32, no.2, pp.253~266.

38) M. F. Goodchild, 1979, "The aggregation problem in location-allocation", *Geographical Analysis*, vol.11, pp.240~255.

39) D. W. S. Wong, H. Lasus, and R. F. Falk, 1999, "Exploring the variability of segregation index D with scale and zonal systems: an analysis of thirty US cities", *Environment and Planning A*, vol.31, pp.507~522.

40) A. J. Curtis and A. D. MacPherson, 1996, "The zone definition problem in survey research : An empirical example from New York State", *Professional Geographer*, vol.48, no.3, pp.310~320.

41) 특히 Openshaw는 AZP(Automatic Zoning Procedure)라는 알고리즘을 개발하였다. Openshaw, 1984, *op. cit.*; S. Openshaw and S. Albanides, 1999, "Applying geocomputation to the analysis of spatial distributions", in Longley, et al. *op. cit.*, pp.267~282.

42) Amrhein, *op. cit.*



수록, 변수간의 상관관계는 높아진다.<sup>43)</sup>

셋째, 합역의 수준이 높아질수록 회귀분석에서 결정계수의 값은 커진다.<sup>44)</sup>

넷째, 합역과 회귀계수와의 관계는 일반화하기 어렵지만 스케일이나 구획 방식에 따라 상당한 정도의 편차가 관측된다.<sup>45)</sup>

다섯째, 합역의 수준이 높아질수록 공간적 자기상관의 값은 작아진다.<sup>46)</sup> 그리고 공간적 자기상관이 높은 변수가 MAUP에 영향을 덜 받는다.<sup>47)</sup>

## 2. 機能地域의 設定과 MAUP의 影響에 對한 實證分析

### 1) 機能地域의 本質과 MAUP

다음 장에서는 위의 다섯가지 일반화에 대해 미국의 기능지역을 사례로 검증해 보고자 한다. 분석에 들어가기 전에 먼저 기능지역과 MAUP의 관계에 대해 몇가지를 지적하고자 한다. 첫째, 기능지역을 구성하는 것은 기능지역이라는 공간단위를 이용하여 최저 합역수준, 즉 개인 자료상에서의 통계적 과정을 추론하기 위함이 아니다. 일반적으로 공간단위를 연구의 분석단위로 사용하는 이유는 개인자료에의 접근이 불가능한데서 기인한 불가피한 선택인 경우가 많다. 이런 경우, 공간단위간의 관계를 통해 개인간의 관계를 추론하는 것은 '生態學的 誤謬(ecological fallacy)'라는 용어로 잘 알려져 있듯 매우 위험한 것이다.<sup>48)</sup> 엄밀히 말해 생태학적 오류는 MAUP의 한 특수한

43) C. G. Amrhein, 1994, "Testing the use of a hybrid regionalisation scheme for confidential tax-filer data", *Canadian Journal of Regional Science*, vol.17, pp.259~274; Gehlke and Biehl, *op. cit.*; Fotheringham & Wang, *op. cit.*; Openshaw, 1984, *op. cit.*; Openshaw and Taylor, *op. cit.*; Wong, *op. cit.*

44) Amrhein, 1995, *op. cit.*; Wong, *op. cit.*

45) Openshaw, 1978, *op. cit.*; Amrhein, 1995, *op. cit.*; Wong, *op. cit.*; Fotheringham & Wong, *op. cit.*

46) Y. H. Chou, 1995, "Spatial pattern and spatial autocorrelation" in A. U. Frank and W. Kuhn, eds., *Spatial Information Theory: A Theoretical Basis for GIS*, Proceedings of International Conference COSIT 95, Semmering, Austria, Berlin : Springer, pp.365~376; 물론 Chou는 다소 다른 측면에서, map resolution과 공간적 자기상관과의 관계를 논의했지만, map resolution과 Moran's I값이 정적인 로그선형(log-linear)관계가 있다는 그의 결론은 합역의 수준이 높아질수록(즉 resolution이 낮아질수록) 공간적 자기상관은 낮아진다는 의미로 차용될 수 있다.

47) C. G. Amrhein and H. Reynolds, 1997, "Using the Getis statistic to explore aggregation effects in Metropolitan Toronto Census data", *Canadian Geographer*, vol.41, no.2, pp.137~149.

48) 생태학적 오류를 최초로 지적한 것은 Robinson이다. W. S. Robinson, 1950, "Ecological

범주이다. 기능지역의 분석에 이와 같은 일반원칙을 적용하는 것은 옳지 않다. 생태학적 오류에 치우친 MAUP에 대한 이해는 당연히 최소의 공간단위를 사용할 것을 권고하겠지만, 기능지역의 경우는 최소단위인 카운티보다 그것의 합역에 의해 생성된 기능지역이 이론적으로나 실질적으로 훨씬 더 유용한 공간단위이다. 사실 기능지역의 설정은 정교한 이론적 작업의 산물이며 그 자체로 하나의 연구주제이다.

둘째, 공간적 자기상관과 MAUP간의 일반적인 연관성을 찾는 것은 적어도 기능지역이 고려대상이 되는 한 매우 어려운 작업이다. 공간적 자기상관과 기능지역의 형성간의 관계에 대해 두가지 가능성이 동시에 존재한다고 말할 수 있다. 첫째, 기능지역은 기본적으로 카운티 간의 유사성에 기반을 두고 있기 때문에 카운티 수준에서의 공간적 자기상관을 흡수하고 있다고 보아야할 것이다. 따라서 잘 구획된 기능지역은 그렇지 않은 기능지역에 비해 낮은 공간적 자기상관을 보일 것으로 기대할 수 있다(합역을 통한 공간적 자기상관의 감소). 둘째, 기능지역이 다른 지역과 구분되는 것은 특정변수에 대해 부분지역이 보여주는 동질성(sameness)이 아니라, 부분지역간의 기능적인 관련성 때문이다. 예를 들어 한 기능지역 내에서 중심카운티와 주변 카운티는 많은 변수들에 대해 뚜렷한 차이를 보여준다. 따라서 이질적인 두 부분의 결합은 평균적인 값을 산출하게 되고 비슷한 과정을 통해 산출된 인접 지역의 값과의 유사성은 증대될 수 있다. 따라서 잘 구획된 기능지역은 그렇지 않은 기능지역에 비해 높은 공간적 자기 상관을 보일 것으로 기대할 수 있다(합역을 통한 공간적 자기상관의 증대).

## 2) MAUP의 영향에 대한 實證分析

스케일 효과를 살펴보기 위해 County-CZ-LMA와 County-BTA-MTA를 비교하였고, 구획의 효과를 살펴보기 위해 비교적 스케일이 비슷한 PSR-

---

correlations and the behavior of individuals", *American Sociological Review*, vol.15, pp.351~357; MAUP과 생태학적 오류와의 관계에 대해서는 다음의 논문을 참고할 수 있다. N. Wrigley, 1995, "Revisiting the modifiable areal unit problem and the ecological fallacy", in A. D. Cliff, P. R. Gould, and N. J. Thrift, eds., *Diffusing Geography : Essay for Peter Haggett*, Oxford : Blackwell, pp.49~71; N. Wrigley, T. Holt, D. Steel, and M. Tranmer, 1996, "Analysing, modelling, and resolving the ecological fallacy", in P. Longley and M. Batty, eds., *Spatial Analysis : Modelling in a GIS Environment*, Cambridge: GeoInformation International, pp.23~40.

MEA-MTA를 비교하였다. 분석에 사용된 자료는 1990 Census Summary Tape File 3A로부터 카운티별로 자료를 추출하고 각 기능지역별 합역양식에 따라 카운티를 합산하였다.<sup>49)</sup> 실증연구에 사용된 변수는 모두 네 가지로 POVFAM(빈곤수준 이하 가정의 비율), HIGSCHL(고졸이상 학력 소비자 비율), BLACK(흑인 인구 비율), 그리고 FEHEAD(여성 단독가정 비율)이다.

(1) 平均과 分散에의 영향

표 2. 각 변수별 기능지역별 평균과 분산

분류	기능지역	County	CZ	BTA	LMA	BEA	PSR	MEA	MTA
	단위공간 수	3,141	741	493	394	172	55	52	51
평균	POVFAM	13.10	13.04	11.84	11.95	11.82	10.89	10.71	10.72
	HIGSCHL	85.48	86.37	87.81	87.51	88.33	89.35	89.51	89.47
	BLACK	8.57	7.59	8.37	9.69	10.28	9.64	10.39	10.45
	FEHEAD	12.78	13.15	14.01	14.29	14.71	14.76	15.13	15.13
분산계수*	POVFAM	0.63	0.48	0.45	0.43	0.39	0.34	0.32	0.32
	HIGSCHL	0.09	0.08	0.07	0.06	0.05	0.04	0.04	0.04
	BLACK	1.67	1.58	1.35	1.23	1.14	0.94	0.82	0.82
	FEHEAD	0.42	0.35	0.26	0.26	0.24	0.18	0.15	0.15

\* 분산계수(Coefficient of Variance) = 표준편차 / 평균

표 2는 각 변수별로 각 기능지역의 평균과 분산을 보여주고 있다. 표준편차에 대한 평균의 영향을 배제함으로써 기능지역별 비교를 용이하게 하기 위해 표준편차를 평균으로 나눈 分散係數(CV; Coefficient of Variance)를 사용했다. 평균값은 합역의 수준이 높아짐에 따라 증가하는 추세를 보인다. 그러나 모든 변수에 대해 그런 것은 아니다. POVFAM의 평균값은 합역의 수준이 높아짐에 따라 오히려 감소하는 추세를 보이고 있다. 그러나 여기서 보다 중요한 것은 분산의 추이다. 모든 변수에 대해, 분산계수는 합역의 수준이 높아짐에 따라 감소하고 있다. 이것은 기존 연구가 보여준 연구결과와 동일하다. 따라서 변수간 상관관계가 합역의 수준이 높아짐에 따라 증가할 것이 기대된다.

49) 자료는 다음의 웹사이트로부터 다운로드할 수 있다. <http://venus.census.gov/cdrom/lookup>.

(2) 相關係數에의 영향

표 3. 기능지역별 POVFAM과 여타변수간의 상관관계

분류	기능지역	County	CZ	BTA	LMA	BEA	PSR	MEA	MTA
	단위공간 수	3,141	741	493	394	172	55	52	51
상관계수	HIGSCHL	-0.579	-0.704	-0.698	-0.719	-0.738	-0.773	-0.754	-0.762
	BLACK	0.373	0.303	0.354	0.394	0.402	0.347	0.352	0.337
	FEHEAD	0.421	0.489	0.527	0.514	0.547	0.424	0.523	0.513

(모든 상관계수 Pearson's correlation coefficient이며 모두 99% 신뢰수준에서 유의하다.)

표 3은 각 기능지역별로 POVFAM과 나머지 변수간의 상관계수를 계산한 것이다. 기존의 연구에 따르면 합역의 수준이 높아질수록 상관계수의 값은 증가한다. 우선 변수별로 살펴보면, HIGSCHL만 일반화된 경향을 보여줄 뿐 나머지 두 변수는 100개 이하의 단위공간수를 가지는 PSR, MEA, MTA에서 상관계수는 오히려 크게 감소한다. 표 2에서 모든 변수에 대해 변량이 합역의 수준이 증가함에 따라 감소함을 확인했음에도 불구하고 이와같은 결과를 보이는 것은 상관계수의 분자 부분을 구성하고 있는 공변량(covariance)이 변수에 따라 달라 질 수 있기 때문이다.

계층적 합역에 의해 형성된 카운티-CZ-LMA와 카운티-BTA-MTA를 비교해 보면, 재미있는 결과를 얻을 수 있다. 카운티에서 LMA로 갈수록 모든 변수들에 대해 상관계수는 증가하고 있다(CZ의 BLACK은 예외). 하지만 카운티-BTA-MTA의 경우는 조금 다르다. 카운티에서 BTA로 갈수록 모든 상관계수는 증가하지만, BTA에서 MTA로 갈수록 모든 상관계수는 감소한다. 이것은 바로 변수별로 합역의 과정에 따른 공변량의 변화가 다양함을 의미한다. 따라서 MAUP의 효과는 '변수특수적(variable-specific)'이다. 이와 관련하여 하나 더 첨가할 수 있는 것은 BLACK의 값이 기능지역별로 가장 변동이 심하다는 것이다. 이것은 자체 분산이 큰 변수가 MAUP에 가장 크게 영향을 받는다는 일반적인 연구결과와 조응하는 것이다.

구획 효과를 살펴보기 위해 PSR, MEA, MTA를 비교해 볼 수 있는데, PSR의 FEHEAD에서 보이는 비교적 낮은 상관계수(0.424)가 특징적이다. 이것은 카운티에 대한 상관계수와 거의 동일한 값인데, MAUP은 스케일 효과뿐만 아니라 구획 효과 역시 드러남을 예증하고 있다.

(3) 回歸係數와 決定係數에의 영향

표 4. 기능지역별 회귀계수와 결정계수

분류	기능지역	County	CZ	BTA	LMA	BEA	PSR	MEA	MTA
	단위공간 수	3,141	741	493	394	172	55	52	51
회귀계수	HIGSCHL	-0.572***	-0.599***	-0.560***	-0.584***	-0.619***	-0.666***	-0.592***	-0.602***
	BLACK	0.035***	-0.083***	-0.082***	-0.066***	-0.078***	-0.002	-0.030	-0.031
	FEHEAD	0.401***	0.673***	0.771***	0.704***	0.710**	0.254	0.452*	0.427*
	intercept	56.536***	56.546***	50.876***	53.673***	56.872***	66.628***	57.201***	58.457***
결정계수		0.427***	0.637***	0.635***	0.659***	0.679***	0.625***	0.619***	0.623***

(\*\*\* 99% 신뢰수준, \*\* 95% 신뢰수준, \* 90% 신뢰수준에서 유의함.)

다중회귀분석에 대한 MAUP의 영향을 살펴보기 위해 POV FAM을 종속변수로, 나머지 세 변수를 독립변수로 하는 다중회귀분석을 실시하였다. 도출된 각 독립변수별 회귀계수와 그것의 유의성, 그리고 결정계수를 기능지역별로 비교하였다. 우선 기능지역별로 보면, CZ, BTA, LMA, BEA는 거의 유사한 결과를 보여주고 있는데 반해, 이것들보다 단위공간의 수가 훨씬 적은 PSR, MEA, MTA과는 상당히 다른 결과를 보여주고 있다. 특히 가장 주목할 부분은 카운티의 BLACK의 회귀계수의 부호이다. 다른 모든 기능지역에 대해 BLACK은 부(-)적인 값을 가졌는데, 카운티에 대해서는 정(+)적인 값을 가진다. 표 3의 상관관계에서 유추해보면 BLACK에 대한 회귀계수는 양수인 것이 옳다. 그렇다면 여타 기능지역들이 보여주고 있는(적어도 CZ에서부터 BEA까지) 부(-)적이면서 동시에 유의한 회귀계수의 값은 무엇을 의미하는가? 아마도 BLACK과 FEHEAD간의 높은 상관관계에 따른 共線性(multicollinearity)에 기인했을 가능성이 높지만, 유의한 회귀계수를 전제할 때 또 다른 해석의 가능성도 완전히 배제할 수 없다. BLACK이 유의하지 않게 산출된 PSR, MEA, MTA의 경우는, 기대와 다른 부호가 나타나는 이유를 공선성에서 찾아야만 할 것이다. 이 회귀분석의 결과에서도 BLACK이 기능지역간에서 가장 큰 편차를 보이는데, 상관관계분석에서도 언급되었듯이 자체 편차가 큰 변수가 MAUP에 가장 큰 영향을 받는다는 일반 원칙을 다시 한번 증명하고 있다.

일반적인 원칙에 따르면 결정계수의 크기는 함역의 수준이 상승할수록 중

가한다. 그러나 표 4를 보면, 그러한 일반화는 과도해 보인다. 이것은 계층적 합역을 통해 생성된 BTA와 MTA를 비교해보면 분명해지는데, 훨씬 적은 수의 단위공간으로 구성된 MTA의 결정계수값이 오히려 낮다.

구획의 효과를 살펴보기 위해 PSR, MEA, MTA의 회귀계수와 결정계수를 비교해 보면, 상관관계분석에서와 마찬가지로 PSR이 나머지 두 기능지역에 비해 특이한 값들을 산출하고 있다. 이것은 다시 구획의 효과를 증명하는 것이기도 하지만, 연구대상이 기능지역이라는 측면에서 보면, PSR이 가지고 있는 구획 자체의 문제에 대해서도 살펴보아야 할 것이다. 실질적으로 PSR은 거의 주의 구획과 흡사하다.

#### (4) 空間的 自己相關과 MAUP

표 5. 기능지역별 공간적 자기상관 지수

분 류	기능지역	County	CZ	BTA	LMA	BEA	PSR	MEA	MTA
	단위공간 수	3,141	741	493	394	172	55	52	51
Moran's I	POVFAM	0.437	0.576	0.535	0.588	0.498	0.386	0.338	0.316
	HIGSCHL	0.683	0.594	0.511	0.538	0.426	0.406	0.458	0.408
	BLACK	0.809	0.857	0.816	0.830	0.751	0.564	0.486	0.483
	FEHEAD	0.560	0.625	0.619	0.670	0.596	0.489	0.447	0.427

(software: S-Plus for ArcView GIS, ESRI & MathSoft)

공간적 자기상관과 MAUP의 관계를 파악하기 위해, 모든 변수에 대해 각 기능지역별로 Moran의 I 지수(Moran's I index)를 계산하였다. I 지수 값은 일반적으로 -1에서 1사이의 값을 갖는데, I가 양수의 값을 가지면서 1에 가까워지면 가까워질수록 유사한 값들이 공간적으로 집중되어 있음을 의미한다.<sup>50)</sup> I를 계산하기 위해 필요한 '空間加重行列(spatial weighting matrix)'은

50) 공간적 자기상관과 그와 연관된 최근의 공간계량경제학(spatial econometrics) 혹은 空間的 自己回歸모델(spatial autoregressive model)의 발달은 이 논문의 범위를 넘어서는 것이다. 아마 이 분야는 1990년대 공간통계학의 최대 성과임과 동시에 가장 빨리 성장하고 있는 분야일 것이다. 공간적 자기상관에 대해 다음의 책이 가장 좋은 입문서 역할을 할 것이다. J. Odland, 1988, *Spatial Autocorrelation*, Newbury Park: SAGE Publications. 공간적 자기회귀분석의 개관을 위해서 다음의 논문을 참고할 수 있다. L. Anselin and A. K. Bera, 1998, "Spatial dependence in linear regression models with an introduction to spatial econometrics", In A. Ullah, ad D. Giles, eds., *Handbook of Applied Economic Statistics*,

경계를 접하고 있는 공간단위의 쌍에 대해 1의 값을, 그렇지 않은 쌍에 대해 0의 값을 부여하는 단순이항가중(simple binary weighting) 방식을 따랐다.<sup>51)</sup> 표 5는 그 결과를 요약하고 있다. 우선, 전반적으로 특정한 패턴을 찾기가 어렵다는 점을 지적할 필요가 있다. 카운티에서 BEA로 갈수록 오르내림을 반복하고 있다. 따라서 최소한 기능지역이 고려되고 있는 한, 합역의 수준과 공간적 자기상관 지수와의 관계에 대한 특정한 일반화를 도출하는 것은 어려운 것 같다. 특히 계층적 합역에 의해 구성된 CZ와 LMA를 비교하면 확연해지는데, HIGSCHL과 BLACK의 값들은 줄어들었지만, POVFAM과 FEHEAD의 값은 증가했다. 그러나, 또 다른 계층적 합역의 예인 BTA와 MTA를 보면, 확연히 MTA의 값이 BTA의 값에 비해 낮다.

또 하나 지적할 수 있는 것은, 100개 이상의 공간단위를 가진 기능지역과 100개 이하의 공간단위를 가진 기능지역간에 일정한 불연속성이 발견된다는 점이다. HIGSCHL을 제외한 나머지 모든 변수에서 공간적 자기상관지수의 값은 BEA에서 PSR로 넘어가면서 현저히 낮아진다. 이것은 집중과 분산을 야기하는 공간과정의 스케일이 미국 전체지역의 대략 1/100정도의 규모로 발생하고 있다는 점을 함축하고 있다. 즉 그 스케일을 넘어 합역되었을 때, 단위공간간 이질성이 급격히 증가하면서 공간적 자기상관지수의 값이 현저히 감소하는 것이다.<sup>52)</sup>

이론적인 의미에서 공간적 자기상관은 명백하게 MAUP과 밀접한 관계에 있다. 그러나 위에서 살펴본 바와 같이 실제자료를 이용한 결과에서 특징적인 패턴을 발견하기 어려웠다. 이것은 여기서 사용되고 있는 예들이 정교하게 만들어진 기능지역들이기 때문에 시뮬레이션을 통해 확률적으로 도출된 단순한 합역체(spatial aggregates)와는 본질적으로 다르기 때문이다. 따라서 MAUP과 공간적 자기상관과의 관계에 대한 일반화를 기능지역의 예에 직접적으로 적용하기는 어려울 것이다.

---

New York : Marcel Dekker, pp.237~289.

- 51) 다양한 공간가중행렬 구성방식에 대해서는 다음의 논문을 참고할 수 있다. D. A. Griffith, 1996, "Some guidelines for specifying the geographic weights matrix contained in spatial statistical models", in Arlinghaus, *et al.*, *op. cit.*, pp.65~82.
- 52) 이 부분은 다른 공간통계학적 방법을 통해 보다 심도 있게 분석될 수 있다. 가장 일반적으로 이용되는 방법은 공간가중행렬을 순차적으로 증가시키면서 산출된 일련의 공간적 자기상관 지수를 이용하는 '空間 遲滯 모델(spatial lag model)'이며 그와 연관된 '공간적 correlogram' 분석일 것이다. 계략적인 이해를 위해 Arbia, *op. cit.*와 Odland, *op. cit.*를 참고할 수 있다.

#### IV. 結 論

이 연구에서는 미국에서 다양하게 규정되고 있는 기능지역의 현황을 살펴 보고, 그것을 MAUP에 대한 기존 연구의 결과들에 적용해 봄으로써 공간정보를 활용하고 분석하는데 있어 필수적으로 고려되어야 할 점들을 점검해 보았다. 다양한 공간단위로 주어지는 정보를 연구주제에 맞게 재조정하는 문제는 공간분석에서 필수적이다. 특히 미국의 경우, 전 미국을 대상으로 하는 지역분석에서는 카운티를 기반으로 하는 기능지역을 구축하는 것이 가장 이상적으로 보인다. 다양한 기능지역에 대한 설정이 이루어져 온 것도 바로 이런 이유에서이다. 카운티 몇 개를 묶어서 하나의 기능지역을 형성하는 과정은, 합역이라고 하는 일반적 과정의 특수한 한 형태이며, 합역에 필연적으로 수반되는 MAUP의 효과로부터 자유로울 수 없다.

7개의 기능지역과 기본단위인 카운티를 대상으로 한 실증연구의 결과는 기존연구의 결과에 조용하기도 하고 그렇지 않기도 했다. MAUP의 분산과 상관계수에의 영향에 대해서는 전반적으로 일치되는 결과를 관찰할 수 있었다. 즉, 합역의 수준이 상승할수록 변수들의 분산은 줄어들고 상관계수의 값은 증가한다. 그러나 회귀계수, 결정계수, 공간적 자기상관 지수에 대한 MAUP의 영향에 대한 결과는 기존의 연구에서 일반화된 원칙에 조용하지 않았다. 우선 변수에 대해 살펴보면, MAUP의 효과는 선택된 변수에 따라 큰 편차를 보인다는 결론을 내릴 수 있다. 즉, 어떤 변수들은 MAUP의 영향에 대한 기존 연구의 결과에 조용하지만, 다른 변수들은 그렇지 않다. 기능지역별로 보면, 100개 이상의 단위공간을 가진 CZ, BTA, LMA, BEA와 100개 이하를 가진 PSR, MEA, MTA이 각각 차별적인 특성을 보여주고 있다. 특히 자기상관 계수의 결과를 보면, 대부분의 변수에 대해 BEA와 PSR의 차이가 BEA와 CZ의 차이보다 훨씬 크다. 이것은, 최소한 본 연구에 포함된 변수에 대해서만 말한다면, 공간적 분포를 일으키는 공간과정이, 대략 미 전 지역이 100개 정도의 지역으로 나누어졌을 때, 그 한 단위의 평균규모에서 발생하고 있다는 것을 함축하고 있다. 즉 그 이상의 규모를 가진 기능지역으로 카운티가 합역되었을 때, 인접 공간단위와의 유사성은 급격히 줄어드는 것이다.

본 연구의 결과는 다양한 시사점을 제공해 주고 있다. 본 연구 결과의 몇몇은 기존 연구들에 의해 성립된 일반화에 반하고 있다. 그러나 그것들이 자



동적으로 그 일반화를 반증하기 위해 사용될 수는 없다. 왜냐하면, 기존 연구의 일반화는 도출 가능한 모든 합역양식에 대한 시뮬레이션을 통해 확률적으로 성립되었기 때문이다. 그러나 기능지역이 시뮬레이션을 통해 도출될 수 있는 모든 가능한 합역양식들 중의 단순한 하나가 아니라는 사실 또한 명백하다. BEA를 예를 들면, 하나의 BEA가 3,141개의 카운티를 172개의 지역으로 합역하는 셀 수 없이 많은(연접해 있는 카운티들만이 합역될 수 있다는 조건하에서도 무수히 많을 것이다.) 양식들 중의 하나임에는 분명하지만, BEA를 제외한 나머지 대부분의 경우들은 기본적으로 ‘기능지역’이 아니며 단순히 카운티의 연결일 뿐이다.

여기서 본 논자는 기능지역과 MAUP의 관계라는 측면에서 새로운 문제제기가 필요하다는 주장을 하고자 한다. 중요한 것은 합역양식에 따라 통계치가 변화한다는 것을 지적하는 것 자체가 아니라, 다양한 통계치 중에서 보다 유의미한 통계치가 무엇이며 그 통계치를 도출한 합역양식이 무엇인지를 평가할 기준을 제시하는 것이다. 다시 말해서, MAUP에 대한 연구는 경쟁하는 기능지역의 합역양식을 평가하고 보다 이론적으로나 공간통계학적으로 의존 가능한 합역양식을 선택하는데 도움을 줄 수 있어야만 한다. 이러한 과제를 해결하는 과정에서 합역양식과 공간적 자기상관 계수의 관계에 대한 모델의 개발은 핵심적인 위치를 차지할 것이다.

# The Delineation of Function Regions and Modifiable Areal Unit Problem

Sang-Il Lee\*

Recently, MAUP has been intensively addressed by researchers, some of whom sought to identify some effects of MAUP on statistical results, and others tried to edify the problems. Obviously, soaring attention paid to MAUP is associated with recent findings and technical advances in spatial statistics, which basically aims at spatializing the general statistics.

The main purposes of the paper were: (1) to review and compare major functional regions generated by aggregating counties in the US especially in terms of different schemes of spatial aggregation; (2) to calibrate some effects of MAUP across the functional regions by reference to general findings derived mainly from simulation approaches.

The major findings of the research can be summarized by several points: (1) the thesis of increase in variances and correlation coefficients in accordance with upward spatial aggregation generating less and less spatial units was evidenced by our data sets; (2) the thesis of increase in regression coefficients and coefficients of determination was hardly supported by the data sets; MAUP seemed to be variable-specific; (3) the relationships between Moran's I and levels of spatial aggregation did not appear generalizable.

Some explanations on the discrepancies between generally accepted findings about the effects of MAUP and those from the real data sets

---

\* Ph. D. student, Department of Geography, The Ohio State University

used for the research might be given by the fact that a functional region is not just one instance of enormous number of spatial aggregates possibly derived from base units. Rather, it is a thoughtfully designed spatial unit which could provide researchers with a much more reliable and sounder basis than the base unit could do. Moreover, in order for the research on MAUP to further itself, it should be able to help evaluate dependabilities among different regionalization schemes for the functional region.