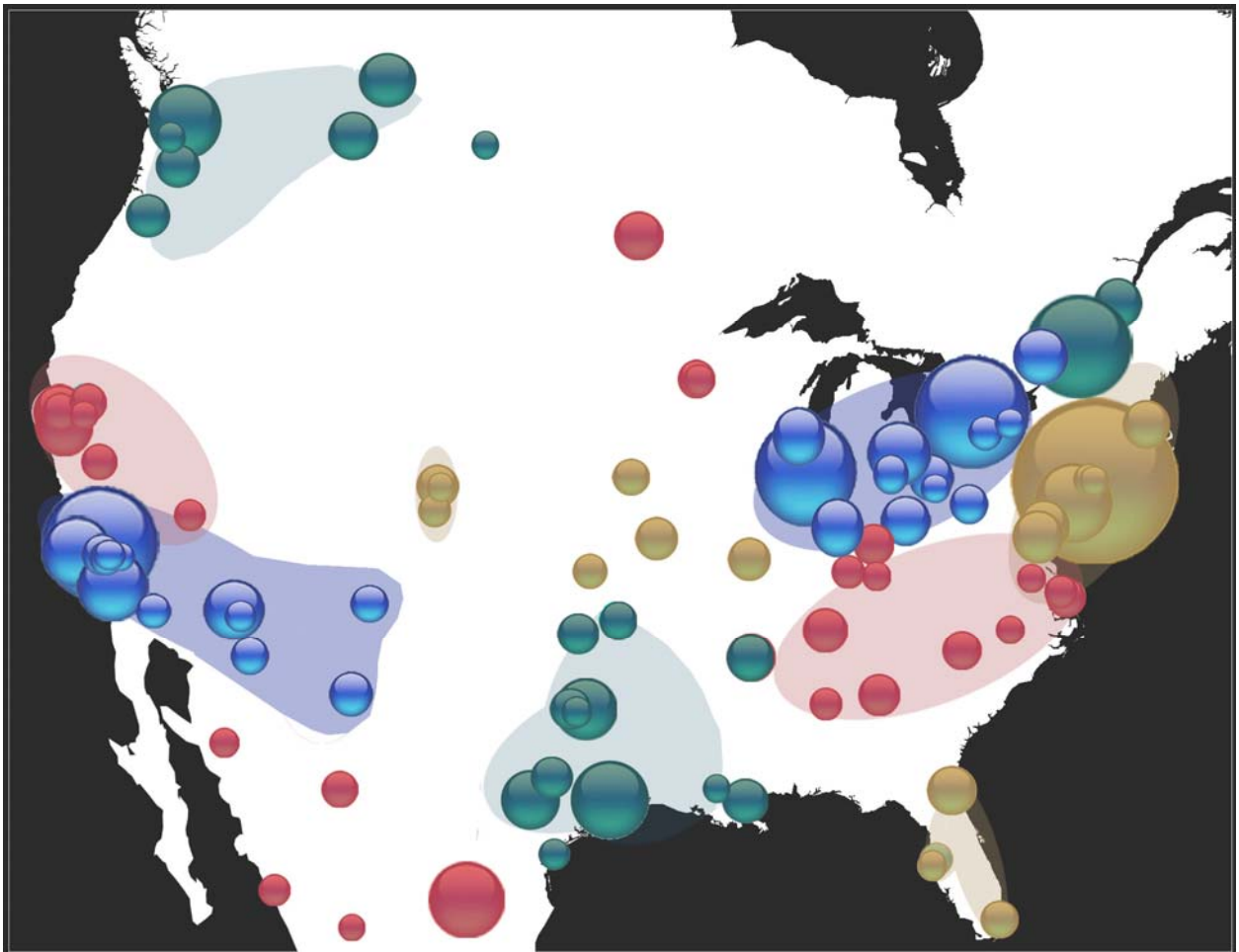


Methods for Planning the Great Lakes MegaRegion

*Urban and Regional Planning Program
University of Michigan, Ann Arbor
April 2006*



*Methods for Planning the Great Lakes
MegaRegion*

Elizabeth Delgado
David Epstein
Yoohyung Joo
Raju Mann
Sarah Moon
Cheryl Raleigh
Erin Rhodes
Daniel Rutzick

*Urban and Regional Planning Program
University of Michigan, Ann Arbor
April 2006*

Table of Contents

	Page
List of Tables	3
List of Figures	3
Introduction	5
Defining the Great Lakes MegaRegion	7
Commute Shed Analysis	15
Commute Shed Analysis - MATLAB Scripts	25
Determining Sprawl	35
Calculating Regional Domestic Product	39
Measuring Residential Economic Vulnerability	43
Calculating Population Growth to 2050	47
Scenarios: Land Use Change	49
Scenarios: Land Use Change - MATLAB Scripts (Sprawl)	61
Scenarios: Land Use Change – MATLAB Scripts (Compact Growth)	71

Source for front cover: US Census, 1990.

List of Tables The Great Lakes MegaRegion

<i>Determining Sprawl</i>	Page
Table 1: Metropolitan Area and Sprawl Score	36
Table 2: Loss in Population Density and Growth in Urbanized Land	37
Table 3: Square Miles of Sprawl	37

List of Figures The Great Lakes MegaRegion

<i>Defining the Great Lakes MegaRegion</i>	Page
Figure 1a: Regional Concentration of Auto Manufacturing Plants	7
Figure 1b: Concentration of Steel Manufacturing Plants	7
Figure 2: Final Great Lakes MegaRegion Map	9
Figure 3: The Great Lakes Watershed	10
Figure 4: Corridor and Galactic Map of the Midwest	11
Figure 5: Morrison's Midwest Mega-Region	12
Figure 6: Doxiadis: The Great Lakes Megalopolis	13
Figure 7: Brookings Institution: the Great Lakes Region	14
<i>Commute Sheds Analysis</i>	
Figure 8: A System for Computing and Displaying Commute Sheds	19
Figure 9: Connections between the MegaRegion and the rest of the US Via Worker Commutes	20
Figure 10: Nearly Everyone Who Works within the MegaRegion also Lives within the MegaRegion	20
Figure 11: Several MSAs within the MegaRegion Appear Joined via Commute Flows	21
Figure 12: Commute Flows into the MegaRegion are concentrated at Two Locations	22
Figures 13 & 14: The Deception of Symbology	24
<i>Measuring Residential Economic Vulnerability</i>	
Figure 15: Metropolitan Residential Economic Vulnerability, Milwaukee-Waukesha PMSA	45
Figure 16: Metropolitan Residential Economic Vulnerability, Indianapolis MSA	45
Figure 17: Metropolitan Residential Economic Vulnerability, Detroit PMSA	46

Calculating Population Growth to 2050

Figure 18:	Great Lakes MegaRegion Population Growth, 1975-2050	47
------------	---	----

Scenarios: Land Use Change

Figure 19:	Future Land Consumption Based on Historic Trends	56
Figure 20:	Future Land Consumption Adjusted for Growth Control Policies	57
Figure 21:	Sprawl Encroaches on Sensitive Areas	58
Figure 22:	Compact Development Reduces Threat to Outlying Ecosystems	59

Introduction

This guide provides information on techniques useful for defining and analyzing MegaRegions. Documenting and sharing these methods has many benefits, including:

- 1) Making comparisons possible across MegaRegions;
- 2) Reducing startup time for new teams beginning MegaRegion planning projects;
- 3) Helping to increase professional rigor behind the MegaRegion concept in the United States, which may lend credibility to policy recommendations;
- 4) Retaining knowledge too technical or detailed for the plans themselves, including the rationale behind decisions and the evolution of ideas;
- 5) Facilitating communication between current students and future students who work on MegaRegion plans by providing another written record;
- 6) Allowing an outlet for constructive critique—including self-critique—that is helpful for advancing MegaRegion planning but not possible within the confines of a final plan that must be clear and concise;
- 7) Providing the basis for eventually standardizing some indicators across MegaRegions.

This document is a Word file where future and current MegaRegion planning teams should feel free to use, critique, and add to the techniques in this guide for the use of others working on the America 2050 project.

Defining the Great Lakes MegaRegion

The first step in planning for the Great Lakes MegaRegion was defining the boundaries of the region. The central question of what defines the Great Lakes MegaRegion was a source of much debate. Hypothetically, different groups working on a plan for the Great Lakes MegaRegion could easily arrive at varying boundary definitions. Key criteria were identified for defining the Great Lakes MegaRegion boundary. These criteria were compiled into a six-step process:

Step 1:

We made a decision to form a Great Lakes MegaRegion boundary using the county as the smallest unit for data gathering. Because Census-defined Metropolitan Statistical Areas (MSA) are clusters of counties, we decided not to divide MSAs. Our initial focus was on the inclusion of the largest metropolitan areas.

Step 2:

During the nineteenth and twentieth centuries, metropolitan areas in states bordering the Great Lakes were connected by shared industrial activity (see Figures 1a & 1b). Coal from the Upper Peninsula of Michigan, steel from Pittsburgh, glass from Toledo, and rubber from Akron arrived in Detroit assembly plants to produce cars.¹ This shared industrial economic base created trade flows and economic connections. This common history created assets and challenges that unite the region today.

Figure 1a. Regional Concentration of Auto Manufacturing

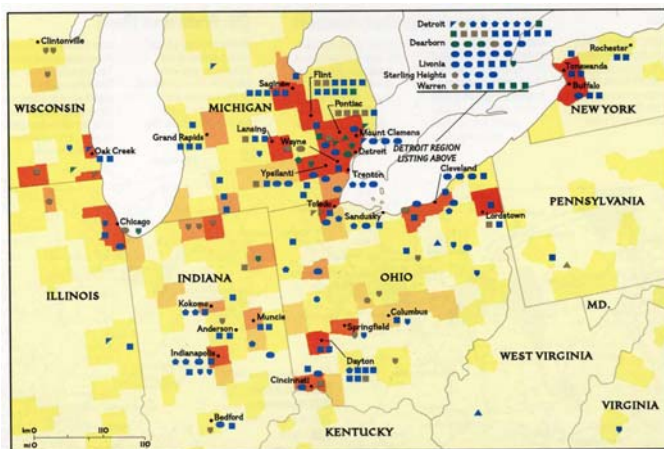


Figure 1b. Concentration of Steel



Source: Historical Atlas of the United States - Centennial Edition. Washington DC: National Geographic Society, 1988.

¹ Jon Teaford, *Cities of the Heartland: The Rise and Fall of the Industrial Midwest*. Bloomington: Indiana University Press, 1993

Defining the MegaRegion
The Great Lakes MegaRegion

A criterion for step two was that the metropolitan areas selected are located in states bordering the Great Lakes, have substantial population size, and share common industrial histories. Based on this criterion, the major metropolitan areas selected were Chicago, Cincinnati, Cleveland, Columbus, Dayton, Detroit, Indianapolis, Milwaukee, Pittsburgh and Toledo.

Step 3:

For this step, counties selected were adjacent to the major metropolitan regions defined in step two and experienced greater than the 8% average population growth rate for the states bordering the Great Lakes.² The growth rate was calculated for 1990 through 2000. The inclusion of these counties within the MegaRegion addresses issues of land consumption and environmental degradation occurring on the fringes of the major metropolitan areas. In the next 10 to 20 years these counties may also become part of a Census defined Metropolitan Statistical Area.

Step 4:

The east and west boundaries of the MegaRegion were formed by the previous two steps. For step four, those counties bordering one of the Great Lakes between the established east and west boundaries were then included. The Great Lakes are one of the region's most important assets. These counties are common recreation destinations for many residents of the cities and suburbs of the region and are therefore connected as a vacation and leisure destination.

Step 5:

Counties surrounded by counties determined to be within the MegaRegion were included to avoid islands.

Step 6:

For the purpose of creating a smooth MegaRegion boundary, we included those counties surrounded on two or more sides by MegaRegion counties inside the Great Lakes MegaRegion's boundary.

² The states of New York and Minnesota were not included in the calculation of the average growth rate because these states are not included in the region.

These six steps lead to the creation of the map below:

Figure 2. Final Great Lakes MegaRegion Map



Source: ESRI GIS Library, Census 2000 Population Data

Exclusion of Major Metropolitan Area

The Minnesota twin cities of Minneapolis and St. Paul were not included within the boundary for this plan. These two cities' metropolitan areas are not tied to the similar industrial pasts of other regional metropolitan areas and do not face similar current challenges.

The Buffalo, New York, metropolitan area shares much of the same industrial past as the large metropolitan areas of the Great Lakes MegaRegion. However, Buffalo appeared directly connected to the metropolitan areas of the Northeast MegaRegion. For this reason, the Buffalo metropolitan area was not included within the MegaRegion boundary.

Despite their proximity to the Great Lakes, Toronto and other Canadian metropolitan areas were also excluded from the MegaRegion. The primary reason for this was the inconsistency in data between the United States and Canada, as well as a lack of time allotted for the project. Toronto and other major Canadian metropolitan areas near the Great Lakes should be considered when expanding this plan's boundary for the MegaRegion.

Other Boundaries Considered

This effort to define the Great Lakes MegaRegion considered several others. These included:

1. Great Lakes Watershed. This definition, based upon natural boundaries, covered all the areas immediately surrounding the Great Lakes themselves but left out much of the land base, such as Indiana and Ohio.

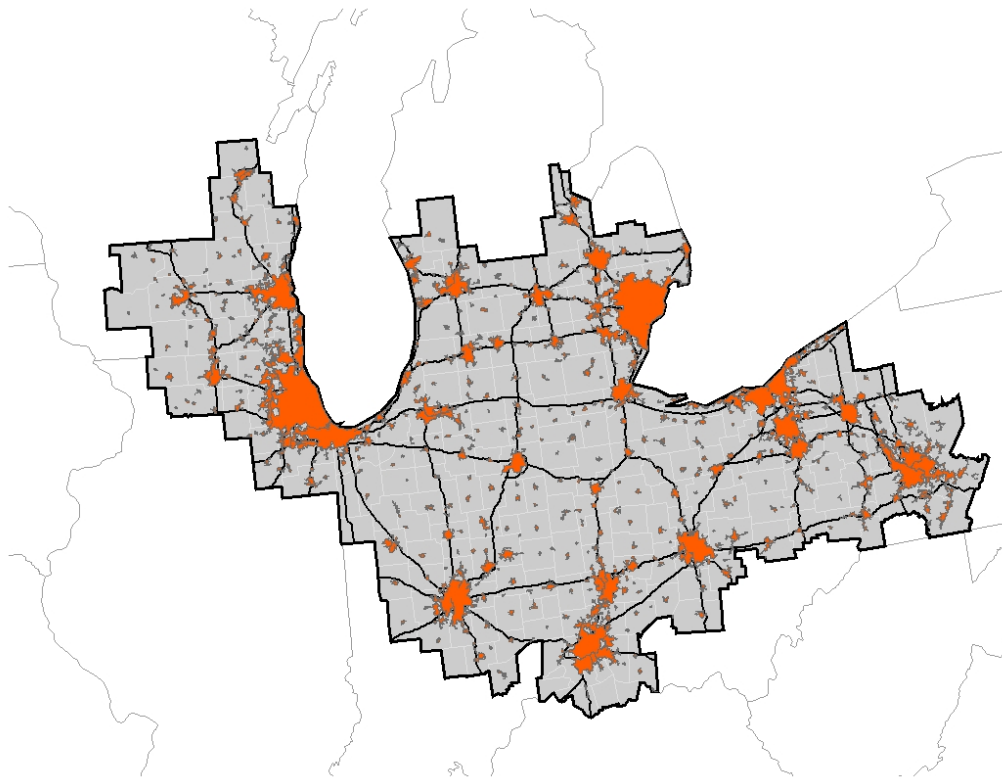
Figure 3. The Great Lakes Watershed



Source: The Great Lakes Information Network. <http://www.great-lakes.net>. Accessed on April 24, 2006

2. Robert Lang & Dawn Dhavale, Virginia Polytechnic Institute and State University, the Great Lakes megapolitan area. Lang and Dhavale defined “megapolitan” areas, integrated networks of metro- and micropolitan areas, as:
 - Derived from contiguous metropolitan and micropolitan areas
 - Constituting an organic cultural region with a distinct history and identity
 - Occupying a roughly similar physical environment

Figure 4. Corridor and Galactic Map of the Midwest



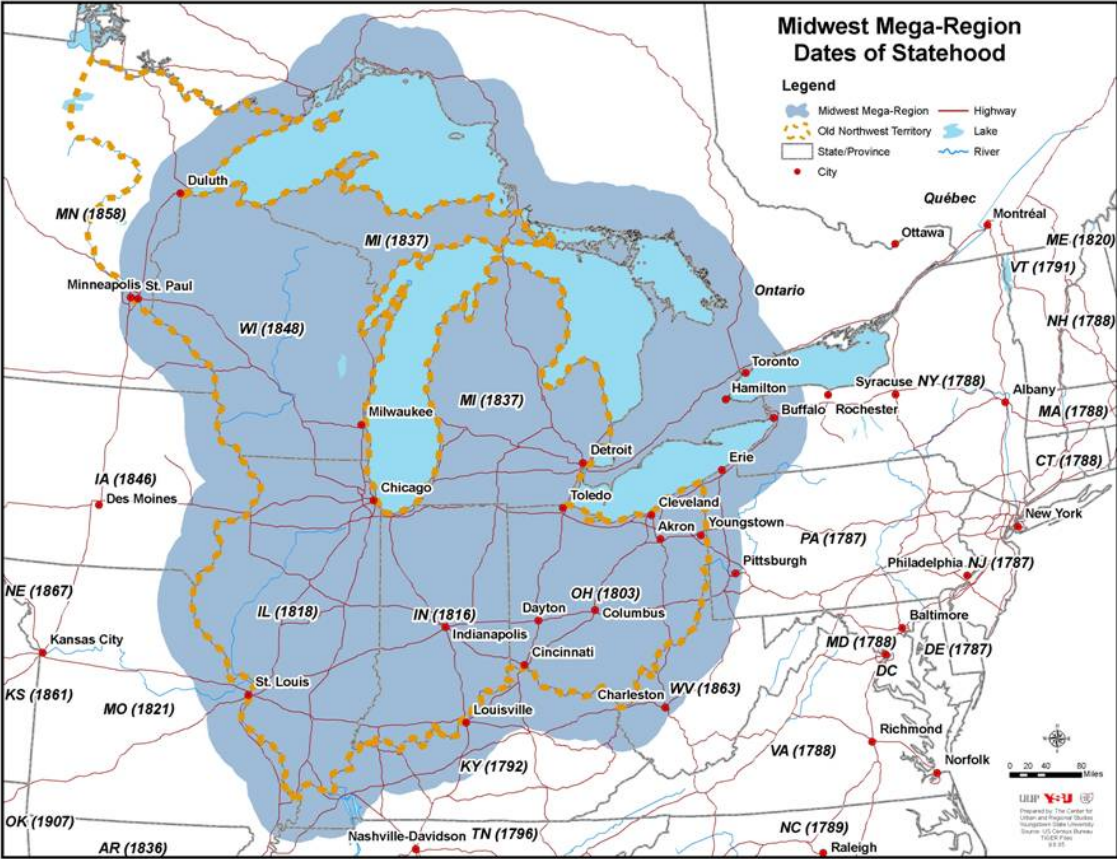
Source: Robert Lang and Dawn Dhavale. “Beyond Megalopolis: Exploring America’s New ‘Megapolitan’ Geography.” Census Report 05:01. Alexandria, VA: Metropolitan Institute at Virginia Tech, 2005.

Figure 3 illustrates Lang and Dhavale’s “corridors” and “galactic” areas. Orange areas represent mega- and micro-politan centers. The big beads (in orange) or galactic areas refer to megapolitan or individual metropolitan areas such as Chicago while the smaller beads indicate micropolitan centers.

Defining the MegaRegion
The Great Lakes MegaRegion

3. Hunter Morrison, Youngstown State University, the Midwest Mega-Region. Morrison’s MegaRegion is defined by environmental features, geography, history, and functional relationships dating from the period of extensive use of the lakes and rivers for navigation. Centered around Chicago, it encompasses the Great Lakes, stretching west to Minneapolis, south to St. Louis, east to Pittsburgh and Buffalo, and north to Nipigon, Ontario (Canada).

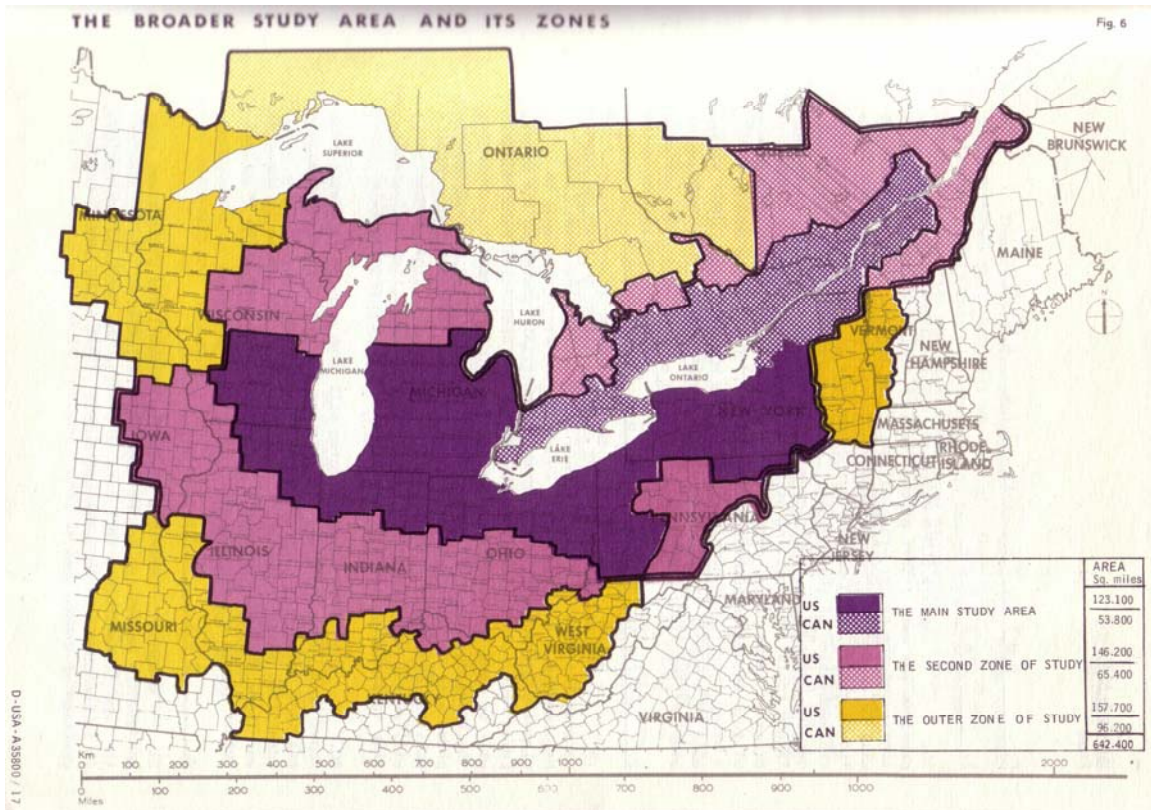
Figure 5. Morrison’s Midwest Mega-Region



Source: Hunter Morrison. PowerPoint Presentation, “Great Lakes Regional Economic Initiative: Defining the Midwest Mega-Region Discussion.” Chicago Illinois, 8 Sept 2005.

4. Doxiadis Plan. The Doxiadis region has a central urbanized area surrounded by secondary and tertiary zones with less connection.

Figure 6. Doxiadis: The Great Lakes Megalopolis



Source: Doxiadis Associates. The Developing Great Lakes Megalopolis Research Project--An Introduction to the Project. Doxiadis Associates. April 1969. Page 9.

- 5. John Austin, Brookings Institution- The Great Lakes Economic Initiative. The Brookings Institution’s work on the Great Lakes economy focuses on five states, parts of four others, and part of Ontario. It is an on-going project that seeks to improve the economic vitality of the Great Lakes region.

Figure 7. Brookings Institution: the Great Lakes Region



Source: John Austin. PowerPoint Presentation “Great Lakes Regional Economic Initiative: Defining the Midwest Mega-Region Discussion.” Chicago Illinois, 8 Sept 2005.

Commute Shed Analysis

Introduction

MegaRegions have been defined as interconnected metropolitan areas.³ These connections may take many forms; these include:

1. Physical connections via rail and road
2. Environmental connections via a shared watershed
3. Production and work-related connections
 - a. Producers and suppliers
 - b. Corporate parent and subsidiaries
 - c. Places of residence and places of work

Places of residence and places of work describe two places that are connected by *commute sheds*. A commute shed is the area from which businesses in a particular location draw employees. When examined using the methods described below, some metropolitan areas appear tied together by counties sending large volumes of trips to multiple employment centers. However, the majority of counties in the Great Lakes MegaRegion are linked only to a single metropolitan area; thus, commute sheds do not serve to unite the region. The analysis had two steps:

1. *Macro-Level Connectivity*: counted the number of people moving into, out-of, and within the MegaRegion for work. This quantified the commute relationship between the MegaRegion and the rest of the United States.
2. *Major Metropolitan Area⁴ Connectivity*: measured the number of people from each county in the MegaRegion traveling to the fourteen largest consolidated metropolitan

³ Armando Carbonell and Robert Yaro. "American Spatial Development and the New Megalopolis." *Land Lines*. Volume 17, Number 2. April 2005. Last accessed April 2006: <http://www.lincolnst.edu/pubs/pub-detail.asp?id=1009>. Their work is based upon the work of European spatial planners, discussed in Andreas Faludi, *European Spatial Planning*. Cambridge, MA: Lincoln Institute of Land Policy. 2002

⁴The following definitions, directly from a glossary provided by the US Census (<http://www.census.gov/geo/www/tiger/glossary.html#glossary> last accessed April 2006, helped to clarify the meaning of these terms:

Metropolitan Statistical Area (MSA) - A geographic entity designated by the federal Office of Management and Budget for use by federal statistical agencies. A metropolitan statistical area (MSA) is a metropolitan area that is not closely associated with another metropolitan area. A metropolitan area consists of one or more counties, except in New England, where MSAs are defined in terms of county subdivisions (primarily cities and towns).

Primary Metropolitan Statistical Area (PMSA) - A geographic entity designated by the federal Office of Management and Budget for use by federal statistical agencies. If an area that qualifies as a metropolitan area has a census population of one million or more, two or more primary metropolitan statistical areas (PMSAs) may be defined within it if they meet official standards and local opinion favors the designation. When PMSAs are established within a metropolitan area; that metropolitan area is designated a consolidated metropolitan statistical area (CMSA).

Consolidated Metropolitan Statistical Area (CMSA) - A geographic entity designated by the federal Office of Management and Budget (OMB) for use by federal statistical agencies. An area becomes a consolidated metropolitan statistical area (CMSA) if it qualifies as a metropolitan area (MA), has a census population of 1,000,000 or more, has component parts that qualify as primary metropolitan statistical areas (PMSAs) based on official standards, and local opinion favors the designation.

statistical area (CMSAs) and metropolitan statistical areas (MSAs) for work. This revealed the counties that tie together the MegaRegion by sending workers to several of the largest metropolitan areas. Conceptually these counties may be considered the “glue” or “fascia” of the MegaRegion.

3. *Micro-Level Connectivity*: was not completed due to time constraints. This step would measure the number of people from each county in the MegaRegion traveling to a particular CMSA or MSA in greater detail. The manner in which the largest metropolitan areas attract fewer workers with greater distance would be determined and compared. Rather than depend on dot density maps and visual inference to judge connectivity, this step would produce numerical results attached to each county that could be queried. An example query could be to highlight all counties that:

Send commute flows to at least three of the largest fourteen metropolitan areas AND the magnitude of each one of those flows is at least 500 people AND the sum of all the flows is at least 50% of the county’s total workforce.

Counties meeting such criteria help to transform the Great Lakes Megaregion into a cohesive whole. People living in these counties may have a greater sense of regional identity.

Data & Methods

This analysis relied upon the following sources:

1. *Journey-to-Work Datafile(s)*.⁵ Provides county-to-county commute flow data for the entire United States in the following format:

1	2	3	4	5	6	7	8	9	10	11
FIPS	FIPS	FIPS	FIPS	county	FIPS	FIPS	FIPS	FIPS	County	Flow
State	County	MSA/	PMSA	name &	State	County	MSA/	PMSA	name	Size
Code	Code	CMSA	home	state	Code	work	CMSA	work	& state	(Count)
home	home	home		home	work		work		work	

The first five columns identify the county of residence, the second five columns identify the county of work, and the final column provides the number of people making that commute. Since the spatial data used in ArcGIS included FIPS numbers, columns 3 and 8 were used to define a commute and column 11 provided its magnitude.

Metropolitan Area is a collective term established by the OMB in 1990 to refer to metropolitan statistical areas (MSAs), consolidated metropolitan statistical areas (CMSAs), primary metropolitan statistical areas (PMSAs), and New England County Metropolitan Areas (NECMAs).

⁵ U.S. Census, http://www.census.gov/population/cen2000/commuting/2KRESCO_US.zip. Last accessed February 2006. Data for individual states are available at <http://www.census.gov/population/www/cen2000/commuting.html>

While requiring intensive computing to process, using the entire United States data file allowed the analysis to capture all commutes into and out of the region, except for international flows and those from New England, which has a different data format.⁶ Downloading individual files for the states within the MegaRegion would have saved disk space, since the average state file is 250 kilobytes (K) and the file for the entire United States is nearly 22 megabytes (M). However, the macro-level analysis required data for at least the states surrounding the MegaRegion in order to measure how it is connected to the United States in workers' commuting.

The data collection can be summarized as:

- a. Download commute flow data for the whole USA;
 - b. Download flow data only for the states in the MegaRegion and neighboring states; and
 - c. Download flow data for only MegaRegion but do not perform any analysis of connectivity with the rest of the United States.
2. *1999 CMSA & MSA List.*⁷ This lists all the CMSAs and MSAs in the 2000 Census along with the counties in each.

Steps in Commute Shed Analysis:

1. *Data Cleaning:* SPSS was used to assign any blank spaces or other markers representing no-data to the value 9999 in the Journey-to-Work file. This was necessary for international commutes and for those from the New England states, which did not follow the same FIPS system for counties. Anything with this no-data marker was still tracked in the macro-level analysis, which tallied flows within, out-of, and into the MegaRegion. However, it was ignored in the major MSA connectivity analysis because no point could be placed on a map showing the location of origin since 9999 was used for all “other” places. The maps produced for this project did not extend as far east as New England so the affect of this omission was not visible. The most significant impact on results is the lack of flow to and from Canada.

2. The CMSA & MSA lists were used to create sets of counties in MATLAB, a mathematical software package and scripting language. The script read in the U.S. flow file, treating each column as a separate “vector”—a series of values. MATLAB scripts are groups of commands issued to the software to perform tasks, such as to load data or to tally columns of numbers. MATLAB allows custom problem solving with large data sets. For example, the U.S. flow file is 170,000 lines long, but Microsoft EXCEL will only read files 65,536 lines long. MATLAB has no such restriction.⁸

⁶ New England is the six states of Maine, Vermont, New Hampshire, Massachusetts, Rhode Island, and Connecticut. CMSAs and MSAs in New England are city-and-town-based, while they are county based in the rest of the US. They were initially treated as international flows for the purpose of prototyping the process. Due to time constraints, they were never readdressed. The New England County Metropolitan Area (NECMA) scheme may allow the included script to handle these states in the future.

⁷ Located from the US Census at <http://www.census.gov/population/estimates/metro-city/99mfips.txt> Last accessed February 2006.

⁸ The scripts for this project were written by David Epstein and are available in the appendix of this methods guide. Contact information: davideps@umich.edu.

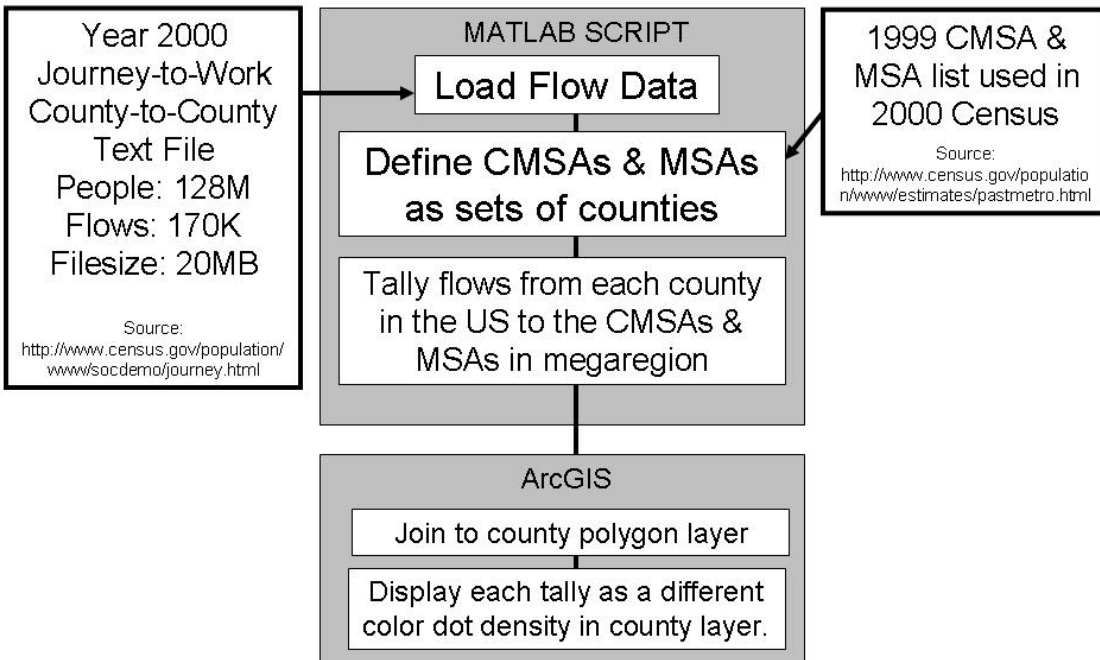
3. For the macro-level analysis, the script did the following:
 - a. Loaded the FIPS codes for all the counties in the MegaRegion;
 - b. Loaded the commuter flow file;
 - c. Traversed the flow file one line at a time, comparing origin and destination counties to the set of counties in the MegaRegion;
 - d. Recorded the count (flow volume) for any commute:
 - i. Into the MegaRegion
 - ii. Out of the MegaRegion
 - iii. Within the MegaRegion
 - e. Ignored flows originating and terminating outside of the MegaRegion; and
 - f. Produced output for the total for d.i,ii, and iii above.

4. For the major metropolitan areas connectivity analysis, the script did the following:
 - a. Defined fourteen sets of FIPS codes, one for each of the largest CMSAs and MSAs in the MegaRegion;
 - b. Loaded the commuter flow file;
 - c. Traversed the flow file one line at a time, comparing destination counties to the set of counties in the MegaRegion;
 - d. Recorded the count (flow volume) of every county in the country to each of the fourteen centers of employment under study;
 - e. Ignored flows terminating in other than these fourteen CMSAs and MSAs; and
 - f. Produced output of a text file with fifteen columns:
 - i. The first column is the FIPS number of the county of origin
 - ii. The next fourteen columns are the flow volume to each of the fourteen CMSAs and MSAs.

ArcGIS matched (“joined”) the FIPS number of the county of origin in the MATLAB output file to a spatially referenced county layer. The result was illustrated as a dot density map in which the location of the dot represented the home county and the color of the dot represented one of the fourteen largest MSAs or CMSAs in the region. Therefore, the map illustrates the volume of people in a particular county who work in a particular center of employment. Counties with more dots send more workers to the largest CMSAs and MSAs. Counties with dots of multiple colors send workers to several metropolitan areas and tie the MegaRegion together through commute sheds. As with the suburban and rural counties, the dots actually located in the metropolitan areas represent people who live there and the color represents where they work. People who work outside the fourteen CMSAs and MSAs are not represented, regardless of where they live.

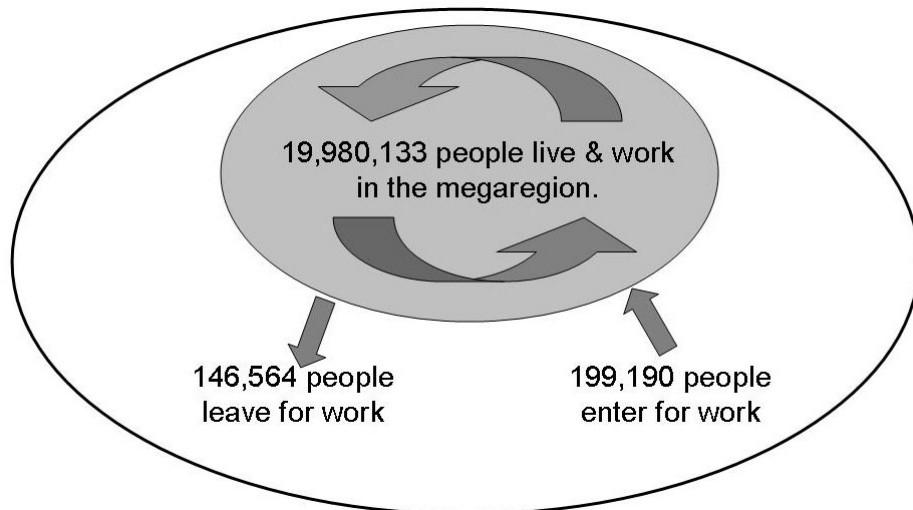
A flowchart of these steps is given below:

Figure 8. A System for Computing and Displaying Commute Sheds



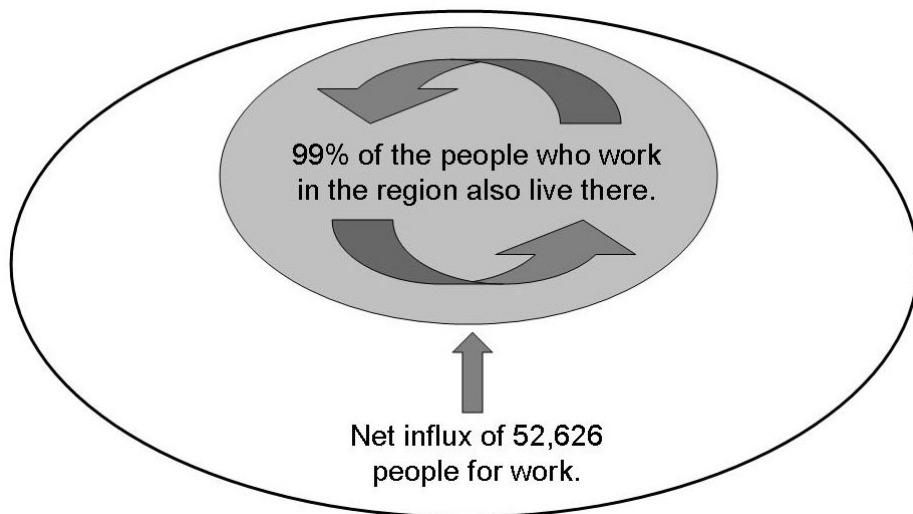
Results of Analysis

Figure 9. Connections Between the MegaRegion and the rest of the US via Worker Commutes



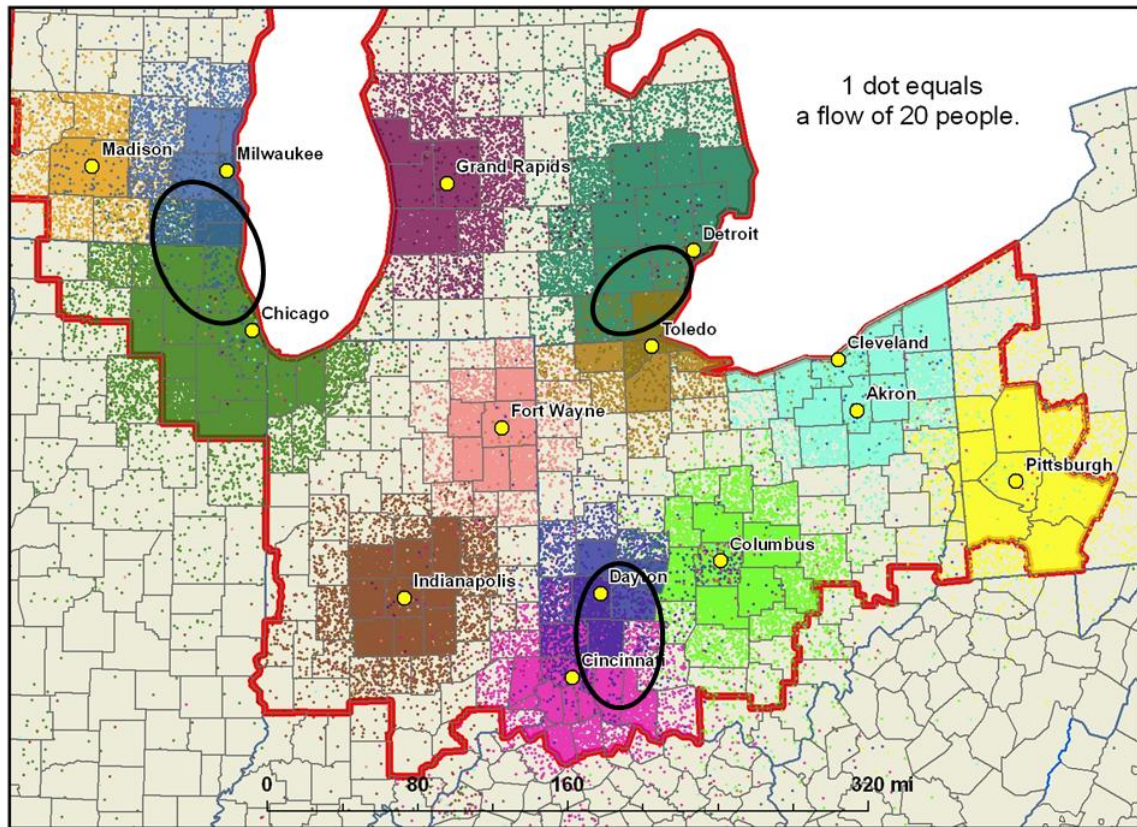
Source: Custom MATLAB scripts, which use the Year 2000 Journey-to-Work County-to-County Worker Flow files for the entire United States, available from the US Census Bureau (<http://www.census.gov/population/www/socdemo/journey.html>) last accessed April 3, 2006.

Figure 10. Nearly Everyone Who Works within the MegaRegion also Lives within the MegaRegion



Source: Custom MATLAB scripts, which use the Year 2000 Journey-to-Work County-to-County Worker Flow files for the entire United States, available from the US Census Bureau (<http://www.census.gov/population/www/socdemo/journey.html>) last accessed April 3, 2006.

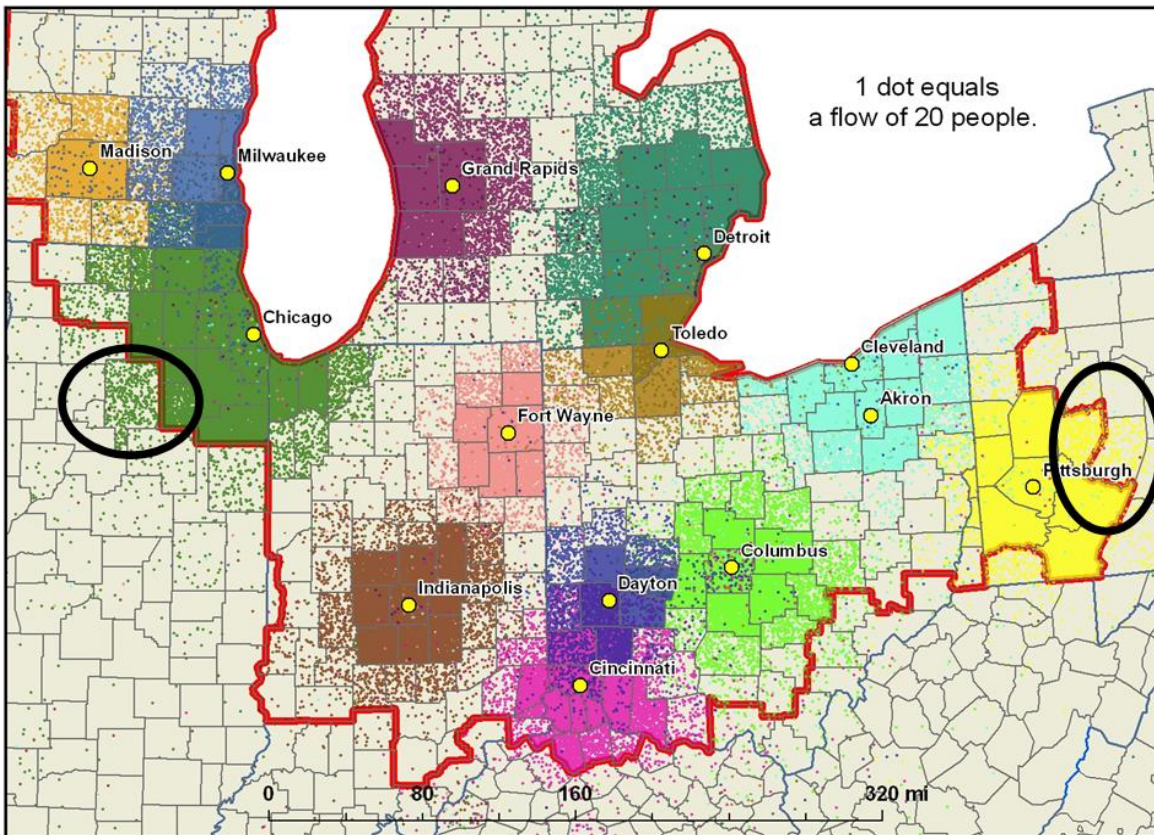
Figure 11. Several MSAs within the MegaRegion Appear Joined via Commute Flows



Source: Custom MATLAB scripts, Year 2000 Journey-to-Work County-to-County Worker Flow files for the entire United States, available from the US Census Bureau (<http://www.census.gov/population/www/socdemo/journey.html>) last accessed April 3, 2006, and ESRI's shapefile library

MSAs such as Chicago and Milwaukee appear connected via commute sheds through the counties circled above, which send a large volume of workers to both centers of employment. The situation appears similar for Detroit and Toledo and Dayton and Cincinnati. However, most of the MegaRegion is fragmented, meaning that most counties send workers to only one of the centers of employment studied—or to none at all. The amount of white space in this figure signifies that many counties are not strongly tied to any of the largest fourteen metropolitan areas in the MegaRegion.

Figure 12. Commute Flows into the MegaRegion are Concentrated at Two Locations



Source: Custom MATLAB scripts, Year 2000 Journey-to-Work County-to-County Worker Flow files for the entire United States, available from the US Census Bureau (<http://www.census.gov/population/www/socdemo/journey.html>) last accessed April 3, 2006, and ESRI's shapefile library.

The circled counties outside the boundary possibly should be added to the MegaRegion in the future due to their ties to the MegaRegion via work. The process used to define the MegaRegion did not subdivide CMSAs and MSAs. This means that commute sheds were considered indirectly since the census considers them in the construction of CMSAs and MSAs. Nevertheless, these counties send high volumes of workers into the MegaRegion.

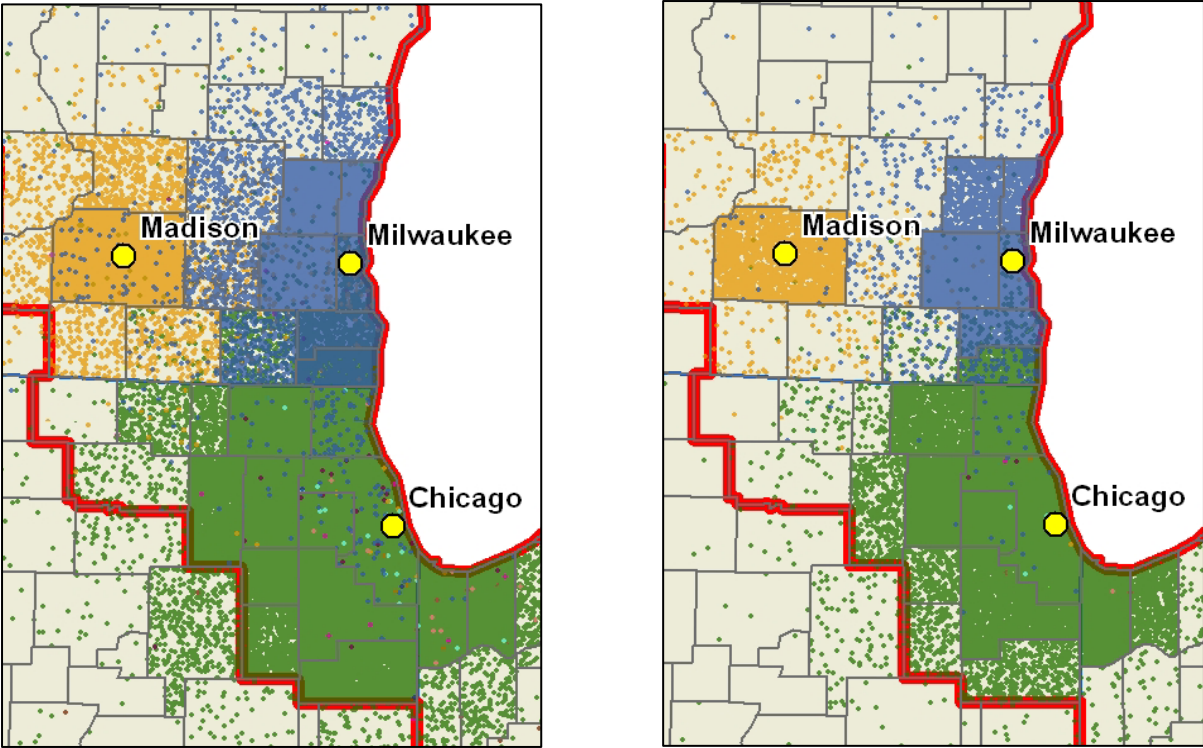
Commentary

The macro-level analysis is a first step in examining commute sheds. The simple diagrams shown above are useful in communicating results. However, the MATLAB script could be improved, allowing it to run faster. Still, any modern computer should be able to process the whole U.S. flow file using the provided script in a matter of minutes.

Two issues with the major metropolitan area connectivity analysis need to be considered:

1. No standard methods exist to measure connectivity in MegaRegions via commute sheds or to interpret results. Is measuring commuting to the largest CMSAs and MSAs important to defining and understanding MegaRegions? This project proceeded with the assumption that such a study would be informative in an analysis of connectivity. Yet, without common indicators and points of comparison, results have very little meaning. For example, if the detailed query described previously were widely adopted as an indicator, comparisons of MegaRegion commute shed connectivity would be possible. Researchers could compare the number or percentage of counties that match such a query in the different MegaRegions.
2. The dot-density mapping technique relies on a heavily computational initial step to measure flows, but interpretation of results is purely through visual inference. *Symbology*, the term used in ArcGIS for the software settings used to render an image, significantly influences the final map. Figures 11 & 12 illustrate this phenomenon. The overlap between Madison and Milwaukee nearly disappears when the symbology changes, but the overlap between Milwaukee and Chicago persists. The same data can make an area look more or less connected through commute sheds. While perhaps a useful tool for data exploration and for illustration, this technique does not qualify as a reliable indicator. More quantitative methods are necessary.

Figures 13 & 14. The Deception of Symbology



Source: Custom MATLAB scripts, Year 2000 Journey-to-Work County-to-County Worker Flow files for the entire United States, available from the US Census Bureau (<http://www.census.gov/population/www/socdemo/journey.html>) last accessed April 3, 2006, and ESRI's shapefile library.

Commute Shed Analysis MATLAB Scripts

NOTE: Those lines that begin with a percentage sign (%) are comments and not actually part of the code.

%SCRIPT 1: MACRO LEVEL COMMUTE SHED ANALYSIS

%David Epstein

clear all;
close all;

% READ IN FULL US JOURNEY TO WORK DATA

% full description of the original formal file is available at:
% "Record Layout for Census 2000 County-to-County Worker Flow Files"
% <http://www.census.gov/population/cen2000/commuting/coxcolayout.txt>

% This script reads in a modified version of the full US 2000 flow file
% All Non-US origins or destinations have been given dummy county numbers
% All US origins or destinations without county codes (New England, etc)
% have been given dummy data as placeholders. Without this Matlab cannot
% read the file

% Read whole file first
disp('loading all US flows...');
flowfile=load('USworkerflow2000.dat');

% Break up the file into columns representing aspects of the flow, using
% variable names from the original file description

resST=flowfile(:,1); %FIPS State Code of Residence
resCO=flowfile(:,2); %FIPS County Code of Residence
resMSA=flowfile(:,3);%FIPS Metropolitan Statistical Area (MSA) or
% Consolidated Metropolitan Statistical Area (CMSA)
% Code of Residence. This field contains the 4-digit
% code for the MSAs and CMSAs designated June 30, 1999.
% 9999 Appears for nonmetropolitan areas. * Appears for
% counties in the six New England states for which more
% than one code is applicable. [GIVEN DUMMY 6 digit
% numbers]
resPMSA=flowfile(:,4);

wrkST=flowfile(:,5); %FIPS State Code of work

Commute Shed Analysis – MATLAB Scripts
The Great Lakes MegaRegion

```
wrkCO=flowfile(:,6); %FIPS County Code of work
wrkMSA=flowfile(:,7);%FIPS Metropolitan Statistical Area (MSA) or
    %Consolidated Metropolitan Statistical Area (CMSA)
wrkPMSA=flowfile(:,8); %FIPS Primary Metropolitan Statistical Area (PMSA)
    %Code of work. OTHERWISE SAME AS resMSA

count=flowfile(:,9); %Number of Workers 16 years old and over in the
    %commuter flow (Right justified; blank filled)

resSTCO=resST*1000+resCO;
wrkSTCO=wrkST*1000+wrkCO;

US_N=length(resSTCO);

US_flow=[resST,resCO,resSTCO,resMSA,resPMSA,wrkST,wrkCO,wrkSTCO,wrkMSA,wrkP
MSA,count];

%           READ IN GREAT LAKES MEGAREGION COUNTIES
% Read whole file first
% megafile=load('MegaRegionCounties.dat');
% Break up the file into columns
% megaST=megafile(:,1); %FIPS State Code of Mega Region County
% megaCO=megafile(:,2); %FIPS County Code of Mega Region County
% megaFIPS=megafile(:,3); %unclear what this FIPS code is for
% megaSTCO=megaST*1000+megaCO;
% mega_N=length(megaSTCO);
% store all flows having Megaregion as origin or destination
% j=0;
% for i=1:US_N
%   if ismember(resSTCO(i),megaSTCO) | ismember(wrkSTCO(i),megaSTCO)
%     j=j+1;
%     megaflow(j,:)= US_flow(i,:);
%   end
% disp(i); disp(j);
% end

% DEFINE CMSAs and MSAs AS SETS OF COUNTIES
Detroit=[26091 26093 26161 26087 26099 26115 26125 26147 26163 26049];
Chicago=[17031 17037 17043 17063 17089 17093 17097 17111 17197 18089 18127 17091
55059];
Cleveland=[39133 39153 39007 39035 39055 39085 39093 39103];
Columbus=[39041 39045 39049 39089 39097 39129];
Indianapolis=[18011 18057 18059 18063 18081 18095 18097 18109 18145];
Milwaukee=[55079 55089 55131 55133 55101];
Pittsburgh=[42003 42007 42019 42051 42125 42129];
```

```
Cincinnati=[18029 18115 21015 21037 21077 21081 21117 21191 39015 39025 39061 39165  
39017];  
Dayton=[39023 39057 39109 39113];  
FortWayne=[18001 18003 18033 18069 18179 18183];  
GrandRapids=[26005 26081 26121 26139];  
Toledo=[39051 39095 39173];  
Madison=[55025];
```

```
% DIMENSION ALL FLOWS TO ZERO
```

```
W(1:100000)=0;
```

```
L(1:100000)=0;
```

```
B(1:100000)=0;
```

```
% LOOP THROUGH US FLOWS AND STORE FLOWS TO DEFINED CMSAs & MSAs
```

```
disp(' ')
```

```
disp('Tallying flows to megaregion CMSAs & MSAs...');
```

```
disp('percent complete:');
```

```
j=0;
```

```
for i=1:US_N
```

```
    if ismember(wrkSTCO(i),Detroit)
```

```
        W(resSTCO(i)) = W(resSTCO(i)) + count(i);
```

```
    end
```

```
    if ismember(resSTCO(i),Detroit)
```

```
        L(resSTCO(i)) = L(resSTCO(i)) + count(i);
```

```
    end
```

```
    if (ismember(resSTCO(i),Detroit) & ismember(wrkSTCO(i),Detroit))
```

```
        B(resSTCO(i)) = B(resSTCO(i)) + count(i);
```

```
    end
```

```
        j=j+1;
```

```
        if j==10000
```

```
            disp((i/US_N)*100)
```

```
            j=0;
```

```
        end
```

```
end
```

```
% PRINT FLOW TOTALS
```

```
disp(' ')
```

```
disp('Total flows to CMSAs & MSAs')
```

```
disp('-----')
```

```
sprintf('Work %.0f',sum(W))
```

```
sprintf('Live %.0f',sum(L))
```

```
sprintf('Both %.0f',sum(B))
```

```
totaldailyFLOW=sum(W)+sum(L)-sum(B)
```

```
disp(' ')
```

%SCRIPT 2: MAJOR METROPOLITAN AREAS CONNECTIVITY ANALYSIS

% David Epstein

clear;

close all;

% READ IN FULL US JOURNEY TO WORK DATA

% full description of the original formal file is available at:

% "Record Layout for Census 2000 County-to-County Worker Flow Files"

% <http://www.census.gov/population/cen2000/commuting/coxcolayout.txt>

% This script reads in a modified version of the full US 2000 flow file

% All Non-US origins or destinations have been given dummy county numbers

% All US origins or destinations without county codes (New England, etc)

% have been given dummy data as placeholders. Without this Matlab cannot

% read the file

% Read whole file first

disp('loading all US flows...');

flowfile=load('USworkerflow2000.dat');

% Break up the file into columns representing aspects of the flow, using

% variable names from the original file description

resST=flowfile(:,1); %FIPS State Code of Residence

resCO=flowfile(:,2); %FIPS County Code of Residence

resMSA=flowfile(:,3);%FIPS Metropolitan Statistical Area (MSA) or

% Consolidated Metropolitan Statistical Area (CMSA)

% Code of Residence. This field contains the 4-digit

% code for the MSAs and CMSAs designated June 30, 1999.

% 9999 Appears for nonmetropolitan areas. * Appears for

% counties in the six New England states for which more

% than one code is applicable. [GIVEN DUMMY 6 digit

% numbers]

resPMSA=flowfile(:,4);

wrkST=flowfile(:,5); %FIPS State Code of work

wrkCO=flowfile(:,6); %FIPS County Code of work

wrkMSA=flowfile(:,7);%FIPS Metropolitan Statistical Area (MSA) or

% Consolidated Metropolitan Statistical Area (CMSA)

wrkPMSA=flowfile(:,8); %FIPS Primary Metropolitan Statistical Area (PMSA)

% Code of work. OTHERWISE SAME AS resMSA

count=flowfile(:,9); %Number of Workers 16 years old and over in the

% commuter flow (Right justified; blank filled)


```
resSTCO=resST*1000+resCO;
wrkSTCO=wrkST*1000+wrkCO;

US_N=length(resSTCO);

US_flow=[resST,resCO,resSTCO,resMSA,resPMSA,wrkST,wrkCO,wrkSTCO,wrkMSA,wrkP
MSA,count];

%           READ IN GREAT LAKES MEGAREGION COUNTIES

% Read whole file first
% megafile=load('MegaRegionCounties.dat');
% Break up the file into columns
% megaST=megafile(:,1); %FIPS State Code of Mega Region County
% megaCO=megafile(:,2); %FIPS County Code of Mega Region County
% megaFIPS=megafile(:,3); %unclear what this FIPS code is for
% megaSTCO=megaST*1000+megaCO;
% mega_N=length(megaSTCO);
% store all flows having Megaregion as origin or destination
% j=0;
% for i=1:US_N
%   if ismember(resSTCO(i),megaSTCO) | ismember(wrkSTCO(i),megaSTCO)
%     j=j+1;
%     megaflow(j,:)= US_flow(i,:);
%   end
% disp(i); disp(j);
% end

% DEFINE CMSAs and MSAs AS SETS OF COUNTIES
Detroit=[26091 26093 26161 26087 26099 26115 26125 26147 26163 26049];
Chicago=[17031 17037 17043 17063 17089 17093 17097 17111 17197 18089 18127 17091
55059];
Cleveland=[39133 39153 39007 39035 39055 39085 39093 39103];
Columbus=[39041 39045 39049 39089 39097 39129];
Indianapolis=[18011 18057 18059 18063 18081 18095 18097 18109 18145];
Milwaukee=[55079 55089 55131 55133 55101];
Pittsburgh=[42003 42007 42019 42051 42125 42129];
Cincinnati=[18029 18115 21015 21037 21077 21081 21117 21191 39015 39025 39061 39165
39017];
Dayton=[39023 39057 39109 39113];
FortWayne=[18001 18003 18033 18069 18179 18183];
GrandRapids=[26005 26081 26121 26139];
Toledo=[39051 39095 39173];
Madison=[55025];
```

Commuter Shed Analysis – MATLAB Scripts
The Great Lakes MegaRegion

```
% DIMENSION ALL FLOWS TO ZERO
DetroitFlow(1:100000)=0;
ChicagoFlow(1:100000)=0;
ClevelandFlow(1:100000)=0;
ColumbusFlow(1:100000)=0;
IndianapolisFlow(1:100000)=0;
MilwaukeeFlow(1:100000)=0;
PittsburghFlow(1:100000)=0;
CincinnatiFlow(1:100000)=0;
DaytonFlow(1:100000)=0;
FortWayneFlow(1:100000)=0;
GrandRapidsFlow(1:100000)=0;
ToledoFlow(1:100000)=0;
MadisonFlow(1:100000)=0;

% LOOP THROUGH US FLOWS AND STORE FLOWS TO DEFINED CMSAs & MSAs
disp(' ')
disp('Tallying flows to megaregion CMSAs & MSAs...');
disp('percent complete:');
j=0;
for i=1:US_N
    if ismember(wrkSTCO(i),Detroit)
        DetroitFlow(resSTCO(i)) = DetroitFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Chicago)
        ChicagoFlow(resSTCO(i)) = ChicagoFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Cleveland)
        ClevelandFlow(resSTCO(i)) = ClevelandFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Columbus)
        ColumbusFlow(resSTCO(i)) = ColumbusFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Indianapolis)
        IndianapolisFlow(resSTCO(i)) = IndianapolisFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Milwaukee)
        MilwaukeeFlow(resSTCO(i)) = MilwaukeeFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Pittsburgh)
        PittsburghFlow(resSTCO(i)) = PittsburghFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Cincinnati)
        CincinnatiFlow(resSTCO(i)) = CincinnatiFlow(resSTCO(i)) + count(i);
    end
    if ismember(wrkSTCO(i),Dayton)
```

```
    DaytonFlow(resSTCO(i)) = DaytonFlow(resSTCO(i)) + count(i);
end
if ismember(wrkSTCO(i),FortWayne)
    FortWayneFlow(resSTCO(i)) = FortWayneFlow(resSTCO(i)) + count(i);
end
if ismember(wrkSTCO(i),GrandRapids)
    GrandRapidsFlow(resSTCO(i)) = GrandRapidsFlow(resSTCO(i)) + count(i);
end
if ismember(wrkSTCO(i),Toledo)
    ToledoFlow(resSTCO(i)) = ToledoFlow(resSTCO(i)) + count(i);
end
if ismember(wrkSTCO(i),Madison)
    MadisonFlow(resSTCO(i)) = MadisonFlow(resSTCO(i)) + count(i);
end
j=j+1;
if j==10000
    disp((i/US_N)*100)
    j=0;
end
end
```

%PRINT FLOW TOTALS

```
disp(' ')
disp('Total flows to CMSAs & MSAs')
disp('-----')
sprintf('Detroit %.0f',sum(DetroitFlow))
sprintf('Chicago %.0f',sum(ChicagoFlow))
sprintf('Cleveland %.0f',sum(ClevelandFlow))
sprintf('Columbus %.0f',sum(ColumbusFlow))
sprintf('Indianapolis %.0f',sum(IndianapolisFlow))
sprintf('Milwaukee %.0f',sum(MilwaukeeFlow))
sprintf('Pittsburgh %.0f',sum(PittsburghFlow))
sprintf('Cincinnati %.0f',sum(CincinnatiFlow))
sprintf('Dayton %.0f',sum(DaytonFlow))
sprintf('Fort Wayne %.0f',sum(FortWayneFlow))
sprintf('Grand Rapids %.0f',sum(GrandRapidsFlow))
sprintf('Toledo %.0f',sum(ToledoFlow))
sprintf('Madison %.0f',sum(MadisonFlow))
disp(' ')
```

%TRANSPOSE FLOWS SO THAT THEY ARE COLUMNS AND NOT ROWS

```
FIPS=[1:100000];
FIPS=FIPS';
ChicagoFlow=ChicagoFlow';
ClevelandFlow=ClevelandFlow';
ColumbusFlow=ColumbusFlow';
```

```

DetroitFlow=DetroitFlow';
IndianapolisFlow=IndianapolisFlow';
MilwaukeeFlow=MilwaukeeFlow';
PittsburghFlow=PittsburghFlow';
CincinnatiFlow=CincinnatiFlow';
DaytonFlow=DaytonFlow';
FortWayneFlow=FortWayneFlow';
GrandRapidsFlow=GrandRapidsFlow';
ToledoFlow=ToledoFlow';
MadisonFlow=MadisonFlow';

% WRITE FLOWFILE OF ALL COUNTIES WITH NON-ZERO FLOWS
disp('Writing flowfile.txt')
j=0;
for i=1:100000
    if
    (ChicagoFlow(i)+ClevelandFlow(i)+ColumbusFlow(i)+DetroitFlow(i)+IndianapolisFlow(i)+MilwaukeeFlow(i)+PittsburghFlow(i)+CincinnatiFlow(i)+DaytonFlow(i)+FortWayneFlow(i)+GrandRapidsFlow(i)+ToledoFlow(i)+MadisonFlow(i)) > 0
        j=j+1;
        cityflow(j,:)= [i ChicagoFlow(i) ClevelandFlow(i) ColumbusFlow(i) DetroitFlow(i) IndianapolisFlow(i) MilwaukeeFlow(i) PittsburghFlow(i) CincinnatiFlow(i) DaytonFlow(i) FortWayneFlow(i) GrandRapidsFlow(i) ToledoFlow(i) MadisonFlow(i)];
    end
end

fid=fopen('14MSAflows.txt','w');
for i=1:length(cityflow(:,1))
    fprintf(fid,'%d %d %d %d %d %d %d %d %d %d %d %d %d %d\n',cityflow(i,:));
end
fclose(fid);

```

Examples of Input Data

This is a piece of the US Journey-to-Work county-level flow file described earlier. “9999” signifies missing data.

8	13	2082	1125	2	185	9999	9999	5
8	13	2082	1125	4	13	6200	9999	54
8	13	2082	1125	4	19	8520	9999	11
8	13	2082	1125	6	1	7362	5775	32
8	13	2082	1125	6	13	7362	5775	10
8	13	2082	1125	6	29	680	9999	9
8	13	2082	1125	6	37	4472	4480	54
8	13	2082	1125	6	59	4472	5945	3

8	13	2082	1125	6	65	4472	6780	8
8	13	2082	1125	6	71	4472	6780	10
8	13	2082	1125	6	73	7320	9999	37
8	13	2082	1125	6	75	7362	7360	29
8	13	2082	1125	6	81	7362	7360	29
8	13	2082	1125	6	85	7362	7400	48
8	13	2082	1125	6	87	7362	7485	8
8	13	2082	1125	6	111	4472	8735	9
8	13	2082	1125	8	1	2082	2080	6137

Examples of Output Data

This is a piece of the final file produced by the Major MSA Connectivity Analysis script. The first column is the FIPS number of the county of origin. The next thirteen numbers are flow tallies to the major MAs (Cleveland and Akron are merged). Joining this file to a county GIS file using the FIPS code allows the flows to be rendered as dot densities.

```

38055 0 0 0 2 0 0 0 0 0 0 0 0 0
38059 0 0 0 4 0 0 0 8 0 0 0 0 0
38065 3 0 0 0 0 0 0 0 0 0 0 0 0
38097 2 0 0 0 0 0 0 0 0 0 0 0 0
38101 9 0 0 0 0 0 0 0 0 0 0 0 0
39001 0 11 50 19 7 0 0 2578 113 0 0 5 5
39003 12 20 88 32 15 0 6 42 133 79 0 118 0
39005 0 2001 142 12 0 0 0 17 10 12 0 18 0
39007 5 44070 19 12 4 0 35 13 25 5 0 37 0
39009 7 66 1308 3 0 0 10 76 13 2 0 25 0
39011 5 4 36 0 0 0 0 17 345 5 0 51 0
39013 9 71 168 0 6 9 382 8 27 0 10 5 0
39015 0 0 26 2 4 0 0 16661 38 8 0 0 0
39017 50 102 129 50 104 0 38 152779 5207 0 0 21 0

```


Determining Sprawl

Urban sprawl is a vague concept. Common definitions of sprawl include⁹:

1. “Sprawl is a low-density land use pattern that is automobile dependent, energy and land consumptive, and requires a very high ratio of road surface to development served;”
2. “Uncontrolled growth of urban development into previously rural areas. Sprawl usually refers to a mixture of land uses occurring in an unplanned pattern;” and
3. “The growth of a metropolitan area through the process of scattered development of miscellaneous types of land use in isolated locations on the fringe, followed by the gradual filling-in of the intervening spaces with similar uses.”

The Great Lakes MegaRegion plan defines urban sprawl as low-density, uncontrolled growth of urban development into previously rural areas. This definition combines the first two definitions above. The combination creates a definition that addresses the two main challenges in the region that attribute sprawl to rapidly urbanizing land outside established cities.

Existing definitions, did not meet the needs of the plan for the Great Lakes MegaRegion for several reasons. The first definition addresses most of the region’s challenges except the urbanization of rural land. Rural land is an asset to the region, and the definition of sprawl addresses the loss of this asset. The second definition does not address the low density-nature of the development, another concern that the region faces. This definition also states that this type of growth is unplanned, but many municipalities have ordinances and codes that mandate planning for low-density. The third definition did not address consumption of rural land or the low-density nature of the region’s development.

Measuring Sprawl

The Great Lakes MegaRegion plan compares two sprawl indices in order to measure the quantity of sprawl that is occurring. The index created by Reid Ewing and that of William Fulton, were chosen for this comparison because they ranked the metropolitan areas in the region in two distinct ways, which show that sprawl can be a problem even in cities that have high densities. For instance, the study by Ewing and his colleagues determines that Chicago is not a sprawling city, but the Fulton, et al., study determines that Chicago is sprawling.

Reid Ewing, Rolf Pendall, and Don Chen measured sprawl in 83 metropolitan areas. They used a “sprawl index based on four factors.” These factors are:

1. Residential density and a diverse housing stock;
2. Availability of jobs and quality of public services;
3. Strong downtowns with civic and cultural amenities; and
4. An easily navigable street network.¹⁰

⁹ Michigan Land Use Leadership Council. Defining Sprawl and Smart Growth. Accessed: http://www.michiganlanduse.org/resources/councilresources/Sprawl_SmartGrowth.pdf, March 2006. Pages 1-2.

¹⁰ Reid Ewing, Rolf Pendall, and Don Chen. Measuring sprawl and its impact. Accessed from: www.smartgrowthamerica.com/sprawlindex/sprawlindex.html, March 3 2006. Page 3.

Metropolitan areas with lower scores are sprawling the most, and the U.S. national average is 100. The following table lists the seven major metropolitan areas in the Great Lakes MegaRegion and their scores according to this sprawl index.¹¹

Table 1. Metropolitan Area and Sprawl Score

Metro Area	Sprawl Score
Detroit, MI PMSA	79.5
Cleveland, OH PMSA	91.8
Indianapolis, IN MSA	93.7
Cincinnati, OH--KY--IN PMSA	96.0
U.S. National Average	100
Pittsburgh, PA PMSA	105.9
Milwaukee, WI PMSA	117.3
Chicago, IL PMSA	121.2

Source: Reid Ewing, et al. "Measuring Sprawl and its impact." Accessed from www.smartgrowthamerica.com/sprawlindex/sprawlindex.html, March 2006.

William Fulton and his colleagues defined sprawl by the amount of urbanized land that an area gained between 1982 and 1997.¹² More specifically, the study examined rapidly growing metropolitan areas that are consuming land for urbanization. It compared how much new land was consumed for urbanization to how much population density of metropolitan areas changed (revised for parallel sentence structure). That is, the study looked at the growth of urbanized land from 1982 to 1997 and the growth in population density over the same time period and compared the two. If an area gained more urbanized land than it did population density, the area was considered sprawling. For example, Chicago, which was not considered sprawling in the Ewing, Pendall, and Chen study, lost 13% of its population density and gained 26% in urbanized land. This difference makes Chicago a sprawling city by this standard. Another example is Pittsburgh, the most sprawling city by this index, because it lost 36% in population density but gained 43% of new urbanized land. That is, in 1997 the population density of Pittsburgh was 36% higher than it was in 1982 and it had 43% less urbanized land in 1982 than in 1997.

¹¹ Ibid. Page 19.

¹² William Fulton, Rolf Pendall, Mai Nguyen, and Alicia Harrison. "Who Sprawls Most?" The Brookings Institution. July 2001. Accessed from: <http://www.brookings.edu/es/urban/publications/fulton.pdf>. March 9 2006. Page 5.

Table 2. Loss in Population Density and Growth in Urbanized Land

Metropolitan Statistical Area Region	Density 1997	Change in Population	Urbanized Land	Density
Milwaukee	3.93	7%	25%	-15%
Chicago	6.02	10%	26%	-13%
Detroit	4.27	5%	29%	-19%
Cleveland	4.03	0%	32%	-24%
Cincinnati	3.77	10%	40%	-21%
Indianapolis	3.58	20%	42%	-16%
Pittsburgh	3.72	-8%	43%	-36%

Source: William Fulton, Rolf Pendall, Mai Nguyen, and Alicia Harrison. . "Who Sprawls Most?" The Brookings Institution. July 2001. Accessed 9 Mar. 2006 <<http://www.brookings.edu/es/urban/publications/fulton.pdf>>.

Other studies have used different indicators for defining and measuring sprawl. For example, one index defines sprawl as "the reduction of rural land due to the increase of the total size of the land area of a city and its suburbs over a particular period of time."¹³ By this standard, Chicago is the most sprawling city out of the region's seven major metro areas, whereas the Ewing index ranks Chicago as the least sprawling. The following table lists the region's seven major metropolitan areas and their sprawl scores as defined by this index. No national average for metropolitan areas' growth in land area was given in this study.

Table 3. Square Miles of Sprawl

Urbanized Area (ranked by amount of sprawl)	Square Miles of Sprawl (growth in land area 1970-1990)
13. Chicago, IL - N.W. Indiana	307.3
18. Detroit, MI	247.4
24. Pittsburgh, PA	181.7
25. Cincinnati, OH-KY	176.6
59. Indianapolis, IN	87.7
76. Milwaukee, WI	55.5
100. Cleveland, OH	- 10.2

Source: Sprawl City. "Defining Sprawl." Accessed 15 Mar. 2006 <<http://www.sprawlcity.org>>

¹³ Sprawl City. "What is Sprawl?" Accessed: <http://www.sprawlcity.org/defining.html>. March 10, 2006

Calculating Regional Domestic Product

Numerous indicators help in analyzing the size of a region's economy. These include, for example, payroll, employment, and the sum of metropolitan areas' gross domestic product. In the Great Lakes MegaRegion plan, the approach was to estimate the region's share of Gross Domestic Product. Gross Domestic Product (GDP) measures the size of a country's economy. The Bureau of Economic Analysis defines GDP as "the market value of goods and services produced by labor and property in the United States."¹⁴ GDP is an important measurement of the strength of a nation's economy; therefore a similar measure would be an important measure of a region's economy.

Gross Domestic Product is disaggregated by state and by metropolitan area.¹⁵ However, the Great Lakes MegaRegion is the sum of many counties, both urban and rural, spanning portions of eight states. Using state-level data would not capture the region's share of gross domestic product. In addition, much economic activity occurs outside of the metropolitan areas but is important to understanding the economy of the Great Lakes MegaRegion as a whole.

Planners working on other MegaRegions have employed several methods to capture each region's gross domestic product. For instance, the University of Pennsylvania planners summed the regional product for major metropolitan areas, excluding other parts of the Northeast MegaRegion.¹⁶ The University of Texas – Austin planners defined the regional product for the Texas Triangle as the sum of the regional product of the four major metropolitan areas that make up the region.¹⁷

The economic structure of the Great Lakes MegaRegion includes areas outside the major metropolitan areas that contribute considerably to the regional economy. Significant economic activity occurs in rural areas in manufacturing, tourism, and agriculture. For this region, a method that includes metropolitan and other counties' contributions to gross product is desirable.

To determine the region's total contribution to Gross State Product (GSP), we first determined the amount of each state's total payroll that takes place within the region's boundary. The aim was to use an economic indicator available at the county level to estimate share of gross product in those counties. An alternative approach would have been to calculate the amount of employment that is generated in the counties of the region. The former method was used to determine the Great Lakes MegaRegion's share of Gross State Product.

The following are the steps used to define the total Gross Regional Product for the Great Lakes MegaRegion using payroll data by county as an indicator of the economic activity of the region.

¹⁴ Bureau of Economic Analysis, Glossary, http://www.bea.gov/nea/glossary/glossary_g.htm, accessed on April 17, 2006.

¹⁵ Gross State Product data are available from the Bureau of Economic Analysis (BEA). Standard & Poor's DRI calculates metropolitan product for all 319 metropolitan areas of the United States.

¹⁶ "Reinventing Megalopolis: The Northeast MegaRegion," University of Pennsylvania, Spring 2005.

¹⁷ University of Texas – Austin. Supercities Conference presentation. 10 Mar. 2006. Fundación Metropoli. Madrid, Spain.

Step 1: Download payroll data by county

Download industry payroll data from County Business Patterns for each state and county in the region.¹⁸ Calculate, by county, the amount of the total state's payroll earned within the region's boundaries.

Step 2: Download Gross State Product

Download Gross State Product (GSP) from the BEA for each state in the region and the nation.¹⁹

Step 3: Calculate the region's share of total State Payroll

Divide the total MegaRegion payroll by the total state payroll for each state in the region. This is the percentage of each state's total payroll earned in the region.

Step 4: Calculate the region's share of Gross State Product

Multiply the total Gross State Product by the region's share of total state payroll for each state in the region. This is the region's share of total Gross State Product, by state.

Step 5: Aggregate region's share of Gross State Product

Aggregate the state's share of Gross State Product for all states in the region. This is the region's total contribution to Gross Product.

Step 6: Calculate region's share of the nation's Gross Domestic Product

Divide the region's Gross Product by total United States Gross Domestic Product. This is the measure of the nation's economic activity attributable to the region.

Critique of Regional Product Calculations

Numerous methods exist to estimate gross regional product; all have shortcomings. Estimating gross product using the MegaRegion's share of state payroll from County Business Patterns has strengths and weaknesses in comparison to other approaches.

- County Business Patterns does not cover all payroll.²⁰ The data exclude self-employed persons, employees of private households, railroad employees, agricultural production workers, and most government employees.²¹ If the region has a larger than the national share of its economy in these sectors, using this method will underestimate gross product attributable to the region. If the region has a smaller than the national share of its economy in these sectors, using CBP data will overestimate gross regional product.

¹⁸ County Business Patterns, Accessed February 2006, <http://www.census.gov/epcd/cbp/view/cbpview.html>,

¹⁹ Bureau of Economic Analysis, Accessed March 2006, <http://www.bea.gov/bea/regional/gsp.htm>.

²⁰ Payroll includes "all forms of compensation, such as salaries, wages, reported tips, commissions, bonuses, vacation allowances, sick-leave pay, employee contributions to qualified pension plans, and the value of taxable fringe benefits. For corporations, it includes amounts paid to officers and executives; for unincorporated businesses, it does not include profit or other compensation of proprietors or partners. Payroll is reported before deductions for Social Security, income tax, insurance, union dues, etc.," from <http://www.census.gov/epcd/cbp/view/genexpl.html>.

²¹ CBP Definitions and Information, <http://www.census.gov/epcd/cbp/view/genexpl.html>, accessed on April 6, 2006.

Employment and payroll data for all levels of government employees are collected by the United States Census of Governments.²² Government payroll data are available annually for the one-month period of March.²³ Careful consideration of merging these two data sets into a complete regional economic activity dataset could be a priority of future MegaRegion analysis.

- Payroll may overestimate regional product in the Great Lakes MegaRegion. Due to the concentration of manufacturing in the region (location quotient = 1.36 in 2003²⁴) and the higher wages associated with this industry, using payroll data might overestimate the regional product. Employment data are more available than payroll and could be used to estimate regional product. Employment numbers for self-employed and government employees are available from the Bureau of Economic Analysis, Regional Economic Accounts.²⁵ On the other hand, using employment numbers might understate regional product. Using payroll data to estimate the gross product of the region for 1999 indicates the region's share of GDP is 16.0%, while using employment numbers yields a regional product estimate of 15.7%.
- Metropolitan areas' regional product would underestimate Great Lakes MegaRegion regional product more than some other regions' regional product. The United States Conference of Mayors estimates that nearly 85% of an area's economic activity, including income, employment, and the production of goods and services, is generated from within metro areas.²⁶ In the case of the Great Lakes MegaRegion, however, only 81.6% of the gross regional product in 1999 was generated from within its metropolitan areas.²⁷

²² State Government Employees include Education Services; Social Services and Income Maintenance; Transportation; Public Safety; Environment and Housing; Liquor Stores (if applicable); Utilities, including public mass transit; and Government Administration; Available from Census of Governments, <http://www.census.gov/govs/www/index.html>.

²³ US Census Bureau, Public Employment and Payroll Technical Documentation, <http://www.census.gov/govs/www/apestechdoc.html#survey>, accessed on April 6, 2006.

²⁴ Location quotient calculated using employment data from County Business Patterns

²⁵ Information is available from <http://www.bea.gov/bea/regional/reis/>.

²⁶ US Conference of Mayors, "City/County Metros Economic Gains," http://www.usmayors.org/citiesdrivetheeconomy/chart5_decade.pdf, accessed on April 7, 2006.

²⁷ The Metropolitan product of all MSAs in the region were aggregated using data compiled from Table 9 in "US Metro Economies: The Engines of America's Growth," <http://www.usmayors.org/citiesdrivetheeconomy/report2000.pdf>, accessed on April 17, 2006.

Measuring Residential Economic Vulnerability

Introduction

The plan for the Great Lakes MegaRegion identified seven major cities within the region—Chicago, Cleveland, Columbus, Detroit, Indianapolis, Milwaukee, and Pittsburgh. The MegaRegion’s major cities differ from their surrounding suburbs in residential economic vulnerability. Residential economic vulnerability is defined here as areas with concentrations of residents who are below the poverty level or could easily find themselves below this level due to slight changes in life circumstances. Analysis of the spatial distribution of residential economic vulnerability in the major cities and suburbs illustrates the inequity across the MegaRegion’s metropolitan areas. The analysis of residential economic vulnerability showed the spatial distribution of highly vulnerable city block groups. A Geographic Information System (GIS) map was produced for each of the seven metropolitan regions to investigate these issues.

Data and Methods

1. U.S. Census 2000 Summary Files (SF-3, Sample Data).

Four indicators of residential economic vulnerability were identified using the block group level, including:

- a. “Poverty Status in 1999 by Age over 17”²⁸: Used to calculate the percent of individuals over seventeen years old in poverty at the block group level as defined by the U.S. Census.
 - b. “Employment Status for the Population 16 Years and Over”²⁹: Used to compute the percent of unemployed residents at the block group level.
 - c. “Educational Attainment for the Population 25 Years and Over”³⁰: Used to determine the percent with no high school diploma at the block group level.
 - d. “Own Children Under 18 Years by Family Type by Age”³⁰: Used to calculate the percent of families with single mothers at the block group level.
2. Downloaded the following shapefile layers by county for each metropolitan area:
 - a Block Groups 2000²⁸
 - b Designated Places 2000²⁹
 - c PMSA Polygons 2000³⁰
 3. Calculated the metropolitan area average for each of the four indicators in order to compare spatial distribution within each of the seven metropolitan areas.

²⁸ As the smallest unit for which the U.S. Census collects sample data, the block group level was effective in assessing neighborhood level distribution of residential economic vulnerability.

²⁹ Provided the city boundaries for the MegaRegion’s major cities.

³⁰ Defined the major cities’ Metropolitan Statistical Areas (Columbus, Indianapolis, and Pittsburgh) and Primary Metropolitan Statistical Areas (Chicago, Cleveland-Lorain-Elyria, Detroit, and Milwaukee-Waukesha).

4. Selected block groups with all four indicators in order to map residential economic vulnerability:
 - a. Above the metropolitan area average.
 - b. 25% higher than the metropolitan area average.
 - c. Twice the metropolitan area average.

5. Assigned a yellow shade to all block groups at or below the average for all four indicators. Darker gradations of the color blue indicate levels of residential economic vulnerability above the average for all four indicators.

Commentary

One drawback to selecting block groups with all four indicators is that certain block groups with above average residential economic vulnerability may be regarded as below average. Such a phenomenon would occur if one of the four indicators was below the metropolitan average. All four indicators must be above their respective metropolitan average to appear on the map as in this situation; the block group would not appear as having above metropolitan average residential economic vulnerability on a map.

A primary goal of this analysis was to determine the concentration of residential economic vulnerability within each metropolitan area's major city. Figure 15 of the Milwaukee-Waukesha PMSA is the most striking example of this phenomenon as only a handful of the metropolitan residential economic vulnerable block groups are located beyond the City of Milwaukee's border. Figure 16 of the Indianapolis MSA and Figure 17 of the Detroit PMSA also demonstrate the concentration of metropolitan residential economic vulnerability within the MegaRegion's major cities' boundaries.

Figure 15. Metropolitan Residential Economic Vulnerability, Milwaukee-Waukesha PMSA

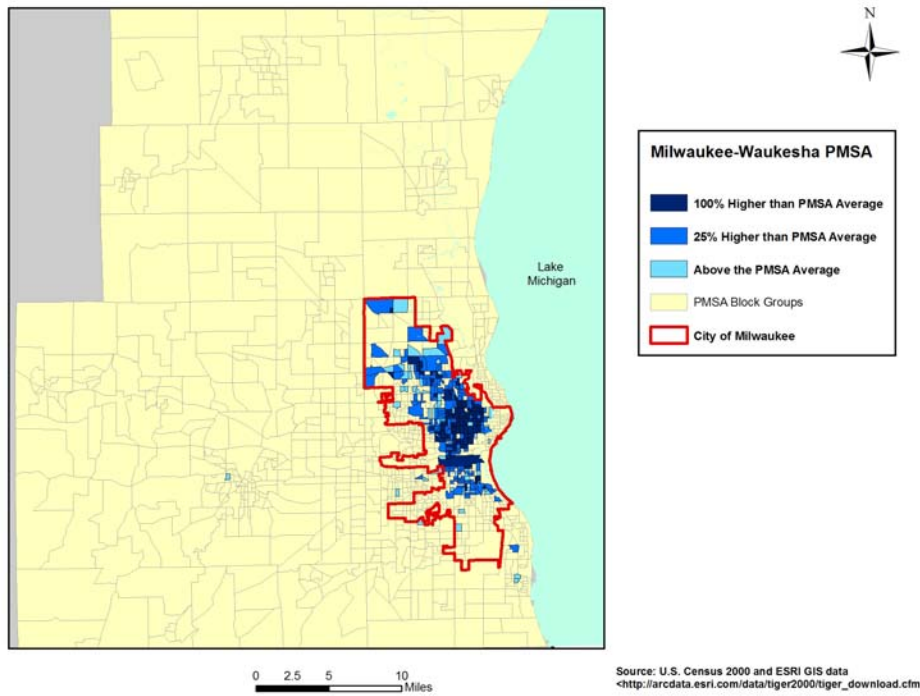


Figure 16. Metropolitan Residential Economic Vulnerability, Indianapolis MSA

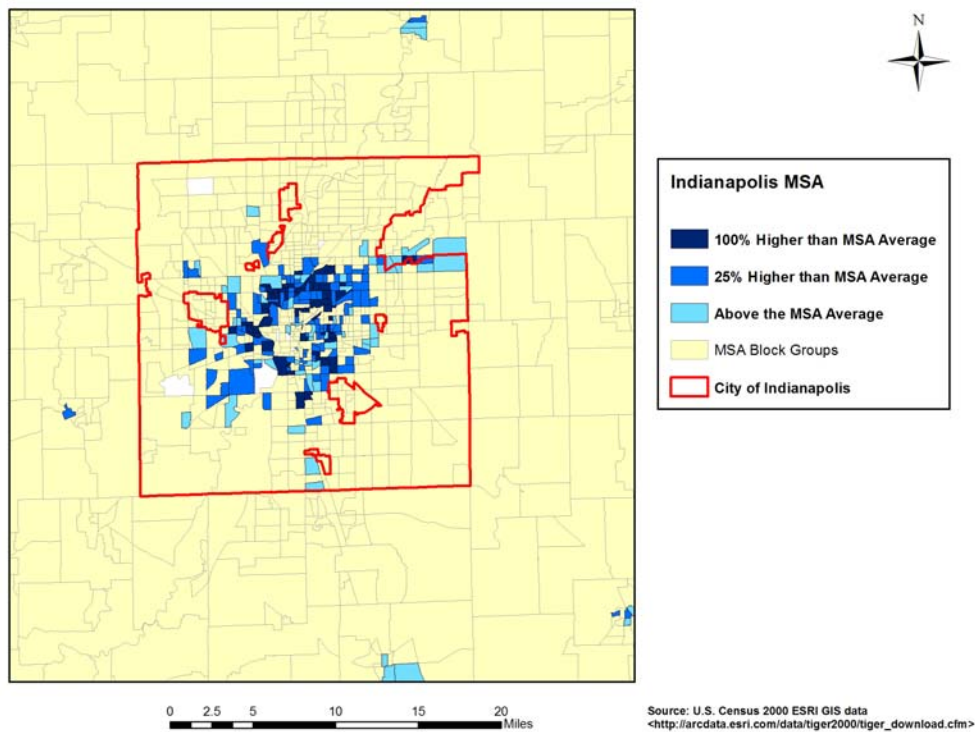
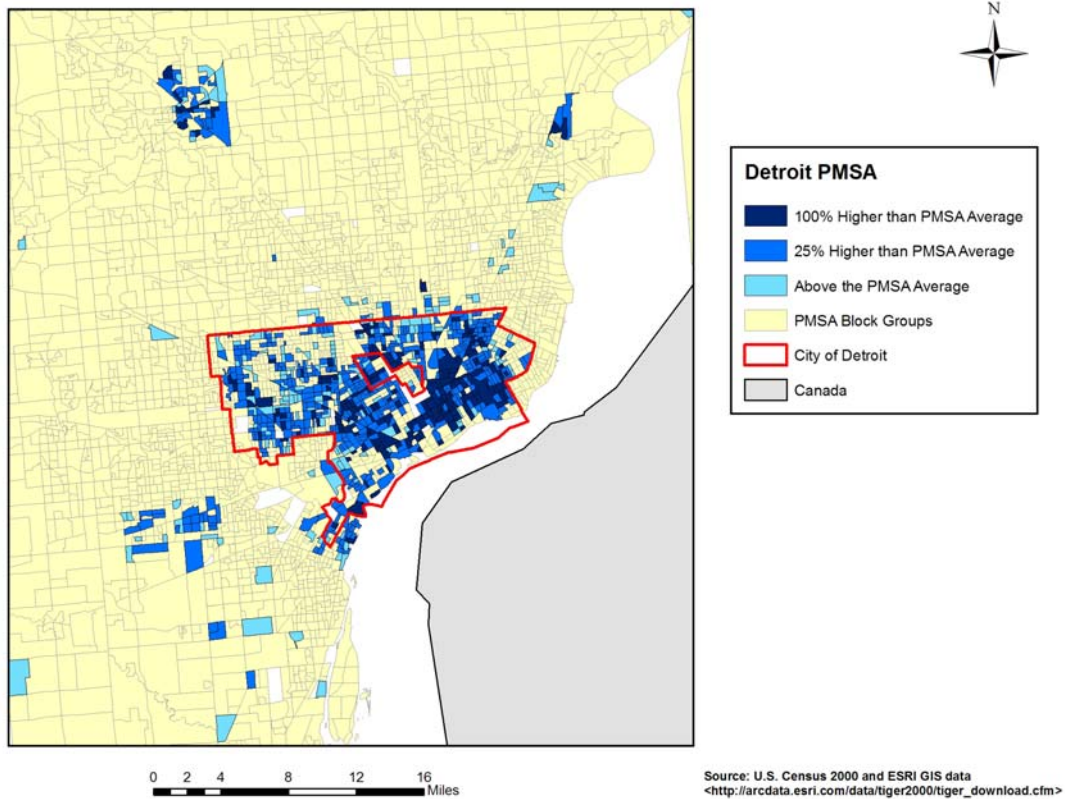


Figure 17. Metropolitan Residential Economic Vulnerability, Detroit PMSA



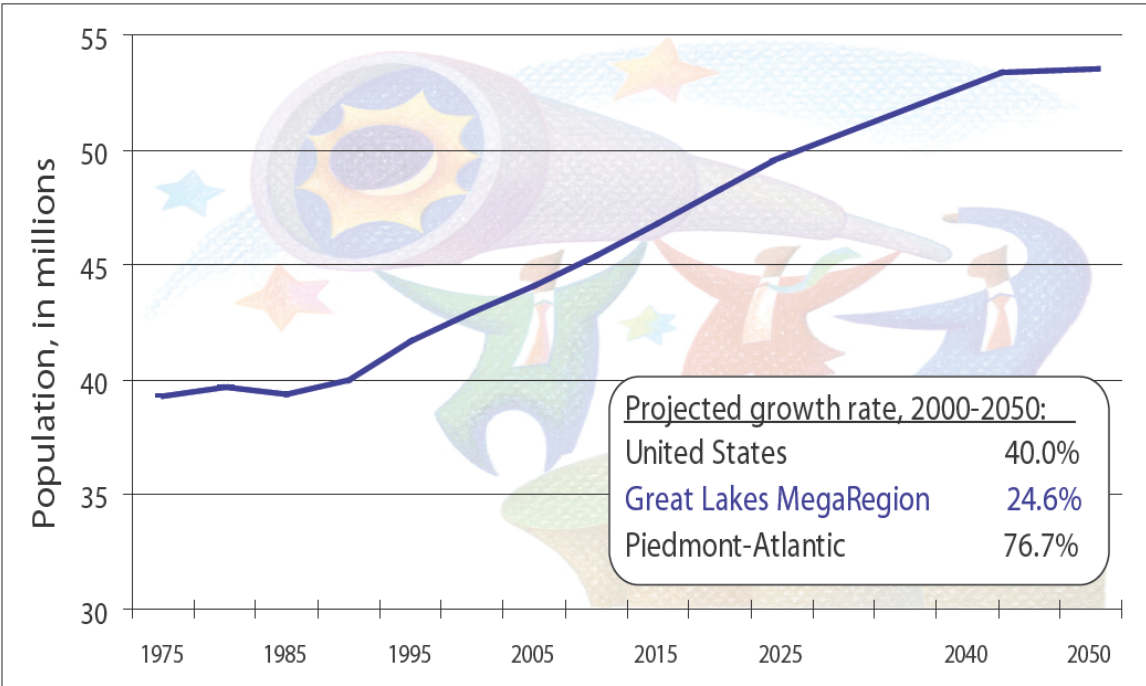
The second primary goal of this analysis was to identify major city clusters of high residential economic vulnerability for prioritizing future urban reinvestment. Figure 15 of the Milwaukee-Waukesha PMSA shows a primary block group cluster of high residential economic vulnerability to the north of the city's center. Figure 17 of the Detroit PMSA shows high vulnerability throughout much of the City of Detroit, especially clustered within the southeast quadrant.

Calculating Population Growth to 2050

The population projections data from Woods and Poole Economics, Inc., were obtained from the Regional Plan Association (RPA). The data from Woods and Poole provide population numbers from 1975- 2000 and population projections by county from 2000 through 2025. Woods and Poole data provide population projections for each year from 2000 through 2010 and for every 5 years 2010-2025.

We selected all the counties in the region and aggregated the population of all counties for the years 2000, 2005, 2010, 2015, 2020, 2025, and 2050 in order to get the population for the MegaRegion for those years. Based on these data points, we constructed a line that represents expected population growth in the region (Figure 17).

Figure 18. Great Lakes MegaRegion Population and Projected Growth, 1975 to 2050



Source: Woods and Poole, Economics, Inc. 2004. "Population and Projections by County, 1975-2000. Obtained as part of the America 2050 project from the Regional Plan Association of New York/New Jersey/Conneticut.

Woods and Poole developed two different projections for the year 2050: 1) the historical growth model and 2) the Woods and Poole growth model.³¹ The historical growth model projected 53.4 million people in the MegaRegion in the year 2050, while the Woods and Poole growth model projected 56.9 million. Based on our understanding of the region's slow population growth and the difficulties of restructuring for a more prosperous economy, we decided to use the historical growth model as the basis for our projections to the year 2050.

³¹ The methods used by Woods and Poole to determine growth patterns were not clear.

Scenarios: Land-Use Change

Introduction

Modeling land use change can provide insight into the ramifications of land use policies and the desirability of different land use scenarios. The Great Lakes MegaRegion planning effort tried two techniques. The first technique, which applied information gathered from an analysis of previous land use change to estimate future land use change, proved difficult to implement. The second technique, driven by per capita land use estimates and based upon David Kooris's work with the Regional Plan Association (RPA), generated the two scenarios in the final plan.³²

The Great Lakes MegaRegion planning effort asserted that a useful model should have the following characteristics:

1. **Stochastic:** The RPA model produced a single outcome for a given set of weights. Since all factors of development cannot be accounted for in a model, several outcomes are possible. Indeed, political issues, regional preferences, and other factors may best be represented by a degree of chance in the final outcome. A stochastic model could be rerun multiple times with the same set of controlling variables to produce a range of development possibilities. Analysis of these outcomes would reveal which scenarios are the most likely.
2. **Iterative:** The *when*, *where*, and *how* of urban development are interrelated. For example the location of previous development has considerable impact upon the location of future development. The model should account for this characteristic of growth by proceeding in steps that become the input of future steps.
3. **Driven by trends:** The model should account for residential preferences reflected in per capita land use. Records show that this figure is increasing from year to year. This criterion especially is based on Kooris's work at RPA. An earlier version of our simulation did not utilize per capita consumption rates but instead increased the amount of urbanized land in a county by the same proportional increase in projected population. This likely resulted in an over-estimation of urbanization in heavily developed areas and underestimation of urbanization in rural areas. It also failed to account for increasing per capita land use in the United States.

³² The Great Lakes MegaRegion project is part of a larger effort by the Regional Plan Association called America 2050. Kooris raised our interest at creating a land use change model at the MegaRegions conference held at the Fundacion Metropoli in Madrid, Spain, from March 6-11, 2006.

4. **Flexible resolution:** The software should not be restricted to a specific resolution. The implementation of the Great Lakes MegaRegion modeling requires that all rasters³³ be the same size, but they can be of any size—provided that adequate computer power is available.
5. **Extendable:** The model should allow skilled users to add new features.
6. **Automated:** Repetitive steps should be performed without the need for user input or supervision.
7. **Educational:** Creating, editing, and using the model should provide information about land use change and land use issues at the MegaRegion scale.
8. **Rationally calibrated:** The controlling variables in the RPA model were determined by visual inference to match notions of how growth would occur. The manner in which development extended along roads from existing urban areas appears to have been a primary concern. While this growth characteristic seems intuitive and correct—more scientific approaches likely exist for building a model. Model output could be compared to a known map to determine necessary adjustments, for example: (1) the model processes an input map for the year 1980 and produces a map for the year 1990; (2) this map is compared to a real land use map for 1990; (3) the model would be adjusted to account for differences. Repeated trials would improve accuracy. The model for the Great Lakes MegaRegion did not meet this criterion due to time constraints and data access problems.

Technique 1: Using Markov Chains to Model Land Use Change

A comparison of urban growth at two different points in time may reveal more accurate weight assignments for development factors such as distance from urbanized land and roads than the more arbitrary method in the RPA model. Markov chains can be used to model land use change by assigning each land use type a percentage chance of turning into every other land use type. Markov processes are designed to model phenomena in which changes of state are only based on the *current* state. For example, using Markov chains implies that the chance for agricultural land to become urbanized land has nothing to do with whether it was under water in the past. The IDRISI software package³⁴ implements tools for using Markov chains, including one that

³³ A raster is a grid (matrix) of numbers displayed as a digital image. Each number in the grid is interpreted as the color of an individual pixel. For the purposes of spatial analysis, each pixel represents a “cell” of a given geographic area. The smaller the geographic area assigned to a single cell, the greater number of cells that are required to model the entire subject area. For example, if the total area of study is 100 square miles, this could be represented by a 10x10 raster with a cell size of 10 square miles and also by a 50x50 raster with a cell size of 2 square miles. The first raster would have 100 cells and the second raster would have 2,500 cells and far more detail—but would require 25 times as much memory and computational time to process. Binary rasters record the existence of a particular phenomenon with a “1” and the absence of the phenomenon with a “0”. For example, a binary raster of urbanized land may assign cells to “1” where land is developed and to “0” where land is not developed. A non-binary urban raster may measure the intensity of urbanization. For example, each cell could be assigned the average height of buildings or percentage of land covered by non-permeable surface, etc.

³⁴ IDRISI is a product of Clark Labs, <http://www.clarklabs.org/>

analyzes two maps to determine the probability for each land use type of turning into every other land use type.

Professor Dan Brown at the School of Natural Resources and the Environment (SNRE) at the University of Michigan made available two datasets for use on this project.³⁵ However, the source maps must be exactly aligned at the pixel level in order to use the IDRISI Markov tools. Different raster cell sizes and possibly also projection³⁶ problems made this accuracy of alignment impossible.

Technique 2: Using Per Capita Land Consumption to Model Land Use Change

Abandoning the previous focus on calibration, this technique duplicated and automated in MATLAB many of the fundamental steps from RPA’s model—with several variations and additions.

PHASE I: Planning, Data Collection, & Pre-processing

Determine units of analysis given project goals, available data, and available computing power. This project used the following units of analysis:

Model Element	Unit of Analysis	Notes
Population projections	County	Source: Woods & Poole
Per capita land consumption rates	State	Urbanization tables from the National Resources Inventory (NRI)
Infill capacity	County (estimated)	The difference between peak county population and current population multiplied by a scaling factor. The scaling factor represents a combination of the usefulness of existing infrastructure and the likeliness of infill based on political realities and personal preferences.
Rasters	One square mile	Source data are available at higher resolution, but the memory (RAM) required to work with multiple high-resolution images is immense. The time required to process such images can also be considerable. For prototyping, lower resolution is preferred.

³⁵ The creation of the data is detailed in the paper, Bergen *et al.* 2005, “Change detection with heterogeneous data using ecoregional stratification, statistical summaries and a land allocation algorithm,” *Remote Sensing of Environment*. Volume 97, pages 434-446.

³⁶ Projections are used to remap a surface on a globe, which is curved, to a flat image for viewing.

1. **Create input rasters at the desired resolution**, using a Geographic Information System (GIS).⁶ Potential input rasters include:

Layer	Known Sources	Notes
Federal lands	TIGER ⁷	See below
Wilderness preserves	TIGER	See below
Water	TIGER and remote sensing ⁸	See below
Existing urban areas ⁹	TIGER (only census defined units: urban clusters & urban areas) and remote sensing	Development pressures decrease with distance from existing urban areas.
Existing road system	TIGER and remote sensing	Development pressures decrease with distance from roads
General land cover	Remote sensing	Marsh and dense forest development are less likely to be developed than open land.
Elevation	USGS	Slope can have a serious impact on land use and can be successfully incorporated into projection models. ¹⁰

⁶ Unless otherwise noted, this model relied on ArcGIS 9.0 to produce source data for the model.

⁷ Topologically Integrated Geographic Encoding and Referencing (TIGER) database of the United States Census Bureau. Available from the USCB in ASCII format at <http://www.census.gov/geo/www/tiger/tiger2k/tgr2000.html> and in shapefile format from ESRI at http://arcdata.esri.com/data/tiger2000/tiger_download.cfm

⁸ In the context of land cover applications, remote sensing involves “sensing” the characteristics of an area through measuring reflected energy in the visible range (photography) or non-visible (infra-red, for instance). This is then interpreted manually or through software to categorize individual cells into specific land categories, such as urban, forest, water. While TIGER data is in a vector (line) format, remote sensing information, such as the USGS and NLCD datasets from <http://www.mrlc.gov/scripts/mapserv.exe?map=d%3A%5Cinetpub%5Cwwwroot%5Clccp%5Cmrlc2k%5Czones%5Czones.map> are in raster format.

⁹ “Urbanized” in this context refers to land used for residential or commercial purposes. Roads and highways are placed on a separate layer that may overlap with urbanized land in places.

¹⁰ Slope proved a major factor in development as presented in the paper, “Dynamic Models of Land Use Change in Northeastern USA” by Mary Tyrrell, Myrna Hall and R. Neil Sampson available from Yale University at http://research.yale.edu/gis/ppf/dynamic_models/index.html

2. **Identify areas completely restricted from development.** These areas will be directly coded into the script as off-limits to urban growth.

Land cover	Notes & Significance
Water	Land area cannot extend over time into the water (land reclamation)
Existing Urban	Infill is discounted or computed by reducing total outward urbanization
Federal land and wildlife preserves	Laws restricting development will hold even if the surrounding area is built out
Roads & Highways	Cell area is small enough that roads are a single cell in width and cannot be turned into urbanized land. For large cell sizes, roads may be recorded as a mile wide, which is ridiculous, but often necessary when dealing with rasters. Depending on implementation, if the urban layer shows no development in such road cells, then perhaps development should be permitted. The negative aesthetic impact could be reduced by always placing the road layer on top.

PHASE II: These steps are computed as part of two internal loops. The outer loop moves from year to year and the inner loop moves through each unit of analysis (county).

3. **Generate propensity value or impedance rasters for the various input rasters.** These terms refer to the same concept from two different points of view. Higher propensity values indicate that development is more likely to occur. Higher impedance values indicate that resistance to development is higher, thus less likely to occur. Depending on implementation, very high impedances (low propensities) may have the same effect as hard-coding areas as not available for development. For example, a cell with a “1” in the federal land raster may be assigned to 1000 in the impedance raster and a “0” may be assigned to a 0 or 1. Cells demarked as treeless grasslands in a land cover raster may be assigned an impedance of 1 and tree-covered areas may be assigned an impedance of 10 to indicate the additional cost of cutting down the trees. Often distance is a factor in these calculations. For example, development pressure is lower as distance from urbanized land and roads increases. Since urban land is increasing at every time step (year), these distance rasters must be recomputed every year.
4. **Calculate relative weights of factors and generate the aggregate raster.** The values in each raster generated in step 4 are weighted to represent their relative impact on development. The weights assigned depend on the range of numerical values in the rasters and knowledge of the urbanization process. Assume that both the rasters for distance from roads and distance from urban areas were computed with an inverse square function and normalized to be between 0 and 1. Then, the urban areas raster should be assigned a larger weight in a model intended to create a more compact urban form and

the distance from roads raster assigned a larger weight in a model intended to simulate sprawl. For simplicity, the relative weights are assumed not to vary in either time or space and can be set before the loop. Only the actual raster aggregation needs to occur inside the loop to account for land developed in the previous time step. Realistically, knocking down trees may be more expensive financially or politically in different counties and at different times.

5. **Determine population growth.** Calculate the number of new people moving into each population projection unit. Straight-line projection between two points in the source data was used for this project. Fitting a curve to the source data and then reading the desired point from the curve is likely more accurate.
6. **Determine population requiring new development.** Counties can have an absorptive capacity based on previous peak population levels. New development is only necessary when this capacity has been surpassed.
7. **Determine per capita land consumption rate.** This will vary from time step to time step and from state to state (or whatever unit of analysis is used).
8. **Determine number of cells of new development.** This is equal to the number of people requiring development multiplied by the per capita land consumption rate (measured in raster cells).
9. **Determine land actually available for development.** This requires subtracting the number of cells in a county that are off-limits to development from the total number of cells in the county—or an equivalent process.
10. **Redistribute demand for land that surpasses land available in the unit of analysis (county).** If the total amount of land available is insufficient to meet the demand for land, then excess demand should be redistributed to adjacent counties or to some range of nearby counties.
11. **Randomly select a cell of available land based on propensity values/impedances.** For example, given that 20 available cells exist and propensity to urbanize is a range from 0 to 1 as in the table below:

# of cells	Value	Chance that an individual cell will be selected for urbanization
5	.001	$.001/5.405 = .00018501$
10	.25	$.25/5.405 = .0463$
4	.5	$.5/5.405 = .0925$
1	.9	$.9/5.405 = .1665$
Total: 20	5.4050	Chance * #cells = 1

Continue to recompute such a table of available cells and their chance to urbanize and to select a cell to urbanize until the final demand calculated in step 11 is satisfied.

12. **Implement a neighborhood analysis for more compact growth.** Non-contiguous growth—called “leap frogging”—still occurred occasionally regardless of the weights assigned to the rasters. This phenomenon could simply reflect the impact of chance or a possible error in the implementation of the model. Instead of summing the weights of rasters, the following steps were taken for the compact growth scenario:
 - a. Create a list of pixels available for urbanization as before.
 - b. For each pixel in the list, calculate the total number of neighboring pixels that are urbanized, resulting in a value from 0 to 8 (north, south, east, west, northeast, northwest, southeast, and southwest).
 - c. Create a subgroup of pixels that have the highest number of urban neighbors.
 - d. If that subgroup has more than one member, select a member at random.
 - e. Urbanize the selected pixel.

13. **Continue for all counties and all time steps.** Since the simulation likely can only compute a single county at a time for each time step, the counties should be processed in a random order for every time step to feign simultaneity.

14. **Analyze results, change propensities and raster weights, and rerun.** This iterative process is necessary to arrive at a reasonable aggregate raster. Without this step, even simulations that could potentially produce useful projections—are little more than algorithmic art machines. This step can likely only be accomplished by using historical data to estimate other historical data and adjusting for differences. This does not necessarily mean the model will be able to tell the future reliably. Since this step was omitted due to time constraints and data availability, modeled outcomes should be considered very rough estimates. The amount of development may be more accurate than the location of development.

15. **Load output into GIS.** A raster generated in MATLAB will have a value of “1” for urban land and a value of “0” for non-urban. This can be saved as a TIFF image, georeferenced¹¹ and re-integrated with existing data layers in GIS. Since the extent of the final image is identical to the extent of all the source rasters, the coordinates are known and can be typed into a supplemental text file. In ArcGIS, such a “world file” must have the same name as the image and a .tfw extension. An example for a raster of cell size of one mile (5280 feet) follows with the last two values representing the specific extent of the subject area.

```
5280
0
0
-5280
23641302.9
1694277.85
```

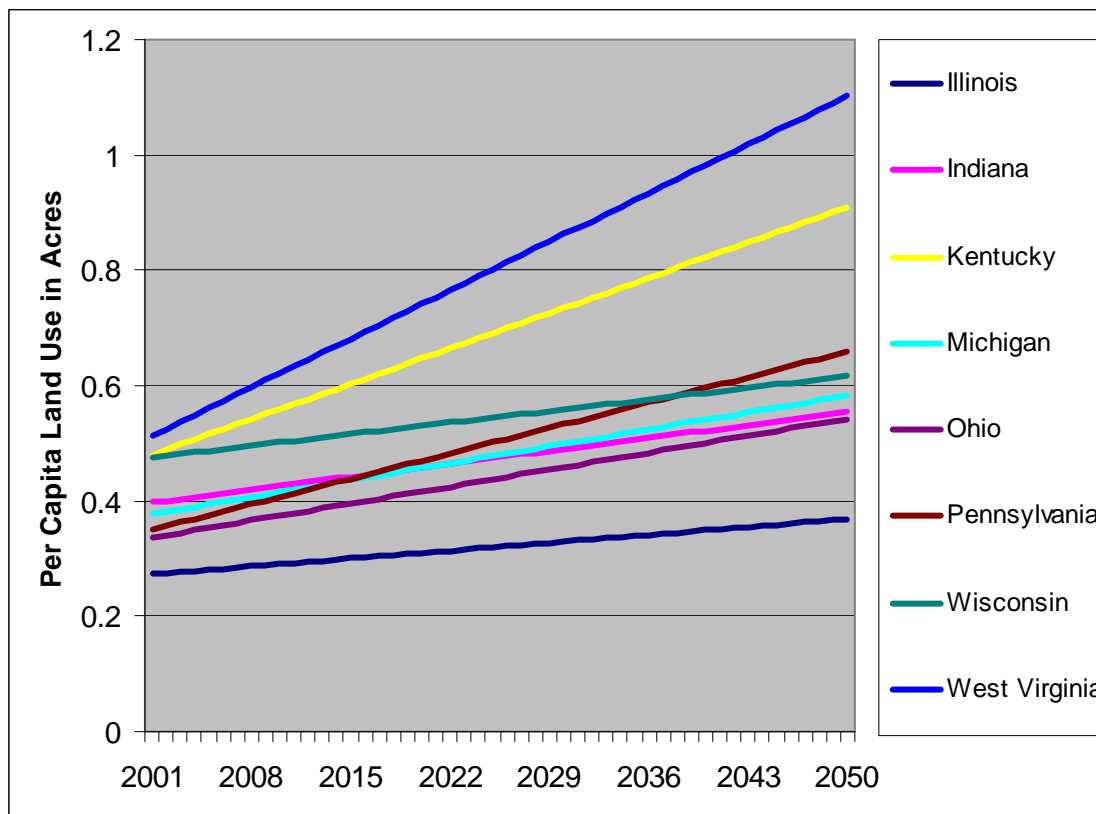
¹¹ Georeferencing is the process of adding spatial information to a raster so that it lines up properly without spatially explicit data.

Parameters for Results Reported in the Plan

Sprawl development

Aggregate propensity map	=	inverse distance squared from roads + inverse distance squared from urban
County infill (absorption) capacity	=	0
Population growth	=	linear interpolation of Woods & Poole data
Per capita land use	=	Please see Figure 1

Figure 19: Future Land Consumption Based on Historic Trends

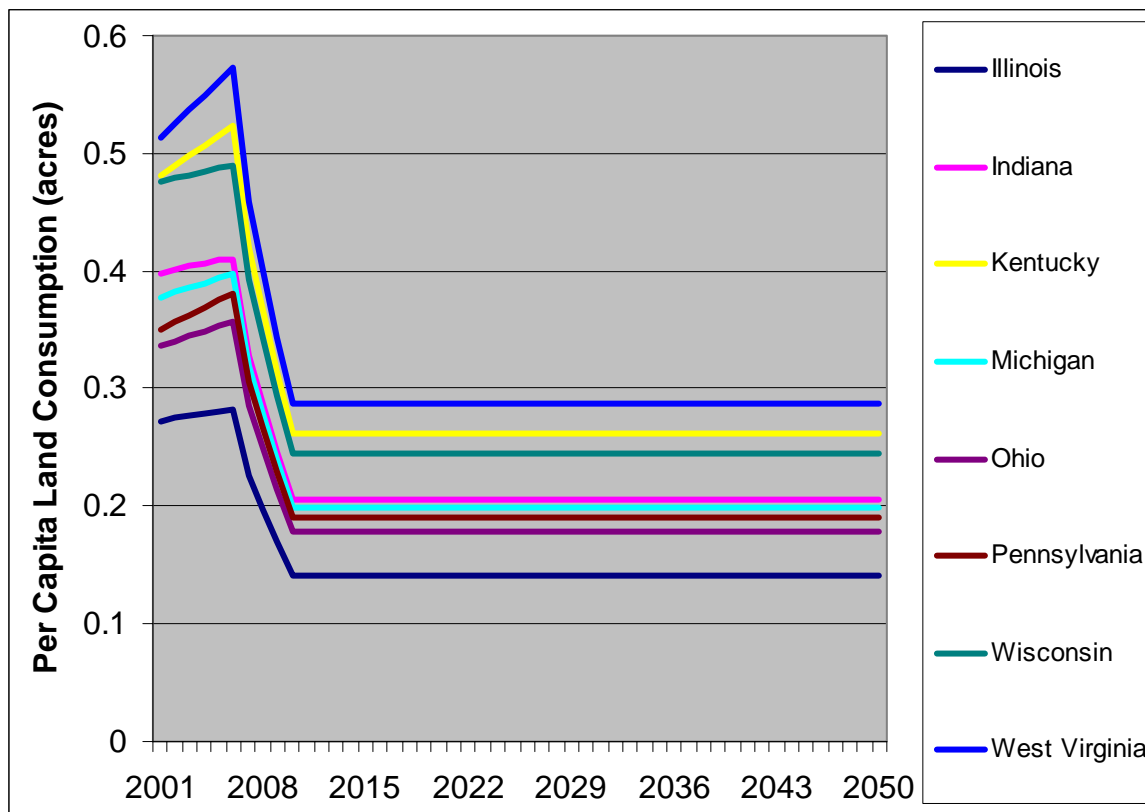


Source: Least squared error linear fit of state urbanized land divided by state population for 1982, 1987, 1992, and 1997, extended to 2050. Urbanized land area from US Department of Agriculture, Natural Resources Conservation Service available at <http://www.nrcs.usda.gov/TECHNICAL/land/meta/t5846.html>. State population data from US Census Bureau Statistical Abstracts available at <http://www.census.gov/prod/www/abs/statab1951-1994.htm>. Both last accessed April 28, 2006.

Compact development

- Aggregate propensity map = replaced with neighborhood function described above in step 13.
- County infill (absorption) capacity = $(\text{peak population} - \text{current population}) * .75$
- Population growth = linear interpolation of Woods & Poole data
- Per capita land use = Please see Figure 2

Figure 20: Future Land Consumption Adjusted for Growth Control Policies

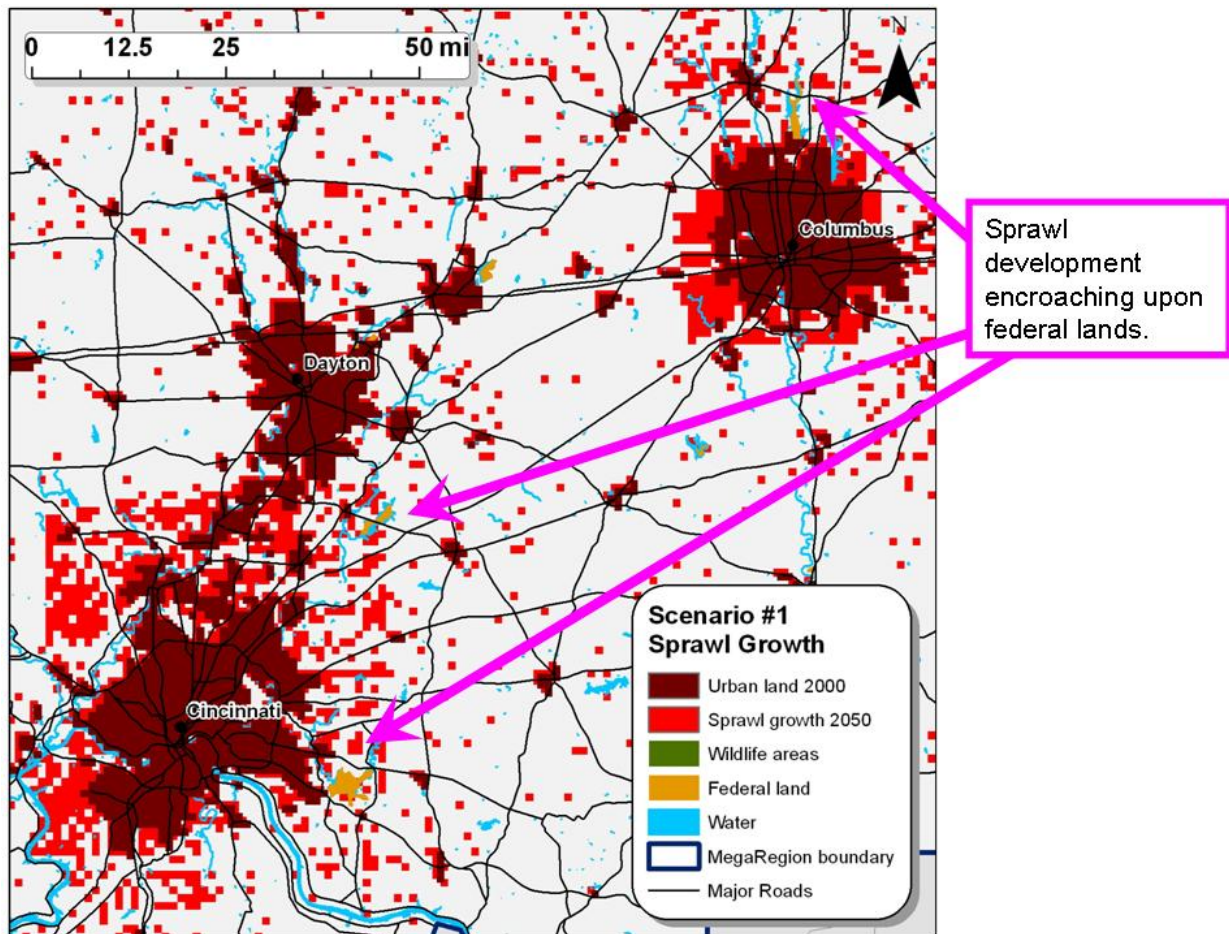


Source: Least squared error linear fit of state urbanized land divided by state population for 1982, 1987, 1992, and 1997, extended to 2006. An (arbitrary) drop to 60% of 2006 levels represents the (anticipated) effect of growth control policies on land consumption. Urbanized land area from US Department of Agriculture, Natural Resources Conservation Service available at <http://www.nrcs.usda.gov/TECHNICAL/land/meta/t5846.html>. State population data from US Census Bureau Statistical Abstracts available at <http://www.census.gov/prod/www/abs/statab1951-1994.htm>. Both last accessed April 28, 2006.

Commentary

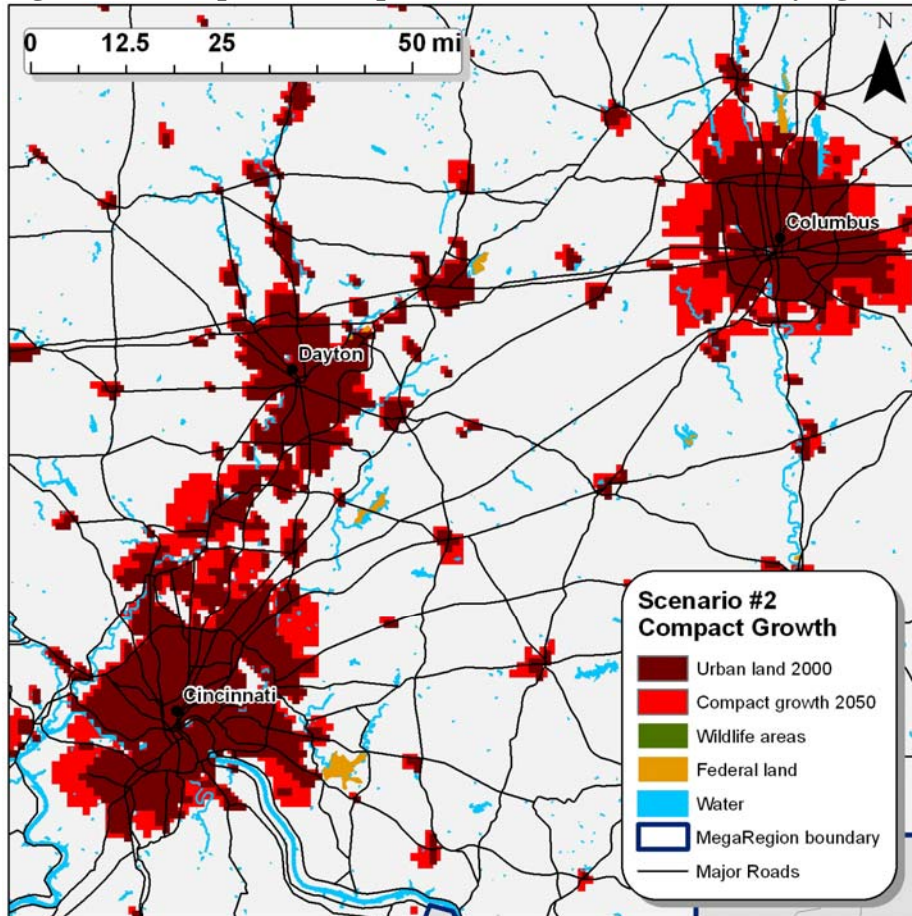
Rendering close-ups of development shows how compact growth reduces the threat to natural areas. While the code explicitly forbids urbanizing federal land and wildlife areas, compact growth also results in less encroachment.

Figure 21: Sprawl Encroaches on Sensitive Areas



Source: Custom software simulation using Census 2000 shapefiles, Woods & Poole county-level population projections, US Department of Agriculture, Natural Resources Conservation Service, US Census Bureau Statistical Abstracts.

Figure 22: Compact Development Reduces Threat to Outlying Ecosystems



Source: Custom software simulation using Census 2000 shapefiles, Woods & Poole county-level population projections, US Department of Agriculture, Natural Resources Conservation Service, US Census Bureau Statistical Abstracts.

Future Direction of Research

This effort is at a very early stage of development. MATLAB provides flexibility but few built in tools specifically for use with modeling land use change. The next step is likely to look more carefully at other general-purpose programming languages and also products specifically geared toward this work.

Scenarios: Land-Use Change MATLAB Script (Sprawl)

```
%LAND USE CHANGE MODEL
```

```
%Great Lakes MegaRegion Plan  
%University of Michigan  
%Urban and Regional Planning  
%April 2006
```

```
%This is a copy of the code that produced the land use change results  
%included in the MegaRegion plan. However, the code is still in an early stage of  
%development and includes some errors. It should eventually be written as a  
%set of functions that can be called from a simpler, shorter script. (do you want to state that lines  
beginning with % are comment lines?)
```

```
clear all;  
close all;
```

```
%Several logs are used to track the modeling process
```

```
genlog = fopen('LOG_general_20.txt','w');  
builtlog = fopen('LOG_builtout_20.txt','w');  
neighlog = fopen('LOG_neighflow_20.txt','w');  
errorlog = fopen('LOG_error_20.txt','w');  
simplog = fopen('LOG_simple_20.txt','w');
```

```
%Load PerCapita Landuse data and create projection curve
```

```
sprintf('Loading state-level per capita landuse data and projecting.')  
data=load('PerCapita.csv');  
PerCapYr=data(1,2:end);           %first row is years associated with per capita, except for 1st  
column which should be skipped  
PerCapSt=data(2:end,1);           %1st column is FIPS code for states  
PerCapData=data(2:end,2:end);     %other rows are data
```

```
StartYr=2001;  
EndYr=2050;  
Incr=1;
```

Land Use Change – MATLAB Scripts (Sprawl)
The Great Lakes MegaRegion

```
NumEst=length(PerCapYr);
NumStates=length(PerCapSt);
EstYr=(2001:1:2050);

for i=1:NumStates
    [p,S]=polyfit(PerCapYr,PerCapData(i,:),1); %fit to straight line, least squared error
    PerCapEst(i,:)=polyval(p,EstYr);
    St=PerCapSt(i);
end

%Load Population Data and create projection curve

sprintf('Loading county-level population estimates and projecting.')
data=load('wp.csv');
PopYr=data(1,3:end); %first two values are column headers for FIPS & Area
PopData=data(2:end,3:end); %first two columns are FIPS,AREA. First row are YEARS--not
data
cFIPS=data(2:end,1); %first entry is column header
cArea=data(2:end,2);
cN=length(cFIPS);
PopData=PopData.*1000; %original data is in thousandths

EstYr=(2001:1:2050);
for i=1:cN
    PopEst(i,:)=interp1(PopYr,PopData(i,:),EstYr,'linear');
end

%FIND PEAK POPULATION
sprintf('Calculating historic peak levels in all counties')
for i=1:cN
    pmax(i)=max(PopData(i,1:6));
end

%Set absorption rate

sprintf('Calculating absorption capacity (infill) for each county')
AbsorbRatio=.5 %percentage of difference between max historic population
%that can be absorbed into the county without the need for
%outward development. The data goes back to 1975

for i=1:cN
    p2000(i,1)=PopData(i,6);
    if pmax(i)<p2000(i)
        PopAbsorb(i,1)=0; %if yr2000 population is the largest, then no absorption possible
```

```
else
    PopAbsorb(i,1)=(pmax(i)-p2000(i))*AbsorbRatio;
end
fprintf(genlog,'county %d can absorb %d more people.\n',cFIPS(i),PopAbsorb(i));
end
```

```
%OPTIONAL Override Absorption for Sprawl
%PopAbsorb(:,1)=0;
```

```
%Load in rasters and other support files
```

```
sprintf('Loading rasters and other support files')
file='counties.tif';
cts=imread(file);
```

```
file='countyboxes.csv';
data=load(file);
cCB=data(:,1); %FIPS from county boundary boxes
ULx=data(:,2);
ULy=data(:,3);
LRx=data(:,4);
LRy=data(:,5);
```

```
file='urban.tif';
urban=imread(file);
urban=urban==1;
urban=dither(urban);
%since urban distance changes with development it is computed within the
%loop
```

```
file='roads.tif';
roads=imread(file);
roads=roads==1;
roads=dither(roads);
roadDI=bwdist(roads); %distance from roads
INVroadDI=1./(roadDI+.1).^2;
```

```
file='wild.tif';
wild=imread(file);
wild=wild==1;
wild=dither(wild);
```

```
file='fed.tif';
fed=imread(file);
```

Land Use Change – MATLAB Scripts (Sprawl)
The Great Lakes MegaRegion

```
fed=fed==1;  
fed=dither(fed);
```

```
file='water.tif';  
water=imread(file);  
water=water==1;  
water=dither(water);
```

```
nogrow=dither(urban+roads+wild+fed+water);
```

```
%LOOK FOR NEIGHBORING COUNTIES
```

```
%county FIPS numbers are recorded as "colors" in the county file.
```

```
%This looks at all pixels around the county and records up to 10
```

```
%neighboring counties FIPS NUMBERS
```

```
sprintf('Calculating neighboring counties for each county')
```

```
neighFIPS(1:cN,1:10)=0;
```

```
for c=1:cN
```

```
    temp=cts(ULy(c)-1:LRy(c)+1,ULx(c)-1:LRx(c)+1);
```

```
    q=unique(temp);
```

```
    if ismember(0,q)
```

```
        q=setxor(0,q);
```

```
    end
```

```
    if ismember(cFIPS(c),q)
```

```
        q=setxor(cFIPS(c),q);
```

```
    end
```

```
    if length(q) > 10
```

```
        q=q(1:10);
```

```
    end
```

```
    if length(q) < 10
```

```
        z=zeros(1,10-length(q));
```

```
        q=[q z];
```

```
    end
```

```
    neighFIPS(c,:)=q;
```

```
end
```

```
%Record Initial Available Space. This step should be coded so that it can be placed within the  
main loop
```

```
%and computed as part of the simulation.
```

```
sprintf('Recording initial space available...');
```

```
okdev=(dither(urban+roads+wild+fed+water)==0);
for i=1:cN
    Avail=(cts==cFIPS(i)).*okdev;
    [AvailR,AvailC]=find(Avail==1);
    InitialAvail(i,1)=length(AvailR);
end

%START MODELING LOOP

unaccounted=0;          %tracks people not accounted for due to time limits, etc

NowPopVect=p2000;
RealPop=PopEst;
NewDevBuffer(1:cN,1)=0;
PerCapConv=640;        %there are 640 acres to the square mile. Per Capita data is in acres
                       %grid pixels are each one square mile

for i=1:EndYr-StartYr+1
    counter=0;
    sprintf('\n\n\n\nSTARTING ESTIMATES FOR YEAR: %d',i)
    urbanDI=bwdist(urban);      %distance from urban recomputed every time step
    INVurbanDI=1./(urbanDI+1).^2;
    propensity=INVurbanDI+INVroadDI; %optionally remove roads for compact development
    calculation
    randcty=randperm(cN); %create random county order
    for ctty=152 %randcty
        counter=counter+1;
        if round(counter/25)==counter/25
            sprintf('processed: %d of %d counties',counter,cN)
        end
        NewDev=0; %inititalize development
        NowFIPS=cFIPS(ctty);
        NowPop=NowPopVect(ctty); %this gives current population starting with 2000
        FutPop=RealPop(ctty,i); %this gives the puplation one step into the future population,
                                %adjusted by run over people. It does not use the direct W&P
                                %estimates because they do not include
                                %the run over people.
        AddPop=FutPop-NowPop;
        State=round(cFIPS(ctty)/1000-.5);
        StINDEX=find(PerCapSt==State);
        PerCap=PerCapEst(StINDEX,i);

        fprintf(simplog,'%d %d %d %d %d\n',i+StartYr-1,ctty,NowPop,FutPop,AddPop);
        fprintf(genlog,'YEAR: %d COUNTY: %d NEW PEOPLE: %d ACRES/PERSON:
%d\n',i+StartYr-1,ctty,AddPop,PerCap);
```

```
    if AddPop > 0
        if PopAbsorb(cty)>=AddPop
            fprintf(genlog,'COUNTY %d NEW POPULATION ENTIRELY ABSORBED:
starting infill space=%d people\n',cty,PopAbsorb(cty));
            PopAbsorb(cty)=PopAbsorb(cty)-AddPop;
            RealPop(cty,i)=NowPop+AddPop;
            AddPop=0;
            fprintf(genlog,'                                ending infill space=%d
people\n',PopAbsorb(cty));
        elseif PopAbsorb(cty)>0
            fprintf(genlog,'COUNTY %d NEW POPULATION PARTIALLY ABSORBED:
starting infill space=%d people\n',cty,PopAbsorb(cty));
            AddPop=AddPop-PopAbsorb(cty);
            PopAbsorb(cty)=0;
            NewDev=AddPop*PerCap;
            fprintf(genlog,'NEW DEVELOPMENT=%d acres. REMAINING INFILL
SPACE=%d people\n',NewDev,PopAbsorb(cty));
        else
            NewDev=AddPop*PerCap;
            fprintf(genlog,'COUNY %d: NO INFILL SPACE! %d acres of NEW
DEVELOPMENT NEEDED!\n',cty,NewDev);
        end
        if NewDev>0
            NewDevBuffer(cty)=NewDevBuffer(cty)+NewDev;
            NumNewPixels=round(NewDevBuffer(cty)/PerCapConv-.5);
            NewDevBuffer(cty)=NewDevBuffer(cty)-NumNewPixels*PerCapConv;
            fprintf(genlog,'County %d: %d new pixels of growth and %d acres in
buffer\n',cty,NumNewPixels,NewDevBuffer(cty));

            if NumNewPixels>0
                % first determine amount of space available for development
                okdev=(dither(urban+roads+wild+fed+water)==0);
                Avail=(cts==NowFIPS).*okdev;
                [AvailR,AvailC]=find(Avail==1);
                NumAvail=length(AvailR);
                fprintf(genlog,'Pixels need: %d. Pixels available: %d
pixels.\n',NumNewPixels,NumAvail);
                % check if some or all of added population must be
                % relocated due to county being built out
                if NumNewPixels>=NumAvail

                    % This next short conditional statement was added to
                    % record when a county becomes built-out. However,
                    % sometimes it fails to function correctly. The best
```

```
%source for the same data is the neighborhood flow file,  
%updated whenever people are redistributed to  
%neighboring counties when one has become built  
%out.  
  
if NumAvail>0  
    fprintf(builtlog,'Year: %d County: %d\n',i+StartYr-1,cty); %keep log of when  
counties are first built out  
end  
  
RunOverPixels=NumNewPixels-NumAvail;  
NumNewPixels=NumAvail;  
RunOverPeople=(RunOverPixels*PerCapConv)+NewDevBuffer(cty);  
NewDevBuffer(cty)=0;  
AddPop=NumNewPixels*PerCapConv;  
RealPop(cty,i)=NowPop+AddPop;  
%now that true number of people to be added has been determined update  
population for following year  
fprintf(genlog,'AREA IS BUILT OUT! Adding only: %d people and  
redistributing %d people\n',AddPop,RunOverPeople);  
  
%Relocated run over people  
neighINDEX=0;  
if ismember(0,neighFIPS(cty,:))  
    neigh=setxor(neighFIPS(cty,:),0);  
else  
    neigh=neighFIPS(cty,:);  
end  
NUMneigh=length(neigh);  
PerNeigh=RunOverPeople/NUMneigh; %divide run over people evenly between  
all neighbors  
for q=1:NUMneigh  
    neighINDEX(q)=find(cFIPS==neigh(q));  
end  
for index=neighINDEX  
    if i<50  
        fprintf(neighlog,'year %d from %d to %d flow %d\n',i+StartYr-  
1,NowFIPS,cFIPS(index),PerNeigh);  
        RealPop(index,i+1)=RealPop(index,i+1)+PerNeigh;  
    else  
        fprintf(neighlog,'year %d from %d to %d flow %d\n',i+StartYr-  
1,NowFIPS,cFIPS(index),PerNeigh);  
        unaccounted=unaccounted+PerNeigh;  
        fprintf(errorlog,'year %d unaccounted total %d\n',i+StartYr-1,unaccounted);  
    end  
end
```

```

        end
    else
        %if county has space then full population is added
        RealPop(cty,i)=NowPop+AddPop;
    end
    for nthpixel=1:NumNewPixels
        fprintf(genlog,'searching for location for pixel #d\n',nthpixel);
        propensityTOT=sum(propensity(AvailR,AvailC));
        chance=propensity(AvailR,AvailC)/propensityTOT;
        select=rand;
        propsum=0;
        count=0;
        sprintf('random select number=%d. Total Pixels Avail: %d',select,length(AvailR))
        while propsum<select          %select random pixel according to propensities.
this part of the code needs to be double checked.
            count=count+1;
            propsum=propsum+chance(count);
            %sprintf('index %d total %d',count,propsum)
        end
        sprintf('selected available vector entry #d: pixel %d x %d to
urbanize',count,AvailR(count),AvailC(count))
        sprintf('max chance=%d. mean chance=%d median chance=%d CHANCE OF
PIXEL=%d',max(chance),mean(chance),median(chance),chance(count))
        urban(AvailR(count),AvailC(count))=1;
        Avail(AvailR(count),AvailC(count))=0; %remove selected pixel from available ones
        [AvailR,AvailC]=find(Avail==1);
        NumAvail=length(AvailR);
    end
end
end
else
    fprintf(genlog,'POPULATION LOSS...nothing to do.\n');
end
end
NowPopVect=RealPop(:,i);    %make the current population vector equal to the current
population after the full phase of development
name=sprintf('PerCapModel_Year%d.tif',i+StartYr-1);
imwrite(urban,name,'TIFF');
end

%Record Final Available Space
sprintf('Recording initial space available...');
okdev=(dither(urban+roads+wild+fed+water)==0);
for i=1:cN
    Avail=(cts==cFIPS(i)).*okdev;

```



```
[AvailR,AvailC]=find(Avail==1);  
FinalAvail(i,1)=length(AvailR);  
end
```

```
%record useful statistics about the run of the simulation.  
stats=[cFIPS InitialAvail FinalAvail NewDevBuffer p2000 RealPop(:,50)];  
csvwrite('SimpleStats_20.csv',stats);  
csvwrite('FullStats_20.csv',RealPop);
```

```
fclose(genlog);  
fclose(neighlog);  
fclose(builtlog);  
fclose(errorlog);  
fclose(simplog);
```


Scenarios: Land-Use Change MATLAB Script (Compact Growth)

% This section replaces the section of the loop that selects a pixel based on the propensity map.

```
for nthpixel=1:NumNewPixels
    % fprintf(genlog,'searching for location for pixel #%d\n',nthpixel);
    pointerlist=randperm(length(AvailR));
    clear tally
    % fprintf(genlog,'random select number=%d. Total Pixels Avail:
%d\n',select,length(AvailR));

    for ct=pointerlist
        Eneigh=urban(AvailR(ct),AvailC(ct)+1);
        Wneigh=urban(AvailR(ct),AvailC(ct)-1);
        Nneigh=urban(AvailR(ct)-1,AvailC(ct));
        Sneigh=urban(AvailR(ct)+1,AvailC(ct));
        NEneigh=urban(AvailR(ct)-1,AvailC(ct)+1);
        NWneigh=urban(AvailR(ct)-1,AvailC(ct)-1);
        SEneigh=urban(AvailR(ct)+1,AvailC(ct)+1);
        SWneigh=urban(AvailR(ct)+1,AvailC(ct)-1);

        tally(ct,1)=Eneigh+Wneigh+Nneigh+Sneigh+NEneigh+NWneigh+SEneigh+SWneigh;
    end

    tallymax=max(tally);
    tallyindex=find(tally==tallymax);
    highrand=randperm(length(tallyindex));
    choice=tallyindex(highrand(1));

    % sprintf('index %d total %d',count,propsum)

    % fprintf(genlog,'selected available vector entry #%d: pixel %d x %d to
urbanize\n',count,AvailR(count),AvailC(count));
    % fprintf(genlog,'max chance=%d. mean chance=%d median chance=%d CHANCE
OF PIXEL=%d\n',max(chance),mean(chance),median(chance),chance(count));
    urban(AvailR(choice),AvailC(choice))=1;
    Avail(AvailR(choice),AvailC(choice))=0; % remove selected pixel from available
ones

    [AvailR,AvailC]=find(Avail==1);
    NumAvail=length(AvailR);
end
```