Potential Cluster Regions: The Case of the Floriculture Industry*

Bruce W. Smith
Center for Regional Development & Department of Geography
Bowling Green State University
Bowling Green, OH 43403

Michael C. Carroll
Center for Regional Development & Department of Economics
Bowling Green State University
Bowling Green, OH 43403

Neil Reid
Urban Affairs Center & Department of Geography and Planning
The University of Toledo
Toledo, Ohio 43606

* Acknowledgement: portions of this paper are based on work supported by the Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under award Nos. CSREES 2003-06230, CSREES 2004-06222, CSREES 2005-02216

Abstract. Most cluster-based economic development programs use co-location to identify cluster areas. Geographic proximity (co-location) is a necessary, but not a sufficient, condition for potential clustering activity. Therefore, assessment of industry location and density patterns becomes the first phase in the identification of potential cluster regions to be included in a cluster driven development policy. This paper uses Getis-Ord Gi* in the identification of potential cluster regions in the U.S floriculture industry. Issues related to the use of alternative spatial weights matrices are reviewed. Also the value of identifying nationwide floriculture potential cluster regions from the perspective of a northwest Ohio greenhouse cluster project is discussed.

Key Words: clusters, spatial statistics, floriculture

1. INTRODUCTION

Cluster-based economic development (CBED) as an alternative economic development strategy has received much attention. This concept has been examined in the academic literature by researchers, the most prominent being Michael Porter (1998). Also it has gained acceptance among practitioners. Akundi (2003) found 40 states to be involved in cluster economic development and the Cluster Initiative Greenbook (Solvell et al., 2003) lists more than 250 projects world wide. Even though the ideas of CBED have become better known, confusion and misunderstanding reigns in terms of cluster definitions, appropriate cluster identification methodologies, and the like. For example, Martin and Sunley (2003) characterized clusters as being popular, but problematic.

Researchers have used various methods to identify existing clusters. Shift-share analyses and location quotients have been used to delineate spatial concentrations (Miller et al., 2001; Hendry and Brown, 2006). Other approaches have incorporated expert opinions in the process (Roberts and Stimson, 1998). Statistical analyses of input-output tables to gauge the interdependences in the regional economy as the basis for empirically deriving clusters have been another strategy. For example, Hill and Brennan (2000) used cluster analysis to identify groups of similar industries in terms of their competitiveness, interindustry linkages, and export characteristics. They then applied discriminant analysis to identify the driver industries. As another example, Feser and Bergman (2000) used principal components analysis on the 1987 U.S. input-output accounts to derive 23 industrial clusters, which they believe can be used as templates in subsequent regional analyses to develop a strategic view of a regional manufacturing economy. Also, clusters have been delineated with spatial statistics (Helsel et al., 2007).

In this paper, the focus is only on identifying the "spatial footprint" of potential clusters regions (PCR). PCR's are areas that potentially can support clustering activities because they contain the necessary concentration of firms in the industry or supply chain. PCR builds on the notion that spatial concentration is a necessary, but not sufficient, condition underlying CBED policy. However, as argued elsewhere, a CBED is a network driven economic strategy that stresses communication between firms in the core industry, local suppliers, local government and support institutions such as universities, think tanks and development agencies (Reid and Carroll, 2007). Consequently, a PCR only has the potential to be a cluster due to the co-location criterion. From this perspective, the examination of industry location patterns to delineate PCR's could be the initial step in a CBED. Elimination of areas without sufficient concentrations of firms will reduce the likelihood of failed cluster projects due to the lack of critical mass.

The purpose of this paper is to use Getis-Ord's Gi* to delineate PCR's in the floriculture industry. In addition, the impact of alternative spatial weights matrices, which are integral to these methodologies, will be examined. The floriculture (NAICS 111422) industry was selected because it is the focus of the authors' ongoing cluster-based economic development project, known as Maumee Valley Growers. The project was initiated in 2003 and assists greenhouse growers in northwestern Ohio cope with a variety of competitive challenges. The original cluster was composed of growers in northwestern Ohio, but there has been recent discussion about expanding the cluster to other areas such as Detroit and southeast Michigan. Moreover, it is desirable to identify clusters across the nation to monitor from the context of competitive trends. Thus, the identification of floriculture PCR's is of more than academic interest.

2. DATA AND METHODOLOGY

The floriculture industry is a subset of the greenhouse industry and includes bedding/garden plants, cut flowers and cut florist greens, foliage plants, and potted flowering plants. This industry was selected because it represents the primary crops of northwestern Ohio growers. The data were obtained from the U.S. Census of Agriculture, 2002 (U.S. Department of Agriculture, 2002). County data were employed because they are the smallest units for which the most complete data are reported.

The Census of Agriculture data has inherent limitations. One problem is the non-disclosure rule which means the Census does not publish data which could reveal information about an individual farm. This rule results in data on the number of operations being reported more frequently than information on the nature of production. For example, the amount of production occurring under glass versus open fields is a common measure of output in the industry. However, such data for over 66 percent of the counties are not reported. The non-disclosure rule also precludes the use of zip code data. While zip codes would provide more geographic specificity, frequently no data on the number of farms are reported due to the non-disclosure rule. For example, in Ohio alone, no data are reported for almost 35 percent of the zip codes. Another major problem is that no sales data were reported in the 2002 Census even though such data were published in previous censuses. Due to these data constraints, the measure of production in each county is the number of floriculture operations.

In this project, Getis-Ord's Gi* is used to identify clusters of floricultural operations. Various researchers have applied this spatial statistic to identify industrial clusters (Feser et al., 2001; Feser et al., 2005; Helsel et al., 2007). This statistic measures spatial autocorrelation at the local level and it identifies "hot spots", or concentrations in spatial distributions in which areal units and their neighbors have similar values of a given phenomena (Mitchell, 2005). A high Gi* value indicates that high values are clustered near each other, whereas a low Gi* value is indicative of low values being near each other (Wong 2006). Gi* is useful in identifying the spatial footprint of clusters of economic activity since it examines patterns of co-location, or clusters, across areal unit boundaries within a specified neighborhood. In contrast, other measures, such as location quotients, examine only the value for a single areal unit without reference to values in neighboring areas.

An important component of local measures of spatial autocorrelation, including Gi*, is the specification of the local neighborhood as defined by the spatial weights matrix. Varying definitions of that matrix will lead to differing neighborhoods and, perhaps produce varying indices of spatial autocorrelation. The spatial weights matrix can be defined using rook or queen's measures of adjacency, distance between county centroids, inverse distance function, inverse distance squared, stochastic weights, and the like (Getis and Aldstadt, 2004; Mitchell, 2005; Wang, 2006; Wong and Lee, 2005).

Selection of a specific spatial weights matrix ideally should be based on some theoretical consideration regarding spatial interactions between counties. Unfortunately, the cluster literature contains no consensus on the spatial extent of a cluster. On the one hand, Porter (2000, p. 16) argued that: "The geographic scope of clusters ranges from a region, a state, or even a single city to span nearby or neighboring countries." On the other hand, May and his colleagues (2001) suggested that a cluster is characterized by firms agglomerating in a region up to 50 miles in radius. When using Gi*, Feser et al. (2001 and 2005) and Helsel et al. (2007) considered only adjacent counties.

The lack of theoretical principles for delineating neighbors is a pervasive problem in the application of these statistics. Paez and Scott (2005) observed that standard rules to guide the selection of distances for analysis in any type of study, not just cluster analysis, are underdeveloped. ESRI (2005), in a White Paper, suggested one try various distance bands using alternative distance functions and then select the one that shows the greatest amount of clustering. In the absence of a theoretical basis or standard rules for defining the appropriate size of a neighborhood for a cluster, pragmatically one can only empirically experiment with various spatial weights matrices. For that reason, we used two different spatial weights matrices. One is an adjacency matrix where the specification of neighbors is based on the queen's case of neighbor relations in which all counties are considered neighbors as long as they touch each other (Lee and Wong, 2001). The other is an inverse distance spatial weights matrix in which the strength of the relationship between counties is inversely proportional to the distance between the counties (Lee and Wong, 2001).

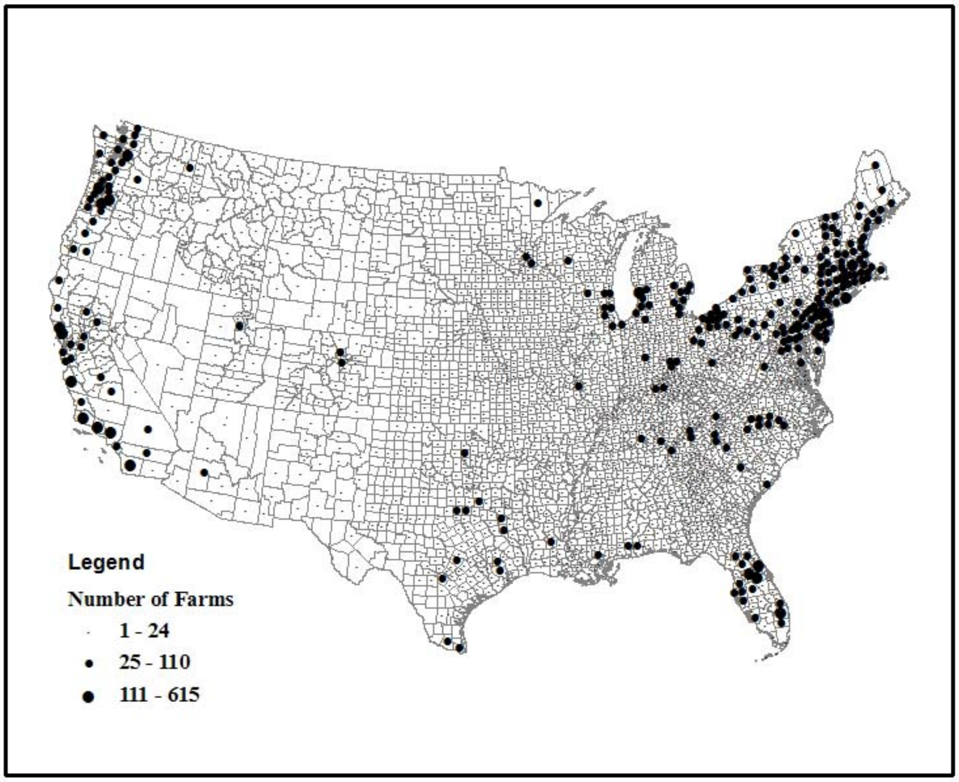
The rationale for selecting spatial weights matrices which emphasize local neighborhoods is based on market characteristics in Ohio. In a 2004 survey, growers in Ohio reported that over 75 percent of their sales were local (Maumee Valley Growers Cluster Project, 2004). Moreover 70 percent reported that they faced high to moderate competition from growers in their home county and over 65 percent faced high to moderate competition from producers in neighboring counties. Thus the growers' sales and competition are located primarily, but not exclusively, in their home county and surrounding counties. Data do not exist to determine to what extent this situation prevails in other parts of the nation outside Ohio, but a study in Monroe County, New York implied that much of the growers sales were local (Brumfield, 2001).

3. RESULTS

Floricultural production is widespread in the continental U.S., with 78.1 percent of the counties containing at least one floricultural operation (Figure 1). However, there

FIGURE 1

LOCATION OF FLORICULTURE OPERATIONS



are several major concentrations of production, depending on how one chooses to delineate them. The West Coast, extending from California to Washington, is one area of production. In fact, California leads the nation in number of floricultural operations, with 9.2 percent of the total, and Oregon and Washington are ranked eighth and tenth respectively. A second concentration extends from southeastern Wisconsin and northeastern Illinois to the East Coast and north from Virginia to Maine. Major states in this area include New York, Pennsylvania, and Ohio, with each containing over 5.0 percent of the nation's floricultural operations. Florida, with 5.9 percent of the country's operations, constitutes yet another concentration. A smaller concentration extends from Tennessee through North and South Carolina, as well as Georgia. This area contains about 10.3 percent of the nation's floricultural operations. Outside of these concentrations, there are smaller areas of production scattered around the nation.

As discussed previously, two spatial weights matrices (adjacency and inverse distance) were used to delineate the PCR's. Euclidean distances, as opposed to Manhattan distances, were employed. The Cluster and Outlier Analysis tool in the ArcGIS 9.1's Spatial Statistics Tools was used to compute the Gi*. For the inverse distance matrix, the Cluster and Outlier Analysis Tool requires a threshold distance beyond which counties are excluded from the computations. In keeping with the localized neighborhood, a threshold of 100 miles was selected as opposed to a larger distance band. The Cluster and Outlier Analysis tool's output is z scores for Gi* which can be thought of as z scores along a normal curve (Wang, 2006). To identify those counties to be included in the PCR, all counties with a z score greater than 1.96, or significant at the 0.05 level, were selected.

PCR's were delineated using both an adjacency spatial weights matrix (Figure 2) an inverse distance spatial weights matrix (Figure 3). In both maps, similar segments of the West Coast, Northeast, and Florida as well as a few outliers are included in the PCR's.

FIGURE 2

PCA'S USING ADJACENCY SPATIAL WEIGHTS MATRIX

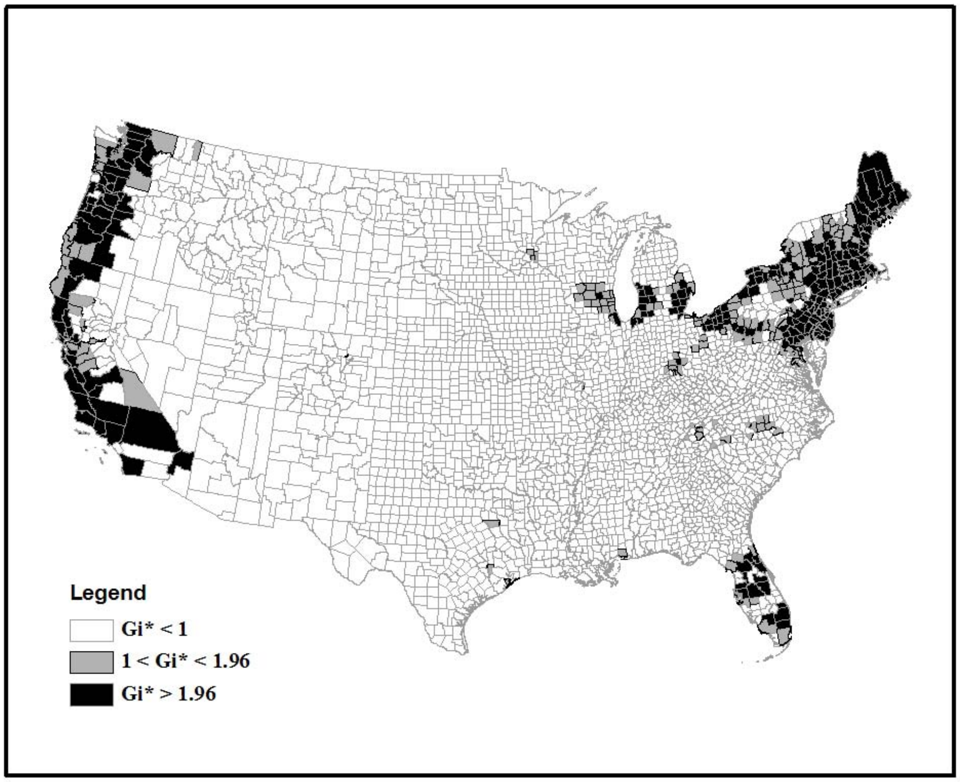
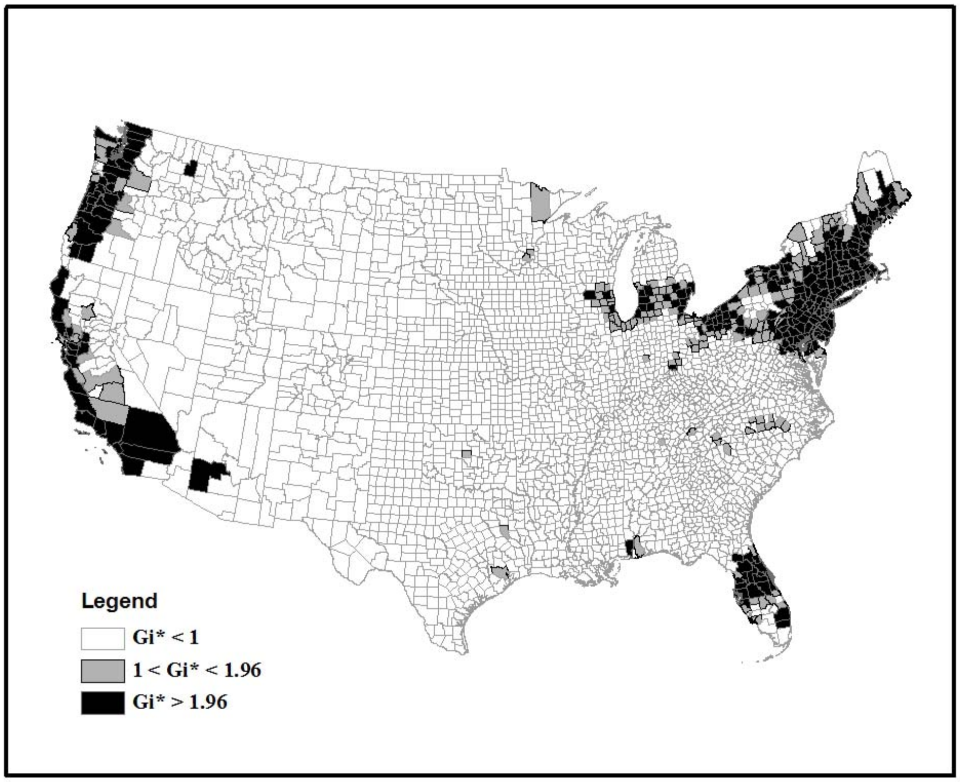


FIGURE 3

PCR'S USING INVERSE DISTANCE SPATIAL WEIGHTS MATRIX



As one would expect, the spatial footprint of the PCR's created with the adjacency weights matrix covers less geographic area than the one created with the inverse distance spatial weights matrix, 239 counties versus 268 counties. Even though the adjacency PCR contained only 7.7 percent of the continental U.S. counties, it contains 18.1 percent of the floricultural operations and 51.8 percent of the wholesalers. In contrast, the PCR created with the inverse distance weights matrix includes 8.6 percent of the counties, 50.4 percent of the floricultural operations, and 65.1 percent of the wholesalers. The number of wholesalers is important because their role in the industry's distribution chain is growing due to the increasing trend of "big box" and grocery stores selling floricultural products (Hall et al., 2005).

In general, Gi* does identify useful clusters of floricultural operations based on patterns of co-location. The extent of those clusters throughout the nation seems reasonable given the pattern of floricultural production in the U.S. The national clusters can be viewed as "geographic benchmarks" for monitoring industry trends. Moreover, the analysis identified several Michigan counties which are "hot spots" that could be joined with the northwest Ohio cluster. It must be emphasized that they are only potential areas. Unless the firms are willing and the necessary social networks can be constructed, co-location is rather meaningless. Wood County, Ohio is a good example of the importance of grower's motivations. Even though Gi* does not include Wood County in the cluster, many Wood County growers have bought into the cluster concept and are major contributors to the cluster. As stated earlier, use of Gi* is only the first step in identifying potential cluster regions.

4. CONCLUSION

Gi* is useful as a first step in identifying clusters, particularly since it examines patterns of co-location. This measure distinguishes concentrations of phenomena, or clusters, across areal unit boundaries within a specified neighborhood. In contrast, other measures, such as location quotients, examine only the value for a single county without reference to values in neighboring counties.

The output of spatial statistics provides a quick method of searching large numbers of areal units for PCR's. Identifying regions that have potential for industrial clustering can be a laborious sort through county data tables. In addition, Gi* is adaptable to varying sizes of areal units, be they counties or metropolitan areas. Moreover, one can incorporate other variables, such as market attributes, supply linkages, and the like, in the initial search for PCR's, as illustrated by the addition of wholesalers in this study. However, this methodology will not indicate the motivation behind spatial concentration, be it the result of historical precedent, a beneficial transportation infrastructure or other types of Marshallian agglomeration. Also, this approach will not identify embryonic or emerging clusters, but it can quickly isolate "hot spots" which may be suitable for cluster based economic policy. Moreover, it is not possible to include those firm behaviors, such as social networking, which are critical to the success of clusters.

The use of spatial statistics would be strengthened by further theoretical development of cluster theory. In this paper we used two spatial weights and produced varying results. If cluster-based economic development had a stronger theoretical base identifying the spatial extent of clusters, it would enhance the empirics. In the absence of theory, the only possible approach to selecting the spatial weights matrix is the "trial and error" strategy suggested in ESRI's 2005 White Paper.

Directly related to the distance issue is variations in the sizes of U.S. counties. The largest county in the continental U.S. is San Bernardino County whose centroid is more than 150 miles from its nearest neighbor. In contrast, there are counties in New Jersey which are less than 30 miles from their nearest neighbors. If one is using some type of distance function, these differential distances are likely to influence the results. If one uses an adjacency definition of neighbors, then one must be aware of the implied differences in distance. This is an issue that the authors are continuing to study by examining the impact of polygon geometry on local indicators of spatial autocorrelation.

In addition, the utility of incorporating a non-symmetric spatial weights matrix is being investigated.

5. REFERENCES

Akundi, K. 2003. Cluster-based Economic Development, Part 1: A Survey of State Initiatives. Texas Economic Development, Business and Industry Data Center. http://www.bidc.state.tx.us/Cluster%20Based%20EconDev%20PART1.pdf. Last accessed April 2007.

Brumfield, R. 2001. Monroe County Greenhouse Industry Market Research Study. http://www.monroecounty.gov/p/farm-GreenhouseReport.pdf. Last accessed April 2007.

ESRI. 2005. Spatial Statistics for Commercial Applications, An ESRI White Paper. http://www.esri.com/library/whitepapers/pdfs/spatial-stats-comm-apps.pdf. Last accessed April 2007.

Feser, E, and E.Bergman. 2000. National Industry Templates: A Framework for Applied Regional Cluster Analysis. Regional Studies 34(1):1-19.

Feser, E., K. Koo, H. Renski, and S. Sweeney. 2001. Incorporating Spatial Analysis in Applied Industry Cluster Studies. http://www.urban.uiuc.edu/faculty/feser/PUBS/EDQ%20Revised.pdf. Last accessed April 2007.

Feser, E., S. Sweeney, and H Renski. 2005. A Descriptive Analysis of Discrete U.S. Industrial Complexes. Journal of Regional Science 45(2): 395-419.

Getis, A., and J. Aldstadt. 2004. Constructing the Spatial Weights Matrix: Using a Local Statistic. Geographical Analysis 36(2): 90-104.

Hall, C., A. Hodges, and J. Haydu. 2005. Economic Impacts of the Green Industry in the United States. http://www.utextension.utk.edu/hbin/greenimpact.html. Last accessed April 2007.

Helsel, J., H. Kim, and J. Lee. 2007. An Evolutional Model of U.S. Manufacturing and Services Industries. In: Enterprising Worlds : A Geographic Perspective on Economics, Environments & Ethics, eds. J. Gatrell, and N. Reid, 83-98. Dordrecht, Netherlands: Springer.

Hendry, C., and J. Brown. 2006. Dynamics of Clustering and Performance in the UK Opto-electronics Industry. Regional Studies 40(7): 707-725.

Hill, E., and J. Brennan. 2000. A Methodology for Identifying the Drivers of Industrial Clusters: the Foundation of Regional Competitive Advantage. Economic Development Quarterly. 14 (1): 65-96.

Martin, R., and P. Sunley. 2003. Deconstructing Clusters: Chaotic Concept or Policy Panacea? Journal of Economic Geography 3(1): 5-35.

Maumee Valley Growers Cluster Project. 2004. Ohio USDA Survey Results. http://www.ohiogreenhouse.com/survey.htm. Last accessed April 2007.

May, W., C. Mason, and S. Pinch. 2001. Explaining Industrial Agglomeration: The Case of the British High-Fidelity Industry. Geoforum 32(3): 363-376.

Miller, P., R. Botham, R. Martin, and B. Moore. 2001. Business Clusters in the UK: A First Assessment. London: Department of Trade and Industry.

Mitchell, A. 2005. The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics. Redlands: ESRI Press.

Paez, A., and D. Scott. 2004. Spatial Statistics for Urban Analysis: A Review of Techniques with Examples. GeoJournal 61(1): 53-67.

Porter, M. 1998. On Competition. Boston: Harvard Business School Press.

Porter, M. 2000. Location, Competition, and Economic Development: Local Clusters in a Global Economy. Economic Development Quarterly 14(1): 15-34.

Reid, N., and M. Carroll. 2007. Collaborating to Compete: The Case of the Northwest Ohio Greenhouse Industry. In: Enterprising Worlds : A Geographic Perspective on Economics, Environments & Ethics, eds. J. Gatrell, and N. Reid, 41-56. Dordrecht, Netherlands: Springer.

Roberts, B., and R. Stimson. 1998. Multi-sectoral Qualitative Analysis: A Tool for Assessing the Competitiveness of Regions and Formulating Strategies for Economic Development. Annals of Regional Science 32(4): 469-494.

Solvell, O., G. Lindqvist, and C. Ketels. 2003. The Cluster Initiative Greenbook. http://www.cluster-research.org/greenbook.htm. Last accessed April 2007.

U.S. Census Bureau. 2002. Economic Census, 2002. On: http://factfinder.census.gov/servlet/DatasetMainPageServlet?_program=ECN&_submenuId=datasets_4&_lang=en&_ts=. Last accessed April 2007.

U.S. Department of Agriculture. National Agricultural Statistics Service. 2002. Census of Agriculture, 2002. http://www.nass.usda.gov/Census_of_Agriculture/index.asp. Last accessed April 2007.

Wang, F, 2006. Quantitative Methods and Applications in GIS. Boca Raton: Taylor & Francis Group.

Wong, D., and J. 2005. Statistical Analysis of Geographic Information with Arcview GIS and ArcGIS. Hoboken: John Wiley & Sons.