An Interdisciplinary Spatio-Temporal Population Model*

Guangqing Chi
Social Science Research Center
Mississippi State University

Stephen J. Ventura
Department of Soil Science
University of Wisconsin-Madison

* Address correspondence to Guangqing Chi, P. O. Box 5287, Mississippi State, MS 39762. E-mail: gchi@ssrc.msstate.edu; Tel: (662)325-9166; Fax: (662)325-9062. The authors thank Paul R. Voss for his guidance on the preliminary analysis of this research.
An Interdisciplinary Spatio-Temporal Population Model

ABSTRACT  Population growth (and decline) has been studied in several fields including demography, human ecology, population geography, environmental sociology, transportation planning, and regional economics. However, little systematic work has been undertaken to synthesize their different approaches to and views on population change. In this study, we examine population change holistically in an attempt to shed light on the mechanism of population change. We first systematize population change’s driving forces and spatial and temporal dimensions from an interdisciplinary perspective. The driving forces are organized and developed into five indices – demographics, livability, accessibility, desirability, and developability. We then test our approach by examining population change from 1970-2000 in Wisconsin at the municipal level. The findings suggest that such an approach helps systematically understand driving factors’ effects on population change, capture their spatial autocorrelation, minimize the multicollinearity, reduce heteroskedasticity, eliminate spatial error and lag dependence, and integrate the two seemingly unmixable approaches of environmental modeling and demographic modeling together.

KEYWORDS: population change; spatial; temporal; demographics.
INTRODUCTION

Population growth (and decline) has been studied in several fields including demography, human ecology, population geography, environmental sociology, transportation planning, and regional economics. These fields adopt different theories to explain and have different approaches to model population change, and thus they view population change differently. Human ecologists model population change within a holistic human-society-environment framework. Population geographers are interested in spatial variations of population distribution, density, composition, and growth. Regional scientists and transportation planners study population via land use patterns. Environmental planners focus on predictions of population growth and consequent impacts. Rural demographers and some regional scientists emphasize the role of natural amenities in affecting population change.

Although existing population studies generally do consider numerous factors in explaining population change, these factors tend to be chosen by an unnecessarily narrow perspective rather than an interdisciplinary perspective informed by other theories and potential data sets. Moreover, the spatial and temporal dimensions of population change are not considered explicitly in many approaches. In the United States context, human settlement processes have experienced distinct patterns through time, such as dispersed colonization followed by city development, and then suburban and finally ex-urban population growth. Incorporating temporal patterns in models may help with understanding of the phenomenon. In the spatial dimension, the fact that population change is spatially autocorrelated is supported by Tobler (1970)’s First Law of Geography, the spatial diffusion theory of population geography (Hudson, 1972), and regional economic theories such as growth role theory (Perroux, 1955). These limitations demand a universal view of population change, which considers population
change’s wide-ranged driving forces and co-variate factors, as well as its spatial and temporal dimensions.

In this study, we examine population change holistically in an attempt to shed light on the mechanism of population change. Following a review of existing disciplines’ views on and approaches to population change, we integrate population change’s driving forces and spatial and temporal dimensions together to propose a synthetic approach to model population change. The hypothesized driving factors are organized and developed into five indices — demographics, livability, accessibility, desirability, and developability. These are evaluated and discussed in the context of population change from 1970-2000 at the minor civil division level in Wisconsin.

**A SYNTHETIC VIEW OF POPULATION CHANGE**

*Existing Population-related Disciplines*

Although several theories have been offered to explain the mechanism of population change, along with numerous empirical studies, little systematic work has been undertaken to synthesize these explanations. Much of this vast literature is scattered across several disciplines, which has resulted in a complex mixture of theoretical and empirical approaches to and views on population change.

Early human ecologists use simple but holistic-oriented models (Duncan, 1964a, 1964b; Commoner, 1972; Ehrlich & Ehrlich, 1990; Ehrlich & Holdren, 1971; Holdren & Ehrlich, 1974; Bilsborrow, 1992), and more recent human ecologists employ the notion of “sustenance organization,” the means of generating livelihoods, to model population (Frisbie & Poston, 1976, 1975, 1978; Hirschl, Poston, & Frisbie, 1998; Poston & Frisbie, 1998). However, familiar models such as the POET (Population, Organization, Environment, and Technology) model and
the IPAT (Impact, Population, Affluence, and Technology) model have been criticized for incorporating few variables, and other complex models are rather descriptive and qualitative (Dietz & Rosa, 1994).

*Population geography* seeks and explains population patterns caused by spatial regularities and processes (Beaujeu-Garnier, 1966; James, 1954; Jones, 1990; Trewartha, 1953; Zelinsky, 1966). Although population geography adopts GIS and statistical models to study the spatial characteristics of population as well as the spatial processes involved, it often ignores socioeconomic considerations and the temporal dimension of population.

*Regional scientists*, especially regional economists, are strong in explaining and modeling the change of land use patterns, which are almost always associated with population change (Cervero & Hansen, 2002; Moore & Thorsnes, 1994; Cervero, 2002; Aschauer, 1990; Cervero, 2003; Boarnet, 1997, 1998). The fact that demographers focus on “population” and regional economists are interested in “employees” and “housing” provides an opportunity for borrowing regional economic theories and models to improve population forecasts. Empirical studies on regional development often stress the interdependencies between household residential choices and firm location decisions (Henry, Barkley, & Bao, 1997), especially in literature identifying causality between employment and population change (Steinnes & Fisher, 1974).

*Environmental planners* focus on how the physical environment and socio-economic conditions encourage or discourage land use change. The approach is generally empirical, typically using geographic information system (GIS) overlay methods such as those supported through the ModelBuilder function (®ESRI) in ArcGIS to answer “what-if” questions. Similar work include developable lands (Cowen & Jensen, 1998), qualitative environmental corridors

*Environmental sociology* attempts to theorize population-environment issues within a political economy context. The neo-Marxist theory claims that development patterns and related population change are the result of capital in pursuit of profit (Hall, 1988). The strength of neo-Marxist theories lies in its theoretical interpretation solely, not modeling of land development and population change.

*Demography*, as a science of population statistics, often interprets statistical information of population rather than explaining the phenomenon theoretically. The fact that most demographic models are data-driven rather than theory-based sometimes forces demographers to borrow theories from other disciplines. In addition, the driving factors of population change tend to be chosen only from the demographic perspective.

In sum, each population-related discipline has its strengths and weaknesses in viewing and modeling population change. An interdisciplinary approach should integrate their strengths by testing the driving factors considered relevant in multiple perspectives, along with spatial and temporal dimensions of population change. For example, several studies called for a holistic approach to study population-environment issues (e.g., Barham, 2001; Belsky, 2002; Luzadis et al., 2002; Machlis, Force, & Burch, 1997; Michaelidou, Decker, and Lassoie, 2002).

*Population Change’s Driving Forces and Covariated Factors*

An exhaustive literature review of population-related theories and empirical studies resulted in approximately 70 variables that were considered significantly relevant to population change. These were within the broad realms of demographic characteristics, socioeconomic
conditions, physical infrastructure, environmental and geophysical factors, cultural resources, and potential legal constraints. However, even these broad realms often are not adequately controlled in multivariate analyses (Lichter & Fuguitt, 1980). A wide range of results is possible by omitting relevant variables from the model (Dalenberg & Partridge, 1997), and such omissions can potentially bias parameter estimates for those variables that are included in the model specification. In the proposed synthetic approach, we attempt to incorporate these variables into the model on the basis of our judgment and data availability.

It has long been understood that demographic characteristics of a population are important determinants of population change. The most important demographic characteristics that may affect population change are population density (Humphrey, 1980; Humphrey et al., 1977; Lutz, 1994a; Moore & Thorsnes, 1994), age structure (DaVanzo, 1981; Humphrey, 1980; Shryock, 1964), racial and ethnic composition (Friedman & Lichter, 1998; Shryock, 1964), institutional populations (DaVanzo, 1981; Humphrey et al., 1977; Mincer, 1978), educational attainments (DaVanzo, 1981; Mincer, 1978), migration (Johnson & Purdy, 1980), household demographic characteristics (DaVanzo, 1981; Mincer, 1978; Shryock, 1964), and sustenance organization (Browning & Gibbs, 1971; Gibbs & Browning, 1966; Frisbie & Poston, 1975, 1976, 1978; Gibbs & Martin, 1958; Hirschl, Poston, & Frisbie, 1998; Poston & Frisbie, 1998; Sly, 1972).

Second, socioeconomic conditions known to have important impacts on population change include school performance (Marcouiller, Kim, & Deller, 2004), crime (Carlino & Mills, 1987; Clark & Murphy, 1996; Deller et al., 2001; Graves, 1979; Marcouiller, Kim, & Deller, 2004; Messner & Anselin, 2004; Schachter & Althaus, 1989; Smith, Tayman, & Swanson, 2000), income (Fuguitt, Brown, & Beale, 1989; Johnson & Beale, 1994; Lyson & Gillespie,
Third, access to transportation and other critical infrastructure is important for local economic growth and development, as well as associated population growth. Accessibility can influence population change indirectly through economic growth, employment change, altered social structure, and environmental change (Chi, Voss, & Deller, 2006). There are many theoretical and empirical works on accessibility in the fields of regional economics, transportation planning, rural sociology, demography, and others. Relevant factors include residential preference (Astone & McLanahan, 1994; Bartel, 1979; Brown et al., 1997; Fuguitt & Brown, 1990; Fuguitt & Zuiches, 1975; Mincer, 1978; Zuiches & Rieger, 1978), highways (Chi, Voss, & Deller, 2006; Humphrey, 1980; Smith, Tayman, & Swanson, 2000; Voss & Chi, 2006), traffic volume (Hobbs & Campbell, 1967), distance to access of highways (Humphrey, 1980; Smith, Tayman, & Swanson, 2000), journey to work, local capital expenditures on transportation (Humphrey, 1980; Humphrey et al., 1977), and others.

Fourth, environmental and natural resource characteristics are known to influence population growth. Dis-amenities (negative influences on population) include landfills and other noxious sites, resource extractions, and propensity to natural disasters. In recent decades, natural resource characteristics such as water features, terrain relief (e.g., viewsheds), and landscape
aesthetics (e.g., regional land use and cover) have been viewed as influences on population change mainly through the role of natural amenities, which are seen as the principal contributor of non-metropolitan population growth (Brown et al., 1997; Deller et al., 2001; English, Marcouiller, & Cordell, 2000; Fuguitt & Brown, 1990; Fuguitt, Brown, & Beale, 1989; Fuguitt & Zuiches, 1975; Humphrey, 1980; Johnson, 1982, 1989; Johnson & Beale, 1994; Johnson & Purdy, 1980; Marcouiller, 1997; Zuiches & Rieger, 1978; Jones et al., 2003).

Equilibrium theory argues that the main determinants of migration come from differences in amenities rather than differences in economic opportunities (Graves, 1979, 1983; Graves & Linneman, 1979). The life-cycle literature suggests that amenity factors become more important as people become older (Clark & Hunter, 1992; Humphrey, 1980). Some regional economists see natural amenities as latent regional factor inputs to the local production of goods and services. They argue that natural amenities play a significant role in affecting economic development and migration (English, Marcouiller, & Cordell, 2000; Graves, 1979, 1983, 1980; Knapp & Graves, 1989; Porell, 1982). So-called “growth engines” in rural areas increasingly are less dependent upon traditional tangible factor inputs (land, labor and capital) and more dependent upon latent factor inputs (such as amenity-based goods and services) (Marcouiller, Kim, & Deller, 2004).

Finally, population growth is limited by the potential for land conversion and development. The land developability of a region is determined by its geophysical characteristics (water, wetland, slope, and publicly-owned lands), built-up lands (existing residential, commercial, and industrial developments, as well as transportation infrastructure), cultural and aesthetical resources, and legal constraints (including land use planning legislation and programs such as comprehensive plans, “smart growth” laws, zoning ordinances, farmland
protection programs, environmental regulations such as Clean Water Act, shoreland and wetland zoning, and others).

Population Change’s Temporal Dimension

A model of population change’s driving forces and co-variate factors must be sensitive to changes in those forces over time. Not all drivers are present or equally influential at all times. Regional scientists, demographers, human geographers, and scholars in other disciplines have studied the history of population distribution and settlement patterns, and explored the determinants of changes.

The industrial revolution that started in the 18th century has significantly transformed human societies. Most urban development models attribute modern urban and suburban trends to technological innovations in transportation, communication, and production, population increase, rising affluence, free markets, and personal choice (Jaret, 1983). The mono-centric city in America was the form two centuries ago. Cities tended to have a single employment center from about 1850 to 1930 (Moore & Thorsnes, 1994). The process of suburbanization started at the beginning of the 20th century when the innovation of transportation tools allowed people to live in a suburb and work in a city center. The early suburban development was exclusively residential, and commercial development was added into suburbs since the 1930s (Pucher & Lefevre, 1996). Regional economists tend to analyze centralization and decentralization on the basis of urban form determinants (Moore & Thorsnes, 1994; Morris, 1994). Recently regional economic studies emphasize the role of transportation and technology in changing urban form.

When transportation infrastructure reaches a certain level at which geographic distance imposes no important effects on local economy, some decision-makers switch to local natural
amenities as drivers of economic growth and development (Marcouiller, Kim, & Deller, 2004). The process of ruralization started in the 1970s when people moved from metropolitan areas to rural areas with substantial natural recreational facilities and natural amenities. Demographers are also interested in turnaround migration in non-metropolitan areas, and seek reasons from the perspective of migration decision-making. Many of them understand the attractiveness of natural amenities as the principal contributor. Centralization re-gained power in the 1980s especially in some metropolitan areas, while suburbanization and ruralization strengthened in the 1990s (Brown et al., 1997).

Human settlement form has experienced dramatic change — survivalization, centralization, decentralization, and ruralization, which are characterized by different determining factors. An important (or unimportant) factor in a time period may become unimportant (important) in another. Thus, a systematic view is important for examining population change.

*Population Change’s Spatial Dimension*

Since its inception a few decades ago, spatial statistics have been applied in numerous fields (Anselin, 1988). However, these tools have drawn demographers’ attention only recently. While demography has a rich methodology, it is largely lacking a spatial perspective (Tiefelsdorf, 2000). It is important to consider spatial effects in demography, not only because several theories consider population change to be spatially autocorrelated, but also because the estimation and inference will be unreliable and estimates of the effects of independent variables may be biased if the spatial effect exists but is not accounted for (Anselin & Bera, 1998; Baller et al., 2001).
Population growth in one unit of geography (the “focal” unit) can be shown to be correlated (autocorrelated) with its neighboring units. This observed pattern is supported by at least three schools of theories\(^1\). Tobler (1970)’s First Law of Geography states that everything relates to everything else, but nearer ones do more so. The spatial diffusion theory of population geography argues that population growth will spread to surrounding areas (Boyce, 1966; Morrill, 1968; Thrall et al., 2001). It implies that population growth is spatially autocorrelated. Regional economic theories such as growth pole theory apply spread and backwash notions to explain the mutual geographic dependence of economic growth and development, which in turn causes population change (Hartshorn & Walcott, 2000; Richardson, 1976).

*Tying Them Together*

Although existing population-environment studies generally do consider numerous factors in explaining population change, these tend to be constrained to by disciplines. A multi- or inter-interdisciplinary perspective draws on all potentially relevant theories and data sets. Moreover, in most traditional approaches, the spatial and temporal dimensions of population change are not considered explicitly. We argue that a systematic view on population change should consider three dimensions – its driving factors, the temporal dimension, and the spatial dimension (Figure 1).

![FIGURE 1 ABOUT HERE](image)

Nevertheless, the large number of independent variables (driving factors) loses the interpretability of the whole picture of population change, and often raises the problem of serious and unnecessary multicollinearity which affects the efficiency of multivariate regression and other models. In order to solve this dilemma and take the strengths of each population-related
approach, we categorize population change’s driving factors and develop five indices: 

demographics (local demographic characteristics), (2) *livability* (a measure of social and
economic conditions), (3) *accessibility* (transportation and community infrastructure), (4)
desirability (a measure of the area’s attractive power from the presence or absence of natural
amenities), and (5) *developability* (the potential for land conversion and development).

DATA AND METHODOLOGY

This study focused on the state of Wisconsin as the research case, and covers population
change from 1970 to 2000 at the Minor Civil Division (MCD) level. The data used in this study
come from a variety of sources, primary or secondary. Population data are from decennial
censuses 1970-2000 and a commercial re-working of the data by Geolytics, Inc. Additional
demographic and socioeconomic data are acquired from the U.S. Census Bureau, the Wisconsin
Department of Public Instruction, the Federal Bureau of Investigation, and the State of
Wisconsin Blue Books. Transportation infrastructure data are provided by the Wisconsin
Department of Transportation, the Wisconsin Bureau of Aeronautics, the National Atlas of the
United States, and the Department of Civil and Environmental Engineering of the University of
Wisconsin-Madison. The data of geophysical factors and natural amenity characteristics come
from the Wisconsin Department of Natural Resources, the U.S. Geological Survey, and the
Environmental Remote Sensing Center and the Land Information & Computer Graphics Facility
of the University of Wisconsin-Madison.

This study is conducted at the MCD level. Wisconsin is a “strong MCD” state and its
MCDs – towns, cities, and villages – are functioning governmental units (with elected officials
who provide services and raise revenues). The MCD geography consists of non-nested, mutually
exclusive and exhaustive political territory. In most parts of the State, census tracts have an average size similar to MCDs and provide an alternative unit of analysis. However, census tracts are geographic units delineated by the Census Bureau only for counting population purpose, and they have no political or social meanings.

MCD boundaries are not stable over time. Boundaries change, new MCDs emerge, old MCDs disappear, names change, and status in the geographic hierarchy shifts (e.g., towns become villages, and villages become cities). In order to adjust the data for these changes, we have set up three rules: new MCDs must be merged into the original MCDs from which they emerge; disappearing MCD problems can be solved by dissolving the original MCDs into their current “home” MCDs; and occasionally, several distinct MCDs must be dissolved into one super-MCD in order to establish a consistent data set over time. In the end, 1,837 MCDs constitute our analytical dataset with an average size of 29.56 square miles.

Figure 2 shows population distributions (in terms of population density) from 1970 to 2000 at the MCD level. In 1970 the population was mainly distributed in the southeast of Wisconsin, where manufacturing and brewing industries were dominant. By 2000 urban areas expanded into their surrounding areas and there has been an important population expansion along the Wisconsin River and to portions of the center of the Wisconsin Northwoods (the northern Wisconsin where forests are the main type of ground cover).

[FIGURE 2 ABOUT HERE]

Population change is the dependent variable, and is expressed as the natural log of population at a census year over the population ten years earlier. For example, the natural log of the 1990 population divided by the 1980 population is our measure of growth rate from 1980 to 1990. This representation helps achieve the normal “bell-shaped” distribution, makes it easier to
interpret population change rate (a positive value means growth; a negative value means decline; and zero means no change at all), and controls the effect of initial population size (Humphrey, 1980; Lutz, 1994b; Smith, Tayman, & Swanson, 2000).

In this study, we first generate five indices — demographics, livability, accessibility, desirability, and developability — out of a candidate set of 33 factors posited to influence population change. The developability index is generated using the ModelBuilder function (®ESRI) in ArcGIS, and the remaining four indices by Principal Component Analysis (PCA). This design not only reduces the number of independent variables, but also integrates the two seemingly unmixable approaches of environmental modeling based on factor combination modeling and population modeling based on regression.

The spatial autocorrelation within each of the five indices is then examined on the basis of Moran’s I statistic. Spatial autocorrelation can be loosely defined as similarity (or dissimilarity) between two values of an attribute with similarity between the two corresponding locations. A positive Moran’s I value indicates positive spatial autocorrelation (i.e., high or low values of an attribute tend to cluster in space), and a negative Moran’s I value indicates negative spatial autocorrelation (i.e., locations tend to be surrounded by neighbors with very different values). Moran’s I statistic for residuals in a model is often used to detect if the model is appropriately specified.

Next, we employ the generated indices and individual variables separately as the independent variables in regression models, to test our hypothesis that indices are more efficient than individual variables in modeling population change. Their performance are evaluated on the basis of several statistics: log likelihood, Akaike Information Criterion (AIC), multicollinearity condition number, Koenker-Bassett test, Moran's I statistic for residuals, robust
Lagrange-multiplier (LM) test for the error term, and robust LM test for the lag term. Finally, we compare several regression models with and without temporal and spatial effects to examine their effects in modeling population change.

**FINDINGS**

*Exploratory Analysis of Indices*

The demographic indices are based on population density, age structure, race, institutional population, educational attainment, migration, family characteristics, seasonal housing, and sustenance organization (agricultural and retail). The first component of PCA explains 26.46% variance in the 1980 dataset, and 23.65% in 1990 (Table 1). Figure 3 illustrates geographical distributions of the generated demographic indices in 1980 and 1990, and shows similar patterns of demographics for both 1980 and 1990. The generated demographic indices have statistically significant spatial autocorrelation as indicated by the Moran’s I statistic (0.2878 in 1980 and 0.4260 in 1990; p-value < 0.001 in both indices), which suggests a need to account for spatial autocorrelation of demographic factors in regression models.

[Table 1 About Here]

[Figure 3 About Here]

The livability indices consist of safety, school performance, public transportation, buses, public water, new housing, county seat status, income, real estate value, and employment rate. The first component of PCA explains 24.75% variance in the 1980 dataset, and 26.78% in the 1990 (Table 1). The generated livability indices have statistically significant and surprisingly strong spatial autocorrelation as indicated by the Moran’s I statistic (0.7849 in 1980 and 0.7860 in 1990; p-value < 0.001 in both indices; Figure 4), which suggests the potential need for using
the index both in its regular form and also as a spatially lagged variable in the regression modeling.

[FIGURE 4 ABOUT HERE]

The accessibility indices are composed of residential preference, highway infrastructure, accessibility to airports, accessibility to highways, and accessibility to workplaces\(^4\). The first component of PCA explains 32.72% variance in the 1980 dataset, and 33.29% in 1990 (Table 1). The generated accessibility indices have statistically significant and strong spatial autocorrelation as indicated by the Moran’s I statistic (0.4639 in 1980 and 0.4882 in 1990; p-value < 0.001 in both indices; Figure 5). This suggests potential spatial autocorrelation in the accessibility variables and the probable need for incorporation of spatial autocorrelation into population modeling.

[FIGURE 5 ABOUT HERE]

The desirability index considers forests, water, lakeshore/riverbank/coastline, golf courses, and slope (Figure 6). The first component of PCA explains 32.46% variance in the data set. The generated desirability index in Wisconsin has statistically significant and strong spatial autocorrelation as indicated by the Moran’s I statistic (0.4089; p-value < 0.001\(^5\)), which suggests the potential need to lag the desirability index in the regression model.

[FIGURE 6 ABOUT HERE]

Different from the above four indices, the developability index is developed using the ModelBuilder function of ArcGIS, which is often employed by environmental analysts to study the interactions between environment, population, land use, and legal constraints at fine pixel sizes. While demographers may be interested in borrowing this approach to study population, the fact that population data are aggregated at rather coarse sizes imposes difficulties in taking
into account environmental variables that generally can only be studied usefully at very fine data resolution. This research employs the ModelBuilder function in ArcGIS to generate a developability index for small polygons via the intermediate use of fine-scale pixels subsequently aggregated to census polygons. The variables include water, wetland, slope, tax-exempt lands (Frentz et al., 2004), and built-up lands. Figure 7 portrays the proportion of developable lands in each MCD. The derived developability index has high spatial autocorrelation as indicated by a visual examination of index distribution (Figure 7), and more formally by the Moran’s I statistic (0.3565; p-value < 0.001). This implies that a spatially lagged version of the index may improve our regression specification.

Regression Comparison between Indices and Individual Factors as Explanatory Variables

We compared how the indices perform in regression models as opposed to individual factors in representing explanatory variables (Table 2). We ran two regression models where we regressed the population change rate from 1990 to 2000 on all 33 individual variables and on the five indices in 1990 separately. In terms of multicollinearity and interpretability, the use of indices achieves the anticipated significant advantage over the use of individual variables. The model using indices has a much lower multicollinearity condition number (9.10) than the one using individual variables (658.92).
There is statistically significant spatial correlation in the model residuals using individual variables (Moran’s I = 0.05, p-value < 0.001), but not in the model using indices (Moran’s I = 0.02, p-value = 0.127). Robust LM tests indicate that both statistically significant spatial lag (p-value < 0.001) and error dependence (p-value < 0.001) are indicated in the model using individual variables, but not in the model using indices (p-value = 0.131 and 0.929 separately). Although each of the five indices has strong spatial autocorrelation, their spatial autocorrelations are explained (or “neutralized”) by the spatial autocorrelation of the dependent variable. In addition, the Koenker-Bassett test points out that the model using individual variables has stronger heteroskedasticity than the one using indices. In other words, the former model’s variance of residuals is more dependent on the independent variables. The only disadvantage in using indices as the explanatory variables is that such model is not as well fitted as the model using individual variables, as indicated by the log likelihood and AIC statistics. This disadvantage is expected since the former has a much smaller number of independent variables.

In sum, the use of indices provides five advantages over individual variables as explanatory variables: (1) minimizing the multicollinearity in the regression model dramatically, (2) reducing the heteroskedasticity in the model, (3) eliminating the spatial error and lag dependence within the standard regression model though all the five generated indices have strong spatial autocorrelation, (4) becoming easier to interpret the regression findings, and (5) integrating the two seemingly unmixable approaches of environmental modeling (by ModelBuilder) and demographic modeling (regression) together.
Confirmatory Analysis

A regression using the indices can have many different specifications: with or without temporal effects of the dependent variable, with or without temporal effects of the independent variables, with or without spatial lag effects, and with or without spatial error effects. Here we examine nine models with different specifications, and we divide them into three groups: no temporal effect at all, considering temporal effects for the dependent variable only, and considering temporal effects for both dependent and independent variables. Each group is estimated by an Ordinary Least Squares (OLS) regression, and by a spatial lag model and a spatial error model (the latter two fits using maximum likelihood).

In the first group of models, the temporal effect is not considered. Based on the standard OLS regression (Eq. 1), each of the five indices is statistically significant in explaining population change 1990-2000 at $p \leq 0.07$ for a two-tail test (Table 3). In the presence of the other four indices, the accessibility index parameter turns negative. The other four indices affect population change positively. The log likelihood and AIC statistics suggest a preference for the spatial lag specification (Eq. 2) over the standard regression, and the spatial error model (Eq. 3) over both the spatial lag model and standard regression – although in each case the goodness-of-fit preference is only mildly stated. The two spatial models do not alter appreciably the OLS coefficients or their levels of statistical significance.

\[
\text{OLS: } \ln \left( \frac{P_{00}}{P_{90}} \right) = X_{90} \beta + \varepsilon \quad [1]
\]

\[
\text{SLM: } \ln \left( \frac{P_{00}}{P_{90}} \right) = X_{90} \beta + \lambda W \ln \left( \frac{P_{00}}{P_{90}} \right) + \varepsilon \quad [2]
\]
SEM:  
\[
\ln \left( \frac{P_{00}}{P_{90}} \right) = X_{90} \beta + \lambda \ln \left( \frac{P_{00}}{P_{90}} \right) - \lambda W X_{90} \beta + \varepsilon \quad [3]
\]

where

- \( P_{00} \) is population in year 2000,
- \( P_{90} \) is population in year 1990,
- \( X_{90} \) is matrix of index values in 1990,
- \( \beta \) is a vector of coefficients of \( X_{90} \),
- \( \lambda \) is a spatial parameter,
- and \( W \) is a neighborhood structure.

[TABLE 3 ABOUT HERE]

In the second group of models, only the temporal effect of the dependent variable is considered (Table 4). The second group of models does not dramatically change the coefficients or their significance over the first group. The log likelihood and AIC statistics still suggest a preference of the spatial lag model (Eq. 5) over the standard regression (Eq. 4), and the spatial error model (Eq. 6) over both the spatial lag and standard models, although the preferences are mild at best – a result of the absence of statistical significance for the spatial parameter in both the SLM and SEM models.

OLS:  
\[
\ln \left( \frac{P_{00}}{P_{90}} \right) = \ln \left( \frac{P_{90}}{P_{80}} \right) + X_{90} \beta + \varepsilon \quad [4]
\]

SLM:  
\[
\ln \left( \frac{P_{00}}{P_{90}} \right) = \ln \left( \frac{P_{90}}{P_{80}} \right) + X_{90} \beta + \lambda \ln \left( \frac{P_{00}}{P_{90}} \right) + \varepsilon \quad [5]
\]
In the third group of models, temporal effects for both the dependent and independent variables are considered (Table 5). Three temporally lagged indices (demographics, livability, and accessibility) are not significant in explaining population change 1990-2000, but have changed the coefficients and their significance of other individual variables dramatically. For example, in the standard regression (Eq. 7), demographics, livability, and accessibility in 1990 become insignificant in the presence of the temporally lagged indices. Once more, the log likelihood and AIC statistics suggest a preference for the spatial lag model (Eq. 8) over the standard regression but just barely (an exception is found with the AIC for the spatial lag model), and the spatial error model (Eq. 9) over other two.

By comparing Group 1 and Group 2 models, the goodness-of-fit improves dramatically when the temporal effect of the dependent variable is considered. However, the goodness-of-fit
improves only slightly when the temporal effect of independent variables is also considered (Group 3 models). In addition, the incorporation of spatial effects simply does not improve the goodness-of-fit much at all, and none of the spatial models show a significant spatial parameter. This result was predicted by the findings in the right-hand panel of Table 2.

Overall, the temporally lagged effect of the dependent variable is much more important than the temporally lagged effect of the independent variables in improving the goodness-of-fit of our regression model. Spatial autocorrelation should be considered in a regression when indicated. Thus it gave our models little added strength as indicated in Table 3 through 5.

**SUMMARY AND DISCUSSION**

Practically, the use of indices provides five advantages over individual variables as explanatory variables: minimizing the multicollinearity in the regression model dramatically, reducing the heteroskedasticity in the model, eliminating the spatial error and lag dependence within the standard regression model though all the five generated indices have strong spatial autocorrelation, becoming easier to interpret the regression findings, and integrating the two seemingly unmixable approaches of environmental modeling (by the ModelBuilder function) and demographic modeling (regression) together. However, the usage of indices loses the interpretability of each variable’s effects on population change. The temporal effect of the dependent variable is much more important than that of the independent variables in improving the goodness-of-fit of a regression model. Spatial autocorrelation should be considered in a model when indicated.
The systematic view of population change challenges conclusions from some existing studies, which examine how variables from one or more of the five indices influence population change without controlling the others. Conclusions from such studies are hardly reliable. The future development of an area is determined jointly by economic, political, demographic, geographic, social, and cultural forces. None of them can independently determine economic development and population change. It behooves us to be extremely careful when interpreting scientific findings, especially when most studies are conducted from a partial perspective. Moreover, the fact that the indices sometimes point in different directions with respect to population growth gives us pause as we embark on the confirmatory analysis, as factors that we had hoped would reinforce one another apparently do not always do so.

This study provides a systematic theoretical framework for understanding population change, and offers a holistic, comprehensive, or rational approach to study population change. The proposed framework builds on existing landscape-scale analyses to examine, specifically and systematically, the ways in which biogeophysical variables and socioeconomic structural variables interact with demographic variables to influence local population growth or decline. Population dynamics should be placed in their social, economic, and political settings, and be linked to environmental constraining/facilitating forces. Population change should be modeled holistically rather than separately within the complex social and economic context. The holistic approach increasingly is becoming critically important as local land-use conflict (in a domestic setting), regional/tribal warfare (on the international scene), and environmental degradation (nearly everywhere) arise as competition for scarce resources is heightened by growing population needs for food and shelter. Environmental problems in modern society are definitely
complex, and it requires a synthetic understanding of historic, cultural, political, economic, and demographic reasons. What we lack are not solutions, but effective solutions.

**ENDNOTES**

1. There might be some social or psychological theories and reasons why population growth is spatially correlated. Such literature needs an exhaustive exploration in order to provide a complete view of the spatial autocorrelation of population growth.

2. Even so, the adjusted MCD boundaries are not perfect, because some boundaries are not clearly known locally or change frequently as a result of annexations.

3. The final scores are of interval rather than ratio type, and there is no comparability between two sets of final scores. Because of that, the final scores in the two maps should be examined individually rather than compared.

4. The measurement of accessibility within Wisconsin and the variables that make up the accessibility index incorporates spatial effects in neighboring states (Minnesota, Iowa, Illinois, and Upper Michigan).

5. To get the most reliable information we need data in 1980 and 1990 for all variables. However, the remote sensing data for the whole Wisconsin are not available both in 1980 and 1990. What are available are the 1992-93 Landsat Thematic Mapper Imagery (LTMI) and the 1971-82 US Geological Survey (USGS) Land Use Data (LUDA). Although they can be used to represent the 1990 and 1980 data separately, only the former is used to derive variables of desirability and developability because of the concern of the LUDA accuracy. The LUDA were collected at different time points from 1971-82 for different parts of Wisconsin. They use different data sources, methodologies, and spatial scales.
6. The developability of a region is determined by its geophysical characteristics, built-up lands, cultural resources, and legal constraints. Ideally we want to use all of the four types of variables to derive the developability index. For practical reasons, we use only geophysical characteristics and built-up lands to generate the developability index.
REFERENCES


### TABLE 1  
**Analysis of Variables by Principal Component Analysis**

<table>
<thead>
<tr>
<th>Variables</th>
<th>1980</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance explained</td>
<td>26.46%</td>
<td>23.65%</td>
</tr>
<tr>
<td>Population density</td>
<td>0.3536</td>
<td>0.3938</td>
</tr>
<tr>
<td>Young</td>
<td>-0.3015</td>
<td>-0.2283</td>
</tr>
<tr>
<td>Old</td>
<td>0.2290</td>
<td>0.1076</td>
</tr>
<tr>
<td>Black</td>
<td>0.1195</td>
<td>0.1553</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.1115</td>
<td>0.1593</td>
</tr>
<tr>
<td>College students</td>
<td>0.2272</td>
<td>0.2894</td>
</tr>
<tr>
<td>High school education</td>
<td>0.1889</td>
<td>0.2429</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>0.3051</td>
<td>0.3636</td>
</tr>
<tr>
<td>Mover</td>
<td>-0.2890</td>
<td>-0.4014</td>
</tr>
<tr>
<td>Female-headed households</td>
<td>0.3533</td>
<td>0.1408</td>
</tr>
<tr>
<td>Seasonal housing</td>
<td>0.0149</td>
<td>-0.0382</td>
</tr>
<tr>
<td>Retail</td>
<td>0.3628</td>
<td>0.3055</td>
</tr>
<tr>
<td>Agriculture</td>
<td>-0.4294</td>
<td>-0.4255</td>
</tr>
<tr>
<td><strong>Livability</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance explained</td>
<td>24.75%</td>
<td>26.78%</td>
</tr>
<tr>
<td>Safety</td>
<td>-0.3529</td>
<td>-0.3242</td>
</tr>
<tr>
<td>School performance</td>
<td>0.2954</td>
<td>0.2734</td>
</tr>
<tr>
<td>Public performance</td>
<td>0.1745</td>
<td>0.0580</td>
</tr>
<tr>
<td>Public transportation</td>
<td>0.1745</td>
<td>0.0580</td>
</tr>
<tr>
<td>Buses</td>
<td>0.0869</td>
<td>0.3242</td>
</tr>
<tr>
<td>Public water</td>
<td>0.0653</td>
<td>0.0495</td>
</tr>
<tr>
<td>New housing</td>
<td>0.2215</td>
<td>0.2315</td>
</tr>
<tr>
<td>County seat</td>
<td>0.0259</td>
<td>0.0114</td>
</tr>
<tr>
<td>Income</td>
<td>0.5544</td>
<td>0.5381</td>
</tr>
<tr>
<td>Real estate value</td>
<td>0.5641</td>
<td>0.5393</td>
</tr>
<tr>
<td>Employment</td>
<td>0.2659</td>
<td>0.2742</td>
</tr>
<tr>
<td><strong>Accessibility</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance explained</td>
<td>32.72%</td>
<td>33.29%</td>
</tr>
<tr>
<td>Residential preference</td>
<td>0.5566</td>
<td>0.5459</td>
</tr>
<tr>
<td>Highway infrastructure</td>
<td>0.6056</td>
<td>0.5919</td>
</tr>
<tr>
<td>Accessibility to highways</td>
<td>-0.0027</td>
<td>0.0047</td>
</tr>
<tr>
<td>Accessibility to airports</td>
<td>0.4874</td>
<td>0.4840</td>
</tr>
<tr>
<td>Accessibility to workplaces</td>
<td>0.2932</td>
<td>0.3426</td>
</tr>
</tbody>
</table>
**Desirability**

- Variance explained: 32.46%
- Forests: 0.5163
- Water: -0.0666
- Lakeshore/Riverbank/Coastline: 0.6495
- Golf courses: 0.1922
- Slope (12.5%-20%): 0.5199

| TABLE 2 |
| Comparison between Indices and Individual Variables as Explanatory Variables |
|---------------------------------|-----------------|-----------------|-----------------|
| Individual Variables            | Indices          |
| stat. p-value                   | stat. p-value   |
| Log likelihood                  | 1099.71 0.000   |
| AIC                             | -2131.42 0.000 |
| Multicollinearity condition number | 658.92 0.000 |
| Koenker-Bassett test            | 78.28 0.000    |
| Moran's I (for residual)        | 0.05 0.000     |
| Robust LM (error)               | 0.000009 0.929 |
| Robust LM (lag)                 | 0.000000 0.131 |

| TABLE 3 |
| Regressions without Any Temporal Consideration |
|---------------------------------|-----------------|-----------------|-----------------|
| Variables                        | Standard regression | Spatial lag model | Spatial error model |
|                                 | Coef. p-value    | Coef. p-value   | Coef. p-value   |
| Constant                        | 0.055 0.000      | 0.048 0.002     | 0.054 0.000     |
| Demographic Index 1990          | 0.018 0.000      | 0.018 0.000     | 0.018 0.000     |
| Livability Index 1990           | 0.011 0.000      | 0.011 0.000     | 0.011 0.000     |
| Accessibility Index 1990        | -0.014 0.000     | -0.014 0.000    | -0.014 0.000    |
| Desirability Index              | 0.006 0.057      | 0.006 0.064     | 0.006 0.066     |
| Developability Index            | 0.064 0.002      | 0.064 0.002     | 0.065 0.001     |
| Spatial parameter (λ)           | /                 | 0.064 0.135     | 0.063 0.147     |

**Measures of fit**

- Log likelihood: 899.45, 900.55, 901.58
- AIC: -1786.9, -1787.11, -1791.16
### TABLE 4
Regressions with Temporal Consideration of Population Change

<table>
<thead>
<tr>
<th>Variables</th>
<th>Standard regression</th>
<th>Spatial lag model</th>
<th>Spatial error model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>p-value</td>
<td>Coef.</td>
</tr>
<tr>
<td>Constant</td>
<td>0.061</td>
<td>0.000</td>
<td>0.056</td>
</tr>
<tr>
<td>Population change 1980-90</td>
<td>0.277</td>
<td>0.000</td>
<td>0.276</td>
</tr>
<tr>
<td>Demographic Index 1990</td>
<td>0.012</td>
<td>0.000</td>
<td>0.012</td>
</tr>
<tr>
<td>Livability Index 1990</td>
<td>0.009</td>
<td>0.000</td>
<td>0.009</td>
</tr>
<tr>
<td>Accessibility Index 1990</td>
<td>-0.011</td>
<td>0.000</td>
<td>-0.011</td>
</tr>
<tr>
<td>Desirability Index</td>
<td>0.006</td>
<td>0.057</td>
<td>0.006</td>
</tr>
<tr>
<td>Developability Index</td>
<td>0.051</td>
<td>0.011</td>
<td>0.051</td>
</tr>
<tr>
<td>Spatial parameter ($\lambda$)</td>
<td>/</td>
<td>/</td>
<td>0.049</td>
</tr>
</tbody>
</table>

**Measures of fit**

<table>
<thead>
<tr>
<th></th>
<th>Log likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population change 1980-90</td>
<td>942.28</td>
<td>-1870.56</td>
</tr>
<tr>
<td>Demographic Index 1990</td>
<td>944.28</td>
<td>-1868.42</td>
</tr>
<tr>
<td>Livability Index 1990</td>
<td>945.60</td>
<td>-1871.19</td>
</tr>
<tr>
<td>Accessibility Index 1990</td>
<td>944.28</td>
<td>-1867.72</td>
</tr>
</tbody>
</table>

### TABLE 5
Regressions with Temporal Considerations of Population Change and Indices

<table>
<thead>
<tr>
<th>Variables</th>
<th>Standard regression</th>
<th>Spatial lag model</th>
<th>Spatial error model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>p-value</td>
<td>Coef.</td>
</tr>
<tr>
<td>Constant</td>
<td>0.056</td>
<td>0.000</td>
<td>0.051</td>
</tr>
<tr>
<td>Population change 1980-90</td>
<td>0.275</td>
<td>0.000</td>
<td>0.274</td>
</tr>
<tr>
<td>Demographics Index 1990</td>
<td>0.004</td>
<td>0.387</td>
<td>0.004</td>
</tr>
<tr>
<td>Demographic Index 1980</td>
<td>0.008</td>
<td>0.067</td>
<td>0.008</td>
</tr>
<tr>
<td>Livability Index 1990</td>
<td>0.009</td>
<td>0.108</td>
<td>0.009</td>
</tr>
<tr>
<td>Livability Index 1980</td>
<td>0.001</td>
<td>0.873</td>
<td>0.001</td>
</tr>
<tr>
<td>Accessibility Index 1990</td>
<td>-0.021</td>
<td>0.121</td>
<td>-0.021</td>
</tr>
<tr>
<td>Accessibility Index 1980</td>
<td>0.010</td>
<td>0.462</td>
<td>0.010</td>
</tr>
<tr>
<td>Desirability Index</td>
<td>0.007</td>
<td>0.025</td>
<td>0.007</td>
</tr>
<tr>
<td>Developability Index</td>
<td>0.057</td>
<td>0.005</td>
<td>0.057</td>
</tr>
<tr>
<td>Spatial parameter ($\lambda$)</td>
<td>/</td>
<td>/</td>
<td>0.049</td>
</tr>
</tbody>
</table>

**Measures of fit**

<table>
<thead>
<tr>
<th></th>
<th>Log likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population change 1980-90</td>
<td>944.21</td>
<td>-1868.42</td>
</tr>
<tr>
<td>Demographic Index 1990</td>
<td>944.86</td>
<td>-1867.72</td>
</tr>
<tr>
<td>Livability Index 1990</td>
<td>945.60</td>
<td>-1871.19</td>
</tr>
</tbody>
</table>
FIGURE 1. Theoretical Framework of Population Change
FIGURE 2. Population Density (persons/km$^2$) from 1970 to 2000 in Wisconsin
FIGURE 3. Demographic Index in 1980 and 1990 in Wisconsin

FIGURE 4. Livability Index in 1980 and 1990 in Wisconsin
FIGURE 5. Accessibility Index in 1980 and 1990 in Wisconsin

FIGURE 6. Natural Amenities and Desirability in Wisconsin
FIGURE 7. Undevelopable Lands and Developability in Wisconsin