The Cascade GIS Diffusion Model for Measuring Housing Absorption by Small Area with a Case Study of St. Lucie County, Florida

*Grant Ian Thrall**
*Charles F. Sidman**
*Susan Elshaw Thrall***
*Timothy J. Fik**

Abstract. **The geographer's spatial diffusion theory is combined with Geographic Information Systems (GIS) technology to provide a new framework for predicting residential single-family development patterns. We refer to the model as a multiple-stage "Cascade" GIS diffusion model. Parameter calibration is done using two-stage least squares. The model predicts new housing built and purchased by small submarket. Our example submarket is at the census tract level; a smaller submarket could have been chosen. The contribution to housing forecasting literature is a structural model that captures the spatial diffusion process at various geographical scales. Model estimation and forecasts are facilitated using GIS technology via a high resolution and high precision database using county property tax rolls.**

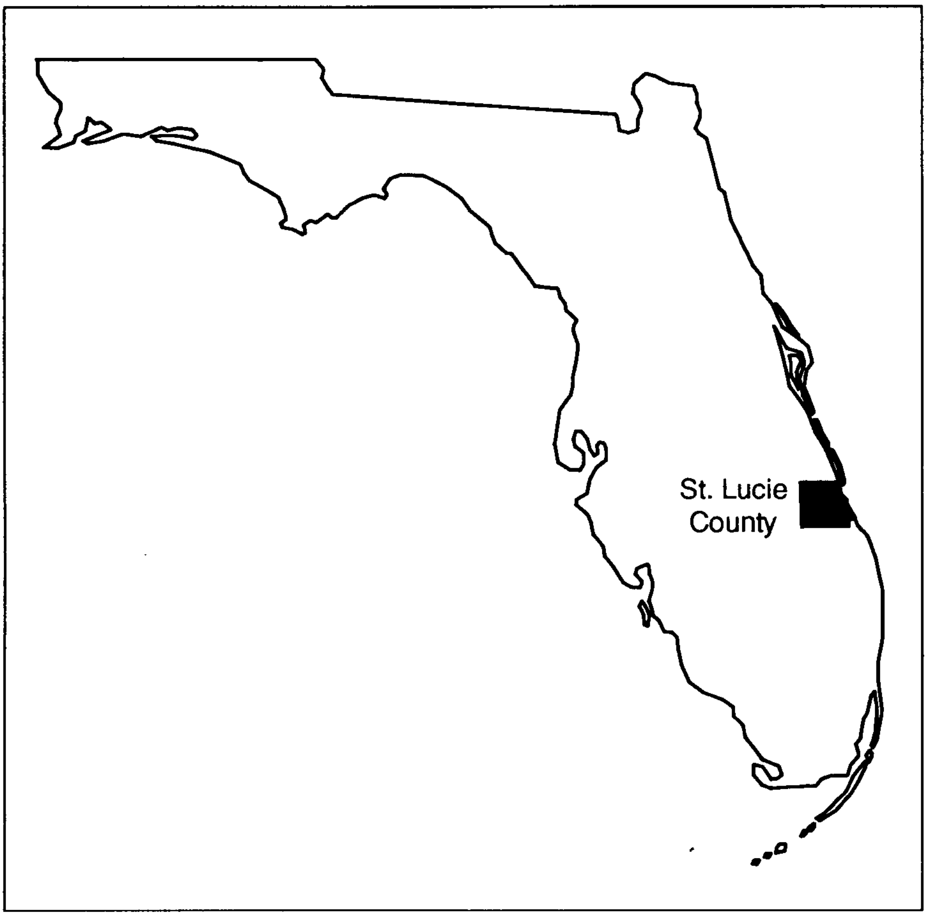
Introduction

This paper introduces a method for forecasting housing absorption for a small target area. We refer to the method as a "Cascade GIS diffusion model." The model builds upon the already extensive geographic literature on spatial diffusion. Spatial diffusion studies are the systematic analysis of the spread of a phenomena across a landscape (Morrill et al., 1988) which are influenced by contagion and/or hierarchical processes. The diffusion phenomena examined here are the development of new single-family dwellings, and the timing and location of that development. The target area could be a quadrat (Boots and Getis, 1988), census tract, census block, transportation analysis zone, planning district, or other small area.

Cascading refers to the mixture of several geographic scales, and the identification of how phenomena at those different scales are linked. For instance, when population increases in the county, some of that population is captured by the neighborhood, and part of the neighborhood increase, in turn, is captured by the target site within the neighborhood. For the purpose of this study, the neighborhood is comprised of all census tracts adjacent to a target census tract. Adjacency is based on shared borders. The geographic impact of population increase is conceptually cascading downward

*Department of Geography, University of Florida, Gainesville, Florida 32611.
**Computer Science, Lake City Community College, Lake City, Florida 32055.
Date Revised—March 1993; Accepted—June 1993.

Exhibit 1
Location of St. Lucie County, Florida



from large scale to small. This Cascade diffusion model builds upon the premise that population growth drives county housing growth, county housing growth subsequently drives the growth of housing in the "neighborhood," which in turn drives the growth of housing in the smaller geographic/target areas.

Geographic Information Systems (GIS) is used to manage geographically referenced data. A GIS is an organized collection of computer hardware, software, geographic data, analytic techniques and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information. A GIS is an integrated system that links data to graphics or digital maps to allow spatial analysis of geographical map features. Until recently, cost and technological limitations prohibited large-scale, data-intensive urban analysis. This limited most empirical analysis and subsequent theoretical advancements to examining and being based primarily upon relative measures of location (e.g., distance from the central business district—CBD). Advancements in computer hardware, GIS software, and spatial reasoning now enable the researcher to process and analyze large data sets where the attributes of the objects are positioned—absolutely—with high precision on a map (Thrall and Marks, 1993).

The Cascade diffusion model presented here can be operationalized using appropriate data for any region, and any number and size of cascading scales of resolution. To illustrate the model, spatially referenced housing data for St. Lucie County, Florida, is used. For the purposes of this study, the county was chosen as the geographic unit of lowest resolution. The intermediate scale is comprised of census tracts that border the target census tract, and the high resolution target area for which the forecasts are made is the census tract. Forecasts are made for all census tracts in St. Lucie County. Parameters for the model are estimated using data up through 1990 and predictions are made for the target areas for 1991. To evaluate the accuracy of the model, 1991 predictions for each target area are compared to the observed or actual housing absorption for 1991 derived from the property appraisal file.

Background—St. Lucie County Study Area

There are about 13 million permanent residents in Florida. Four million persons are projected to be added between 1990 and the year 2000. Much of this growth is occurring in southern coastal areas (*Florida Statistical Abstracts*, 1990). St. Lucie County is in the core of the state's high growth region, located along the population dense southeastern coast.

There are 143,214 persons in St. Lucie County. From 1980 to 1989, St. Lucie County sustained a 72% increase in the number of households, and a 56% increase in total population. This translates into a sustained average increase in population of approximately 5,600 residents per year during the past decade. It is anticipated that 200,000 people will reside in the County by the year 2000, suggesting that there will be a 40% increase in St. Lucie County's population during the 1990s. A permanent population of 318,650 (382,380 including seasonal population) is projected for the year 2015 by the comprehensive plan of the County. The population is dispersed across the County as shown in Exhibit 2.

Exhibit 2
Population by Incorporated Unit in St. Lucie County

City	1980	1990	% Change
Fort Pierce	33,802	39,057	15.5
Port St. Lucie	14,690	48,051	227.1
Unincorporated	38,099	55,505	45.7

Source: Florida Bureau of Economic and Business Research

Background—Theories of Land Use

A variety of theories have been used to *describe, explain* and *predict* urban morphology. When combined with appropriate calibration and data analysis, these theories can be used to *prescribe* actions that fit within an envelope of objectives. These theories can be classified into two general categories: (a) those that depict the change in land use and land values in terms of absolute locations (as measured by latitude and longitude or a coordinate system); and, (b) those that use relative locations in relation to some reference point, such as x miles from another important location on the map.

The literature that uses *absolute spatial location* to measure land use change includes the work of Burgess (1925). His concentric zone model has rings of distinct land uses radiating outward from the central business district. The "sectorial" model introduced by Hoyt (1939) shows urban form as shaped by land use development that proceeds along transportation corridors, usually radiating out from the urban core. Hoyt's model depicts a high level of homogeneity of land use within each corridor or sector. Harris and Ullman's (1945) "Multiple Nuclei" model allows for a more complicated pattern of land use: growth arises from spatially distributed activity centers, or nodes, and land use patterns—perhaps like those advanced by Burgess or Hoyt—unfold around each node.

Literature that uses *relative spatial location* to measure land use and land value change includes that of Alonso (1964), Muth (1969) and Thrall (1987, 1991). The approach of each author is different, but the overall effect is to provide a stylized model of land value and land use change. These models generally include the specification of an exogenous parameter or instrument (e.g., Thrall identifies the instrument typically as either a "budget constraint shifter" or a "utility shifter"). The values of the endogenous variables of the model can be changed by altering the value of definition of the instrument. As the values change, the general trajectories of the endogenous variables are revealed giving rise to a heuristic explanation and expectation of how land use and land values change. Based upon relative measures of location, the trajectory of change may be in one direction at the CBD, and in the same or other direction at a node located some distance from the CBD. These models are useful for heuristic purposes and until recently have been our best window into understanding urban morphology.

In addition to the above models, a variety of alternative approaches have been offered to explain the underlying processes that influence the evolution of the urban landscape. Lowry (1964) examines the interaction of land use with transportation to forecast residential land uses. Essential to Lowry's model is the assumption that demand for residential property increases with accessibility to employment opportunities. Lowry's model is driven by population change and a variety of economic indicators.

The "settlement spheres" approach taken by Sargent (1972) suggests that land use patterns are shaped by land investors taking advantage of the creation of transportation networks. The transportation network provides the feasible urban boundary, while land speculation provides the sequence, location and timing of development. Sargent (1972, p. 357) believes residential land use to be "commonly the leading edge of urban expansion," only to be followed later by commercial activity.

Background—The Geographer's Paradigm of Spatial Diffusion

The models discussed above have not proven entirely satisfactory for prediction as they lack components that link spatial and temporal aspects of processes in a geographic hierarchy. The geographer's diffusion paradigm, according to Morrill et al. (1988), can readily be applied to the study of urban growth as an alternative approach. When mapped, the diffusion process might look in some respects similar to the concentric zone model proposed by Burgess (1925). However, the diffusion process is dynamic in that it incorporates both spatial and temporal components. Interestingly, few researchers have applied diffusion theory to the study of metropolitan growth (see Berry, 1972, on growth diffusion between regions). It is in this application of diffusion theory that Morrill et al. (1988) believe there to be the greatest potential. Morrill et al. (1988) explicitly call for the diffusion paradigm to be used to monitor and forecast urban growth, and it is this subject area, they maintain, that should receive priority for further advancement of the diffusion literature. Our research here is an acceptance of the challenge posed by Morrill et al. (1988).

Diffusion as a Theoretical Framework: Hägerstrand's Model

Many geographers have contributed to the development of diffusion literature over the past one hundred years. Most important to the formulation of the mathematical approach followed here is the pioneering work of Hägerstrand (1952, 1965, 1967a, 1967b). Hägerstrand formally modelled the diffusion or spread of a given phenomenon, from a place of origin over space and through time (see Morrill et al., 1988), explicitly including the temporal and spatial elements of the diffusion process. Hägerstrand's analysis captures the "spatial bias" of the "neighborhood effect," where once the phenomenon is introduced onto the landscape, the phenomena spreads "contagiously" and outward from where it was introduced (much like the rippling effect of water when a stone pierces its surface).

Morrill was one of the early geographers to use diffusion theory to explain and predict urban land use and land value change, and to apply these concepts to urban housing. Morrill (1965) (see also Alves and Morrill, 1975) demonstrated that populations inhabiting housing submarkets expand or contract their areas of dominance in a manner that could be explained and predicted using the general processes of diffusion (see also O'Neill, 1981). Further advancements have also been made by Brown (1968, 1975, 1981) who introduced macro-oriented diffusion modelling, and Boyce (1966) who provided extensions of Hägerstrand's general theoretical framework, drawing upon population density gradients such as those proposed by Clark (1951) to explain urban expansion. Boyce's "Wave Theory Analog Approach" likens urban growth to that of ocean waves—successive flows of people spreading outward in decreasing densities from the current urban periphery. Morrill (1968) also characterized the diffusion process as a "wave-like phenomenon." These approaches can be applied to both commercial real estate site selection studies and analysis of real estate markets.

Summaries of the effect of diffusion processes are generally made using a logistic "S" curve such as

$$HT = k/(1 + e^{\alpha + \beta t}) , \tag{1}$$

where HT can be the cumulative sum of houses built by time period t within a census tract. The component k is the asymptote measuring, say, the carrying capacity, or maximum number of housing units that can be built in a given time period because of market conditions, zoning, or engineering considerations. Parameters α and β, which characterize the dynamics of the diffusion process, are to be estimated. The literature provides many examples of diffusion processes conforming to a logistic "S" function (Griliches, 1957; Casetti, 1969; Morrill, 1970; O'Neill, 1981; Thrall, 1983).

Since neighborhood or locational effects are a central feature of real estate and real estate development, they must be explicitly represented in any real estate diffusion model. Equation (1) must then have structural error since its summary glosses over neighborhood effects giving rise to parameter estimates that are biased in an unknown direction (Thrall, 1988). Structural error of equation (1) results in haphazard forecasts. In addition, an accurate measurement of k, or the carrying capacity, is difficult to obtain. This is especially problematic when counties and cities oversupply residential land. Estimation error arises when k is much greater than the cumulative development, which is the usual case. To overcome the structural problems associated with equation (1) for small area real estate and related analyses, we offer the Cascade diffusion model.

The Cascade Diffusion Model

The Cascade diffusion model is a variation on Hägerstrand's innovation diffusion model. The model is parsimoniously comprised of the following elements:

P = population estimates for the county;
HC = cumulative sum of single-family houses in the county;
HN = cumulative sum of houses in the neighborhood excluding the target tract (the neighborhood is defined as all tracts adjacent to the target tract);
HT = cumulative sum of houses in the target tract;
α = the existing number of homes at the beginning of the study period (intercept);
β = the development rate over time (slope).

The non-recursive Cascade diffusion model is expressed structurally as

$$\ln HC = \alpha_1 + \beta_1 {*} \ln P , \tag{2}$$

$$\ln HN = \alpha_2 + \beta_2 {*} \ln HC , \tag{3}$$

$$\ln HT = \alpha_3 + \beta_3 {*} \ln HN . \tag{4}$$

In equations (2) through (4) the dependent variable of equation (2) becomes an independent variable in equation (3), and the dependent variable of equation (3) becomes an independent variable in equation (4). This simultaneous system of three equations is estimated using two-stage least squares (2SLS). Essentially, in 2SLS, parameters estimates for HC in equation (2) are estimated; the resulting estimated HC is used as an independent variable in equation (3). In turn, estimated values of HN in equation (3) are used as an independent variable in equation (4).

Equation (2) postulates that the product added (single-family dwellings) in the county-wide area is proportional to the population growth of the county. Equation (3) expresses HN as proportional to HC thereby implying that each neighborhood of the county captures part of the total product added. Lastly, equation (4) expresses the small target area as capturing part of the total development of its immediate surrounding neighborhood, while the residual neighborhood development is spatially distributed elsewhere in the neighborhood.

Using equations (2) through (4), the reduced-form equation can be shown to be:

$$\ln HT = \alpha_3 + \beta_3\{\alpha_2 + \beta_2[\alpha_1 + \beta_1(\ln P)]\} . \tag{5}$$

The reduced-form equation is nonlinear with population remaining as the sole independent variable. The value and arrangement of the parameters in equation (5) allows for forecasts to take into consideration the neighborhood effect. Projections for yearly absorption are calculated by estimating coefficients of equations (2) through (4), and substituting the appropriate values into equation (5). Forecasts are created by substituting estimates for P for the appropriate year for which the forecast is desired.

Because neighborhood effects are a central ingredient to growth of a small area, then parameter estimates and forecasts that are based upon a model that does not include neighborhood effects like equation (1) will produce parameter estimates that are biased in an unknown direction. Likewise, because the important neighborhood effect is not accounted for, a transformation of the model in equation (5) to be

$$\ln HT = \alpha + \beta(\ln P) , \tag{6}$$

or

$$\ln HT = \alpha + \beta(\ln t) , \tag{7}$$

would also produce spurious parameter estimates and unreliable forecasts. To sum up, a model of the phenomena investigated here is erroneous if the model does not include geographic scale effects.

The Data

The data on the number of single-family housing units built and sold by year and location comes from the St. Lucie County property tax appraisal computer tapes. Property appraisal data were georeferenced and imported into Caliper Corporation's GIS software **GisPlus**; the method of georeferencing and database design used here is detailed elsewhere (Thrall and Elshaw Thrall, 1991). The database was created with property information current as of April 1992. The "tax tapes" contained data on every real property in the county. Single-family dwellings were selected out from the larger database using the revenue code for the property. There were 51,644 single-

Exhibit 3a
St. Lucie County 1980 Census Tracts

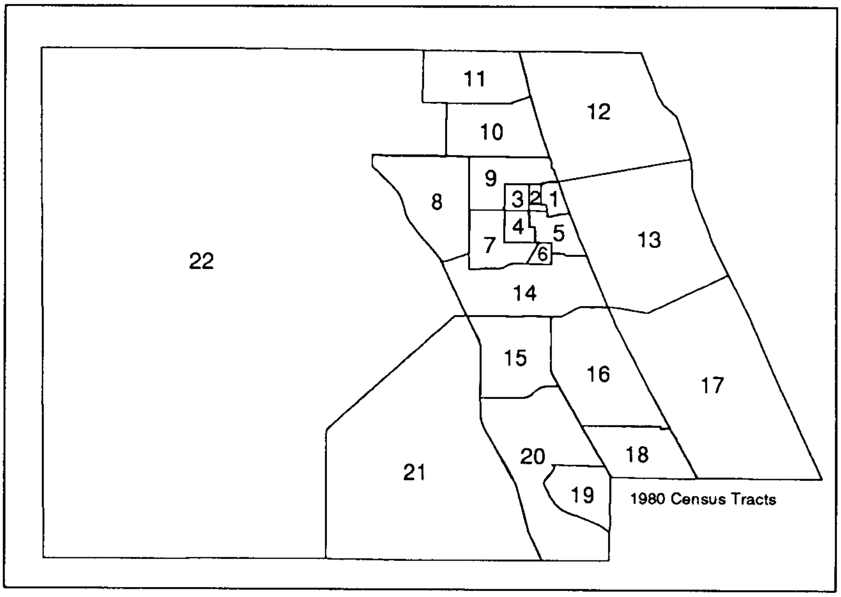
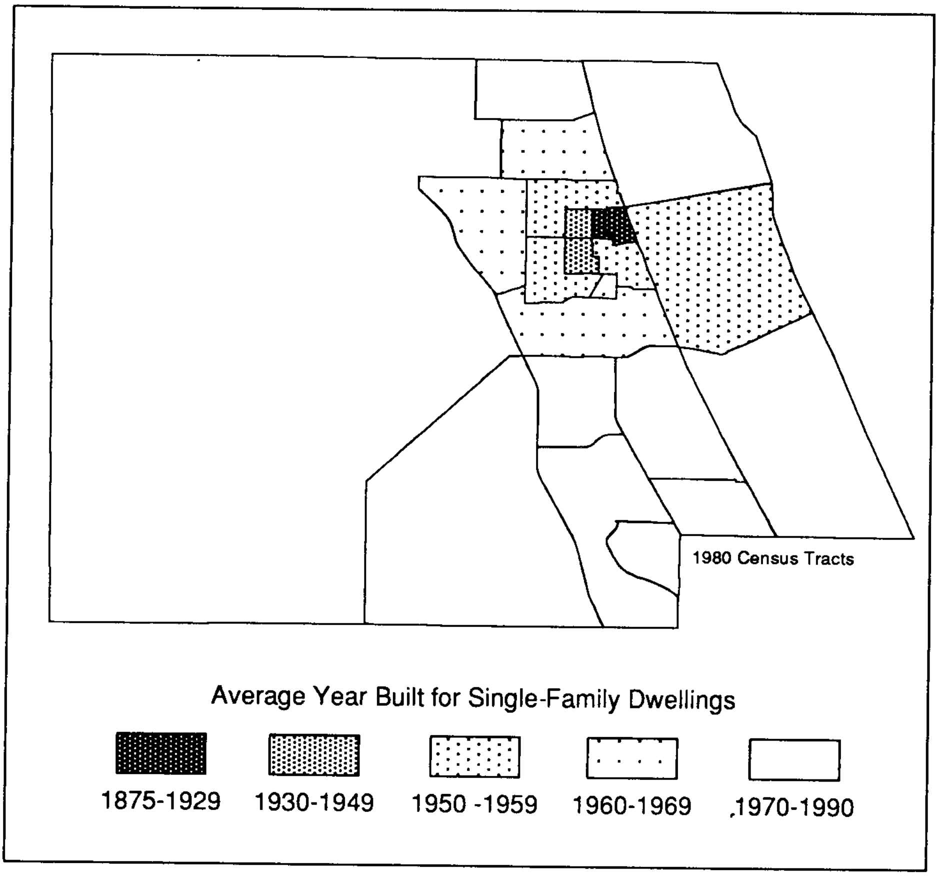


Exhibit 3b
Individual Tracts and Their Respective Neighborhoods

Census Tract	Nearest Neighbors (Neighborhood)
1	2, 5, 9, 13
2	1, 3, 5, 9
3	2, 4, 9
4	2, 3, 5, 7
5	1, 2, 4, 6, 13, 14
6	4, 5, 7, 14
7	4, 6, 8, 9, 14
8	7, 9, 10, 14, 22
9	1, 2, 3, 7, 8, 10, 12
10	8, 9, 11, 12, 22
11	10, 12, 22
12	9, 10, 11, 13
13	1, 5, 12, 14, 17
14	5, 6, 7, 8, 13, 15, 16, 22
15	14, 16, 20, 21
16	14, 15, 17, 18, 20
17	13, 16, 18
18	16, 17, 19, 20
19	18, 20
20	15, 16, 18, 19, 21
21	15, 20, 22
22	8, 10, 11, 14, 21

Exhibit 4
Contagion Diffusion Process of Housing Development



family dwellings at the end of the 1991 study period, all of which were georeferenced and included in the analysis.

United States Census Bureau TIGER/line files provided the geographic objects necessary for the analysis. This included census tract, municipal and county boundaries. These geographic objects are an integral part of this GIS analysis. Census tracts were selected as the geographic unit of measure because the areal unit is known to both developers and planners. Instead of 1990 census tracts, the 1980 census boundaries were used because they offered sufficient resolution for the purposes of illustrating small area forecasting; and, there were only twenty-two tracts in 1980 as compared to thirty-seven tracts in 1990 (for a larger discussion on this issue see Thrall, 1992). Each census tract target area has its own unique set of neighbors. Exhibit 3a shows the location of the census tracts, and the set of census tracts that form the neighborhood for each individual target census tract for St. Lucie County. The GIS software was used to count the number of homes built each year, by census tract, and to aggregate the counts together as necessary to form counts for neighborhoods (see Exhibit 3b).

The time sequence of development of census tracts is shown in Exhibit 4. The average age of all single-family dwellings is calculated using the GIS for each census tract. Exhibit 4 clearly reveals a contagion diffusion process in which growth emanates outward from a central urban core. Housing development began initially in the core of the City of Fort Pierce, and then expanded outward. Observation of a spatial pattern like that of Exhibit 4 is sufficient justification for that growth being represented as a spatial diffusion process (Morrill et al., 1988). The spatial description of Exhibit 4 and the calculations required for the model estimation could not have been performed without GIS technology. For example, to calculate age of dwellings by census tract the GIS is first used to geoposition the observation within the census tract. Afterwards, the GIS is used to average all observations that fall within the borders of the census tract (a point-polygon operation).

Many census tracts had no, or very few, single-family houses prior to a recent base year. Each target tract was therefore assigned a base year to reflect when the census tract began some level of development. Modeling the timing of census tract takeoff will be the subject of subsequent research. For this study, years prior to the observed period of takeoff were truncated from the analysis. All census tracts were observed to fall into one of three "base year" groups; census tracts were assigned a base year of 1970, 1973 or 1978. The number of observations per census tract then ranged from the most of 1991–1970 = 21, and from the least of 1991–1978 = 13. To reduce bias that could enter a time series analysis covering only a few time periods the Beach and Mackinnon (1978) maximum likelihood estimation procedure was used to calculate the 2SLS regression coefficients; their algorithm also minimizes the effects of positive serial autocorrelation of the error terms.

The data were further divided into an historical period (covering the years 1970 through 1990) and a validation period (1991). For most census tracts, 1 through 14, and 16 through 19 (as illustrated in Exhibit 5), data truncated to a base year of 1970 ($n = 21$) provided acceptable validation period predictions. Validation period results were greatly improved when truncation resulted in the number of time periods being reduced to $n = 18$ (tracts 15, 20, 22), and $n = 13$ (tract 21). *HC, HN* and *HT* for the base year of the analysis were derived by calculating the total number of houses that had been constructed up through the base year. With our technology, we could have used 1900 or an even earlier base year. However, the shorter time sequence reduced positive serial autocorrelation and was found to provide better housing absorption estimates for the 1991 validation period. An implication is that detail of the 1900 housing market has little bearing on the 1991 housing market, and, since the model is driven by cumulative population estimates and development figures, the effects of historical population and housing growth are automatically factored into each successive time period. Essentially, our "base year" demarcates when the census tract began development takeoff.

A future research agenda is to expand this line of geographic reasoning to include forecasts as to when small areas such as census tracts begin their development take-off. Another agenda is to include a variety of other structural variables in our parsimonious model that take into account the demographic and income profile, and changes in economic base; this should allow for forecasts not only of housing counts, but also of housing by product type (price range, square footage, and so on).

Historical population figures prior to 1970 were obtained from the *1980 Census of*

Exhibit 5
Cascade Diffusion Model: First Structural Equation

$$HC = e^{\alpha_1} P^{\beta_1}$$
$$\ln HC = \alpha_1 + \beta_1 \ln P$$

Census Tract	α_1	β_1	R^2	D–W
1–14 and 16–19	−4.17101	1.25211		
*$n=21$	(−8.521)	(29.172)	.990970	1.59585
15, 20, 22	−4.67397	1.29537		
**$n=18$	(−6.506)	(20.726)	.987613	.91108
21	−3.42094	1.18904		
***$N=13$	(−6.961)	(28.088)	.993610	.50708

*years 1970–1990 used in evaluation
**years 1973–1990 used in evaluation
***years 1978–1990 used in evaluation
t-values are in parentheses.

Population: Florida Chapter A: Number of Inhabitants, Vol. 1, Pt. II, 1982, United States Department of Commerce. Yearly estimates from 1970 through 1991 are from *Florida Statistical Abstracts*. Projected population estimates are from *Population Projections by Age, Sex and Race for Florida and its Counties, 1991–2010* (1992), Vol. 25, Nos. 3–4. The geometric mean was used to estimate population for time periods in which data was not available (Norcliffe, 1977, p. 44).

Empirical Results

Coefficients for equations (2) through (4) were calculated in the manner described above. Exhibits 5, 6, 7 summarize the results of calibrating the Cascade diffusion model. The coefficients α_i and β_i for $i=1,2,3$, from Exhibits 5, 6 and 7 are substituted as appropriate into the reduced form equation (5) to derive the forecasts.

In Exhibits 5 through 7, t-values are shown in parentheses beneath the estimated coefficient. With the number of degrees-of-freedom here, the regression coefficients can be regarded as being significant.

The R^2 statistic is usually drawn upon to show how well an estimated regression line fits the data. However, in the 2SLS technique the R^2 is not defined. We follow the convention here that when using 2SLS the R^2 statistic is reported as calculated by a standard regression, independent of the other equations in the system. Values of R^2 close to 1.0 are indicative of a good fit of the model to the data, heteroscedasticity and related issues aside.

A Durbin-Watson "D" statistic is also provided in Exhibits 5, 6 and 7. Durbin-Watson values greater than 1.08 (for $n=15$ observations), 1.16 ($n=18$ observations), and 1.22 (for $n=21$ observations) indicate little or no serial autocorrelation at the 95% confidence level (Chatterjee and Price, 1991).

Exhibit 6
Cascade Diffusion Model: Second Structural Equation

ln $HN = e^{\alpha_2} HC^{\beta_2}$

Census Tract	a_2	b_2	R^2	D-W
1	5.85334 (18.708)	.235985 (7.746)	.934745	.45399
2	6.56534 (27.688)	.181541 (7.744)	.933132	.17570
3	5.10239 (15.287)	.299131 (9.064)	.949675	1.28067
4	7.31328 (85.037)	.118968 (13.991)	.980033	1.26433
5	4.77886 (55.240)	.357429 (41.776)	.991533	2.12274
6	5.62909 (58.796)	.293175 (30.962)	.990156	2.00208
7	4.44984 (17.740)	.399872 (16.118)	.977780	1.61545
8	3.65741 (8.656)	.496609 (11.885)	.965616	1.30852
9	6.38965 (47.073)	.203414 (15.156)	.981375	1.69390
10	−.118696 (−.149)	.831955 (10.542)	.963297	1.29674
11	−.828455 (−.571)	.814457 (5.678)	.896769	.97960
12	1.35522 (2.735)	.679862 (13.878)	.979749	1.45907
13	3.92225 (31.522)	.406177 (33.005)	.991801	1.69471
14	−.405409 (−.727)	.950756 (17.233)	.977388	1.40031
15	−4.67397 (−6.506)	1.29537 (20.726)	.987613	.91108
16	−6.67268 (−9.709)	1.55184 (22.831)	.987387	1.64260
17	−8.57603 (−6.757)	1.63151 (12.998)	.972138	1.44043
18	−7.77532 (−5.086)	1.63504 (10.824)	.968230	1.35737
19	−9.54902 (−6.093)	1.76738 (11.409)	.974757	1.55217
20	−8.46289 (−8.050)	1.69657 (16.492)	.978132	.79570
21	−4.88537 (−9.138)	1.37435 (26.696)	.991478	.50708
22	−3.65852 (−8.278)	1.17553 (27.165)	.990793	1.48046

t-values are in parentheses.

Exhibit 5 provides estimated 2SLS parameters for equation (2) where $HC = f[P]$; also reported are the degrees of freedom (number of years of data remaining after the truncation) for each target census tract, the Durbin-Watson "D" statistic for the transformed data, and the R^2 statistic. Exhibit 6 shows the 2SLS parameter estimates

Exhibit 7
Cascade Diffusion Model: Third Structural Equation

$$HT = e^{\alpha_3}\, HN^{\beta_3}$$
$$\ln HT = \alpha_3 + \beta_3 \ln HN$$

Census Tract	a	b_2	R^2	D–W
1	2.21800	.312944	.917530	1.64247
	(6.599)	(7.704)		
2	3.17392	.391529	.977553	1.53437
	(15.532)	(16.093)		
3	5.12253	.267138	.953600	.79849
	(23.029)	(9.757)		
4	−16.7637	2.80406	.964968	1.93788
	(12.985)	(18.496)		
5	5.265	.231133	.987391	1.89035
	(72.418)	(26.675)		
6	−.550776	.825871	.885976	1.36986
	(−.430)	(5.536)		
7	2.88629	.523881	.956498	1.29511
	(6.492)	(10.004)		
8	−3.58578	.986968	.984538	1.88356
	(−12.188)	(29.109)		
9	−8.99166	1.93010	.860613	−.08745
	(−3.049)	(5.527)		
10	2.81204	.417769	.969746	1.49255
	(7.899)	(9.740)		
11	−.828455	.814457	.896769	.97960
	(−.571)	(5.678)		
12	1.3522	.679862	.978749	1.45907
	(2.735)	(13.878)		
13	3.92225	.406177	.991801	1.69471
	(31.522)	(33.005)		
14	−.405409	.950756	.977388	1.40031
	(−.727)	(17.223)		
15	.288179	.841812	.963162	1.06639
	(.449)	(11.983)		
16	−2.78602	1.14180	.938727	1.09177
	(−2.340)	(8.679)		
17	−8.57603	1.63151	.972138	1.44043
	(−6.757)	(12.988)		
18	−5.55127	1.29581	.886890	1.12408
	(−2.602)	(5.338)		
19	2.04208	.534791	.974757	1.55217
	(6.213)	(13.569)		
20	−.207214	.975881	.949640	.73457
	(−.217)	(9.132)		
21	−12.9750	2.11704	.964939	1.07859
	(−7.824)	(11.986)		
22	−1.44404	1.0005	.752412	1.01852
	(−.658)	(3.818)		

t-values are in parentheses.

for equation (3), and Exhibit 7 shows the same for equation (4). A strong and positive correlation is shown between the population of the county and the number of houses built in the county, between the houses built in the county and the houses built in the neighborhood market, and between the neighborhood housing market and that captured by the target small area.

Autoregressive Integrated Moving Average (ARIMA) models were also estimated for each census tract. Although the ARIMA model can provide an effective forecast over short periods, its accuracy deteriorates quickly as the span of the forecast period is increased. On the one hand, the short-term-forecasting ARIMA approach may be simpler than the Cascade model as it only requires a univariate time series. On the other hand, ARIMA has an inability to capture spatial processes, socioeconomic conditions, and demographic change (e.g., geographic spillover effects, economic cycles, and population growth) all within the same model; this contributes to reducing ARIMA's applicability as a medium- and long-term estimator.

The multidimensional and multivariate features of the Cascade model makes it more intuitively appealing than the ARIMA model. Potentially, the Cascade model is a much better overall forecasting tool. An obvious advantage that the Cascade (equations (2) through (4)) has over logistic (equations (1) and (7)) and ARIMA models is the Cascade's inclusion of the neighborhood effect. However, if neighborhood effects are not significant, then the Cascade may not yield results superior to the ARIMA model. Indeed, the Cascade approach will not perform as well as an ARIMA approach when the target area does not adhere to normal localized phenomena, including when the target area is independent of the growth processes influencing the county and neighborhood. This can also arise when an error is made in the selection of the scale (size) of the target area. If the size of the target area is too large, the target area can encompass those neighborhoods that would otherwise be influencing its development had the target area been defined with a higher resolution (smaller area). Exacerbating this problem would be adjacent areas included in the calibration of equation (3) that do not have a strong bearing on the development of the target area. *As the size of the target area increases, the target area moves beyond the scale of the process itself.*

To determine whether the Cascade model yields results superior to the logistic "S" curve of equation (1), parameter estimates of the linear transformation of equation (1) were first made. Values for the asymptote k were derived from area measurements taken from the county-wide growth management plan as required by the State of Florida Growth Management Act. The Durbin-Watson procedure indicated high positive serial-autocorrelation. Both the Cochrane-Orcutt and maximum likelihood transformation procedures were unable to correct for the severely autocorrelated error structure of the logistic "S" curve. Direct estimates of the nonlinear form of equation (1) can potentially mask the effects of serial autocorrelation; but, parameter estimates were shown to be highly sensitive to starting values required for the nonlinear estimation procedures. Structural and estimation problems of the logistic "S" curve of equation (1) leave that formulation suspect for yielding haphazard results.

The Cascade model did not have the estimation problems inherent with the logistic approach. Importantly, the Cascade model better captures the principles of the diffusion process as it incorporates both the "neighborhood effect" and historical county growth, and it is shown to minimize serial autocorrelation since it avoids the

Exhibit 8
Cascade Diffusion Model: Validation Period Results

	Cumulative Number of Homes as of 1991			
Census Tract	Predicted	Observed	Difference	% Difference*
1	128	127	.7	.55
2	673	670	2.9	.4
3	1,553	1,546	7.3	.47
4	1,548	1,507	40.8	2.7
5	1,424	1,420	4.1	.29
6	822	796	25.9	3.3
7	1,770	1,763	6.9	.39
8	203	219	15.7	7.2
9	1,856	1,877	79.4	4.2
10	675	672	2.6	.39
11	2,853	2,722	131.0	4.8
12	360	350	9.7	2.8
13	596	569	26.5	4.7
14	1,865	2,068	202.0	9.8
15	6,525	8,780	2,254.0	25.6
16	6,131	6,943	811.5	11.7
17	564	250	314.0	125.0
18	1,398	1,385	13.2	.96
19	1,259	1,127	132.6	11.8
20	12,440	11,436	1,004.0	8.8
21	3,223	3,600	376.9	10.4
22	2,013	1,817	196.0	10.8

*difference as a percentage of the observed

direct use of time as an actual regression variable. Instead, the time dimension is indirectly included in the Cascade model by way of the cumulative addition of county, neighborhood and census tract housing and population variables.

Summary results for the 1991 validation period for the Cascade model are shown in Exhibit 8. This table compares predicted values with the actual cumulative number of houses built in each census tract. The absolute difference and percentage difference between predicted and observed are measures of the ability of the model to forecast.

The above discussion stressed the importance of selecting a target area the size of which does not exceed the scale of the process that the Cascade model is intended to measure. Therefore, a test was designed to determine if census tracts represent an appropriate scale of resolution for the Cascade analysis. The test also can determine if some phenomena other than county population increase and neighborhood effects are driving the development of the target area. The test compares the results from the Cascade model to a Box and Jenkins ARIMA model. The ARIMA model is very similar to a model like that of equation (7) where only time is an independent variable, and no neighborhood effects are considered. Predictions using the ARIMA model for the 1991 validation period are displayed in Exhibit 9.

The ARIMA model generates predictions based solely on the historical growth trend of the subject area itself. ARIMA predictions were generated using the Number Cruncher Statistical Systems (NCSS) version 5.3 software program. When the time

Exhibit 9
ARIMA-Time Series Model: Validation Period Results

Census Tract	Cumulative Number of Homes as of 1991 Predicted	Observed	Difference	% Difference*	Auto-Regressive Parameters**	R^2	t-values
1	127	127	0	0	.4107	.9926	.35
2	670	670	0	0	.7458	.9991	10.50
3	1,545	1,546	1	.06	.8687	.9994	15.50
4	1,507	1,507	0	0	.5857	.9978	6.82
5	1,420	1,420	0	0	.7764	.9989	11.62
6	796	796	0	0	.8270	.9989	13.75
7	1,768	1,763	5	.28	.8737	.9993	15.14
8	219	219	0	0	.6126	.9966	6.13
9	1,889	1,877	12	.63	.8528	.9991	15.14
10	672	672	0	0	.2976	.9932	2.92
11	2,753	2,722	0	0	.9838	.9989	20.50
12	359	350	9	2.6	.8374	.9964	8.81
13	574	569	5	.28	.8511	.9984	11.30
14	2,043	2,068	25	1.2	.8350	.9982	12.67
15	8,916	8,780	136	1.5	1.2812	.9970	23.66
16	6,944	6,943	1	.06	1.1187	.9982	26.00
17	250	250	0	0	.0930	.8987	.44
18	1,386	1,385	1	.06	.7145	.9892	9.14
19	1,129	1,127	2	.18	.5856	.9866	4.08
20	11,513	11,546	33	.29	.9529	.9975	13.52
21***	3,655	3,600	55	1.5	AT1 − .6405	.3655	−3.55
					AT2 − .3013		−1.67
22	1,821	1,817	4	.22	.6338	.9932	7.36

*difference as a percentage of the observed
**ARIMA (1, 1, 0) 1st difference model gave best results
***ARIMA (2, 2, 0) 2nd difference model gave best results

horizon of the forecast was short, the ARIMA model provided absorption estimates competitive with the Cascade model. However, in fast-growth tracts, the ARIMA model tended to grossly overpredict, compounding the error for long-term forecasts. The ARIMA model predicts 923,000 homes will be built in tract 15 by the year 2010; this figure translates to a density of 186 residential single-family homes per acre. The Cascade model instead predicts a more reasonable 15,634 houses, translating to a density of 3.15 houses to the acre for the year 2010. This lends further credibility to the Cascade's usefulness in producing reasonable long-term projections.

Overall, the Cascade diffusion model provided good results; the results are especially strong and appealing since the Cascade model is parsimonious; the median error measured by the "% Difference" of "Observed" and "Predicted" columns in Exhibit 8 does not exceed 5%. Parsimony lends credibility to the Cascade approach since the model is based solely upon the spatial relationships of neighborhoods, population estimates and housing counts. The Cascade model lends itself to improved accuracy of forecasting and explanatory power since it can be extended to include indicators other than population such as income, economic indices, interest rates, and the demographic makeup of the households.

The results, shown in Exhibits 8 and 9, indicate that both the Cascade and the ARIMA models provide reasonable estimates for the city of Fort Pierce and the other northern census tracts (tracts 1 through 13). However, the Cascade model was not as reliable for the census tracts comprising the city of Port St. Lucie and its immediate neighborhood (tracts 15, 16, 17, 19, 20, 21). The hypothesis test formulation then suggests that two types of growth patterns are responsible for the manner in which development occurs in St. Lucie County; in other words, for select census tracts, the census tract may not be an adequate representation of the scale of the growth process.

Growth in the northern part of the County, including the city of Fort Pierce, corresponds to a "natural" diffusion process that is captured within the census tracts by the Cascade model. But, development patterns in census tracts within the southern portion of the County do not share the same diffusion process as the northern portion of the County. Instead, a propagated diffusion process controls residential development in the southern portion of the County. Brown (1981) provides the framework for analyzing propagated diffusion patterns.

In the early 1960s all of the land comprising the city of Port St. Lucie was purchased and platted by a single firm, General Development Corporation (GDC). GDC was a multinational firm, marketing and advertising its housing product through its branch offices in the U.S. Midwest, U.S. Northeast, as well as England, Germany and other European countries. Migration to Port St. Lucie was therefore propagated. Customers purchased the house and lot sight unseen; GDC—not local market conditions—then controlled where and when new houses were built. Lots were developed more as a function of the sequence when customers retired than on the basis of where the lot was relative to other development.

An analysis of GDC-held properties in the property appraisers file suggested a correlation between the census tracts in which GDC operated and the tracts where the Cascade model validation estimates suffered. Brown describes propagators like GDC as "Diffusion Agency Establishments" or "Coordinating Propagators" (Brown, 1981, p. 52). The coordinating propagator, in this instance GDC, controls the amount, location, and size of the development. The gross pattern of diffusion is centrally controlled by GDC. The Cascade model instead relies upon diffusion coming about solely through the aggregation of individual actions and decentralized decisionmaking (Brown, 1981, p. 53) The Cascade model is then better suited for areas that experience normal growth. Port St. Lucie's growth has been recent and rapid; it was not incorporated by GDC until the early 1970s. The ARIMA model is better suited to those situations where growth is outside the norm, is controlled by a central force, and when time is the only force (or appropriate surrogate variable) influencing the phenomena. The ARIMA model is a good alternative to the Cascade model for those areas where interfering mechanisms are disrupting the natural growth processes. As noted earlier, projections by the ARIMA model in such a context are limited to the very short term.

Also, the Cascade model as presented here does not include unusual characteristics that would serve as barriers or inducements for development. The coastal islands (tracts 12, 13 and 17) serve as an example of how the environment can *induce* development at a rate greater than expected by looking at the pattern of growth of the larger system. Those coastal islands are part of a different growth process dominated by high-rise multifamily condominiums. Census tract 22 is an example of a barrier to development; the tract is not in the urban service area as defined by the State's

Growth Management Act. Since infrastructure to support housing development will not be installed in census tract 22 in the near term, citrus production is likely to remain as the predominant use of land. Greater explanation, and forecasts of greater accuracy, can be obtained from the Cascade model by installing inducements and barriers in the framework.

One approach for installing inducements and barriers in the Cascade model would be to assign probability-like scores to areas (census tracts) that are adjacent to the target area (census tract). Each target area would be accompanied by a unique array of scores assigned to its neighbors. The greater (lesser) the barrier between a target area and its neighbors, the lower (greater) the score. The expected spillover effect from the neighbor would then be measured as the product of the score and, say, the number of houses in the neighboring tract. The above data analysis has implicitly assumed neither inducements nor barriers thereby implying the value of the score to be equal to 1. Inducements may be captured by assigning a score greater than 1.

The Cascade model, when combined with a knowledge of local conditions, can then provide dependable estimates of growth patterns and can give insight as to why small areas deviate from that which would otherwise be expected.

Conclusion

This study extends the geographer's paradigm of diffusion to the role of being a tool for forecasting expansion in housing markets. This study demonstrates that small area forecasts can be accurately done using a combination of diffusion theory and geographic information systems (GIS) technology. The Cascade diffusion model can be used for monitoring and predicting growth patterns of residential single-family development, but is not limited to that real estate product.

The Cascade diffusion model is flexible and can be applied to commercial, industrial and multifamily land uses. Furthermore, additional explanatory factors can be added to its formulation to include, for example, household income, interest rates, age profile and other demographic characteristics, as well as local economic variables. The Cascade model can be adjusted to include "inducements" and "barriers" to growth. The Cascade model can be extended to include more than three scales of resolution—upwards to include larger regions such as groups of counties, downwards to target areas smaller than a census tract such as a subdivision, and to intermediate scale layers such as transportation analysis zones. The Cascade model can also be extended to forecast product by subcategory, such as number of houses by price range.

The applications for this information are wide-ranging. Developers, lending institutions, and real estate investors can significantly benefit from the use of the methodology presented here. An understanding of spatial market forces reduces development risk; the results affirm the importance of geographic reasoning to forecasting urban growth and real estate-related phenomenon.

References

Alonso, W., *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge, Mass.: Harvard University Press, 1964.

Alves, W. R. and R. L. Morrill, Diffusion Theory and Planning, *Economic Geography*, 1975, 51, 290–304.

Beach, C. M. and J. G. MacKinnon, A Maximum Likelihood Procedure for Regression with Autocorrelated Errors, *Econometrica*, 1978, 46, 51–58.

Berry, B. J. L., Hierarchical Diffusion: The Basis of Developmental Filtering and Spread in a System of Growth Centers, in N. Hanson, editor, *Growth Centers in Regional Economic Development*, New York: Free Press, 1972.

Boots, B. N. and A. Getis, *Point Pattern Analysis* in G. I. Thrall, editor, *Scientific Geography Series*, Vol. 8, Newbury Park, Calif.: Sage Publications, 1988.

Boyce, R. R., The Edge of the Metropolis: The Wave Theory Analog Approach, *British Columbia Geographical Series*, 1966, 7, 31–40.

Brown, L. A., *Diffusion Processes and Location: A Conceptual Framework and Bibliography*, Bibliography Series Number Four, Columbus: Ohio State University, Regional Science Institute, 1968.

———, The Market and Infrastructure Context of Adoption: A Spatial Perspective on the Diffusion of Innovation, *Economic Geography*, 1975, 51, 185–215.

———, *Innovation Diffusion: A New Perspective*, New York, N.Y.: Methuen, 1981.

Burgess, E. W., Growth of the City, in R. E. Park, E. W. Burgess and R. D. McKenzie, editors, *The City*, Chicago: University of Chicago Press, 1925.

Casetti, E., Why Do Diffusion Processes Conform to Logistic Trends? *Geographical Analysis*, 1969, 1, 100–105.

Chatterjee, S. and B. Price, *Regression Analysis by Example*, New York: John Wiley, second edition 1991.

Clark, C., Urban Population Densities, *Journal of the Royal Statistical Society*, Series A, 1951, 114, 490–96.

Colenutt, R. J., Linear Diffusion in an Urban Setting: An Example, *Geographical Analysis*, 1969, 1, 106–14.

Griliches, Z., Hybrid Corn: An Exploration in the Economics of Technological Change, *Econometrica*, 1957, 25, 501–22.

Hägerstrand, T., *On The Propagation Of Innovation Waves*, Lund, Sweden: Lund Studies In Geography B, No. 44, 1952.

———, Aspects of the Spatial Structure of Social Communication and the Diffusion of Information, *Papers, Regional Science Association*, 1965, 16, 27–42.

———, *Innovation Diffusion as a Special Process*, Chicago: University of Chicago Press, 1967a (1953) (Alan Pred, trans.).

———, On Monte Carlo Simulation of Diffusion, in W. Garrison and D. Marble, editors, *Quantitative Geography*, Evanston, Ill.: Northwestern University Studies in Geography, No. 13, 1967b.

Harris, C. and E. L. Ullman, The Nature of Cities, *Annals of the American Association of Political and Social Science*, 1945, 242, 7–17.

Hoyt, H., *The Structure and Growth of Residential Neighborhoods in American Cities*, Washington, D.C.: Federal Housing Administration, 1939.

Lowry, I. S., *A Model of Metropolis*, Santa Monica, Calif.: Rand Corp., 1964.

Morrill, R. L., The Negro Ghetto: Problems and Alternatives, *Geographical Review*, 1965, 55, 339–61.

———, Waves of Spatial Diffusion, *Journal of Regional Science*, 1968, 3, 1–17.

———, The Shape of Diffusion in Space and Time, *Economic Geography*, 1970, 46, 259–68.

———, G. L. Gaile and G. I. Thrall, *Spatial Diffusion*, in G. I. Thrall, editor, *Scientific Geography Series*, Vol. 10, Newbury Park, Calif.: Sage Publications, 1988.

Muth, R., *Cities and Housing*, Chicago: University of Chicago Press, 1969.

Norcliffe, G. B., *Inferential Statistics for Geographers*, New York: John Wiley, 1977.

O'Neill, W. D., Estimation of a Logistic Growth and Diffusion Model Describing Neighborhood Change, *Geographical Analysis*, 1981, 13, 391–97.

Sargent, C. S., 1972. Toward a Dynamic Model of Urban Morphology, *Economic Geography*, 1972, 48, 357–74.

Thrall, G. I., The Proportion of Household Income Devoted to Mortgage Payments: A Model with Supporting Evidence, *Annals of the Association of American Geographers*, 1983, 73, 220–30.

———. *Land Use and Urban Form*, London: Routledge/Methuen, 1987.

———, Statistical and Theoretical Issues in Verifying the Population Density Function, *Urban Geography*, 1988, 9, 518–37.

———, The Production Theory of Land Rent, *Environment and Planning A*, 1991, 23, 955–67.

———, Using the JOIN Function to Compare Census Tracts, *Geo Info Systems*, 1992, 2:5, 78–81.

——— and S. Elshaw Thrall, Reducing Investor Risk: A GIS Design for Real Estate Analysis, *Geo Info Systems*, 1991, 1:5, 78–81.

Thrall, G. I. and A. P. Marks, Functional Requirements of a Geographic Information System for Performing Real Estate Research and Analysis, *Journal of Real Estate Literature*, 1993, 1, 49–61.