Survival Analysis in Land Change Science: Integrating with GIScience to Address Temporal Complexities

Li An* and Daniel G. Brown†

*Department of Geography, San Diego State University
†School of Natural Resources and Environment, University of Michigan

Although land changes are characterized by dimensionality in both space and time, and a multitude of methods and techniques have been developed to model them, the temporal dimension has seldom been adequately addressed by commonly used methods. In the context of temporal complexities represented in different space–time data models, this study aims to establish a framework for applying survival analysis theory and techniques to geographical land change modeling. Our efforts focus on (1) introducing basic concepts in survival analysis and their connections to space–time data commonly used in land change analysis, (2) using survival metrics to describe temporal patterns that are not easily detected by other methods, and (3) applying survival analysis methods to disclose effects of varying temporal patterns and uncertainties. Our findings suggest that survival analysis, coupled with geographic information systems (GIS) and remote sensing data, can effectively disclose relationships in land changes, and in many instances excel in shedding light on the temporal patterns of land changes. *Key Words: geographic information systems, land change modeling, remote sensing, survival analysis, temporal complexity.*

Land change modeling has attracted increasing attention from geographers in the last three decades, as a rapidly growing human population has dramatically altered land cover around the globe, resulting in many environmental problems. There exist three major questions in this arena: why (which factors, in what ways), where (in which locations), and when (at what rate or in what time) do land changes occur (Lambin 1997; Brown, Pijanowski, and Duh 2000)? By land changes we mean changes in land use, land cover, or both, where land use is defined to be the ways humans use the land for various activities and land cover is the physical or biotic attributes of the land (e.g., Lambin, Arounsevell, and Geist 2000). In our case study, we use data on land-use changes, but the issues (e.g., modeling approaches) we address in this article can apply equally to either of them. In this article we focus on addressing the third major question in connection with the first two; that is, at what time and at what rate do land changes occur? Our major emphasis, however, is to introduce survival analysis to the land change modeling literature and show how survival analysis can be coupled with various space–time data models to help in answering this question.

This article begins with a brief review of the space–time data models, focusing on how temporal spatial information is saved and retrieved. Then the applications, strengths, and weaknesses of traditional statistical models in geographical land change modeling are brought up, emphasizing how well they utilize the information contained in such data models. Following this review, we introduce survival analysis, an alternative type of model often used in biomedical studies. Our goals are to demonstrate that survival analysis has substantial potential to address time-related complexities, to demonstrate the power of survival analysis when integrating with geographic information systems (GIS) and remote sensing data, and to explore its potential to better understand the temporal components in both the response and explanatory variables. To better achieve these goals, we illustrate survival analysis using a case study from southeastern Michigan, a region that is home to 5.5 million people and experiencing significant deconcentration of the human population (Brown et al. 2005), resulting in a range of settlement densities from the urban settings of Detroit, Pontiac, Flint, and Ann Arbor, through suburban and exurban environments, to low-density agricultural land. We conclude this article by pointing out the strengths, caveats, and future directions of survival analysis in land change science.

Background

Contemporary land change models address spatial heterogeneity well, and most frequently use statistical tools that include Markov chain models,

Annals of the Association of American Geographers, 98(2) 2008, pp. 1–22
Initial submission, March 2006; revised submission, March 2007; final acceptance, July 2007
Published by Taylor & Francis, LLC.

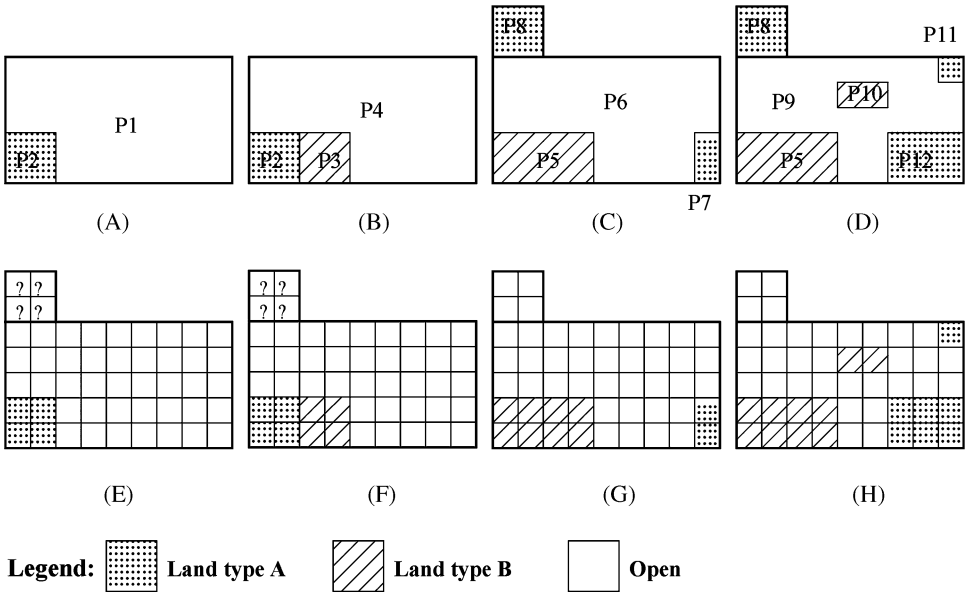


Figure 1. Illustration of land change over time. (A) through (D) represent spatial distributions of land types O, A, and B at four times, T1 through T4. (E) through (H) represent the data represented by the space–time snapshot model.

logistic-function models, and multivariate linear regression models. Following a brief overview of the space–time data models in GIScience, we point out the limitations of these three types of models in conducting space–time analysis.

Space–Time Data Models

Land changes take time to occur. Development of a subdivision may take a couple of years, and several days or weeks may suffice to convert a forest patch (e.g., three acres) to farmland. An issue of time measurement resolution may arise: If the measurements largely match the scale at which changes operate (e.g., the conversion of the forest patch is measured in days or weeks, or development of subdivisions in years), we may call it a precise measurement. On the other hand, if the timing of land changes is measured at a much coarser level, we may call it a coarse or imprecise measurement (e.g., the conversion of the forest patch is measured at annual intervals, or the development of subdivisions at intervals of twenty or more years). In this context, we give a land change example that incorporates many common land change events, such as merging and splitting of spatial units and type transition. The timing of such land changes will be considered first as precise measurements (e.g., obtained from governmental documents of land parcel transactions or developments), then as coarse measurements (e.g., derived from remotely sensed images at chosen times).

Assume there is a piece of land that has the potential to be used for three different purposes, labeled O (open), A, and B. For the simplicity of spatial illustration, we use squares or rectangles as boundaries of the land parcels in evolution (Figure 1A–1D). Also for the simplicity of illustration, we assume that at or between four time points (T1–T4), land changes could be measured, corresponding to a precise or coarse measurement of timing, respectively. At the beginning (T1), the land has two parcels: Parcel 1 for O and 2 for A. At or prior to T2, Parcel 1 is split into Parcel 3 (B) and 4 (O). At or prior to T3, we are able to collect new information about Parcel 8, for which no data are available at earlier times for various reasons such as clouds in a satellite image or a missing land parcel document. Also, Parcel 4 is split into 6 (O) and 7 (A), and Parcels 2 (A) and 3 (B) are merged into 5 (B). At or prior to T4, Parcel 6 is split into Parcels 9 (O) and 10 (B), and part of it is absorbed by Parcel 7, forming Parcel 12 (A). We use this example as we introduce different space–time data models next. Our focus is to show how spatiotemporally changing information is contained in, and can be retrieved from, such data models for use in land change analysis.

The most popular space–time model is the snapshot model, where the space is usually gridded into a collection of equal-sized cells, and such grid-based maps are used repeatedly to represent the changing status of the space over time (Armstrong 1988). Although criticized for reasons such as lack of efficiency (e.g., repeated storage of the same grids over time), the snapshot model

Composites	T1	T2	T3	T4
P5	A, O	A, B	B	B
P8	N/A	N/A	A	A
P9	O	O	O	O
P10	O	O	O	B
P11	O	O	O	A
P12	O	O	A,O	A

(A)

ST Atoms	Time & Land type (Precise)	Time & Land Type (Coarse)	ST Objects
P1	[T1, T2) O	[T1, T2) O	O
P2	[T1, T3) A	[T1, T3) A	A
P3	[T2, T3) B	(T1, T3) B*	B
P4	[T2, T3) O	(T1, T3) O	O
P5	[T3,) B	(T2,) B	B
P6	[T3, T4) O	(T2, T4) O	O
P7	[T3, T4) A	(T2, T4) A	A
P8	[T3,) O	(T2,) A	A
P9	[T4,) O	(T3,) O	O
P10	[T4,) B	(T3,) B	B
P11	[T4,) A	(T3,) A	A
P12	[T3,) A	[T3,) A	A

(B)

Figure 2. The data in Figure 1 represented by (A) the space–time composites model and (B) the spatiotemporal object model.

is still the predominant space–time data model, which may arise from its straightforwardness and ease of comprehension and interpretation. Using this model, the preceding land change example can be represented as in Figure 1E through 1H, and we can easily retrieve the land change history of any cell or parcel. If we are interested in, for instance, the land change history of Parcel 10, we simply go to the four maps, locate the associated parcels at each map, and record their land type, which results in the trajectory O → O → O → B. An inherent issue with this data model is that the number of snapshots (or time points) will be relatively small; otherwise the workload (e.g., data storage and retrieval) will become increasingly high. Therefore this data model is more appropriate for space–time data measured at a coarse time resolution, or if the number of land change events is relatively small.

An improvement was made by Peuquet and Duan (1995), who developed the event-based spatiotemporal data model (ESTDM), where only one initial grid map (here at T1) needs to be saved, and cells (identified by their x- and y-coordinates) with land-type changes are saved in different lists at each later time. With this modification, ESTDM can effectively handle space–time data measured at either precise or coarse temporal resolutions. To identify the land change history of Parcel 10, the model simply visits the initial map and finds its type as O. Then it moves to T2 and T3, checking the lists of cells where the land type has changed. As the cells for Parcel B are not in such lists, the model indicates that no changes have occurred at these steps, thus keeping its status unchanged as O for both T1 and T2. For the last step (T4), the cells associated with Parcel B should be in the list for land type B, thus its status will be decided as B. Still a trajectory of O → O → O → B will be found.

The space–time model by Langran and Chrisman (1988), the space–time composites model, breaks the space over time into increasingly small fragments, identifying a maximum number of units with *spatial* homogeneity at each time. The land changes in our example can be represented using this model (Figure 2A), where we can easily find the land change history for Parcel 10 is that for polygon P10; that is, O → O → O → B. This model can handle space–time data measured at both precise and coarse temporal resolutions; if at a precise resolution, there should be many time measurements (e.g., T1, T2, . . . , T200), expanding the table with more columns. This model, however, only records the historical types of the parcels (i.e., spatial composites) at the latest time, ignoring the transitional histories. At a coarse resolution, the interpretation should be different. For instance, the trajectory O → O → O → B for P10 suggests that the parcel is developed to land type B any time after T3, rather than at T4.

The spatiotemporal object model by Worboys (1992), on the other hand, can avoid this problem (Figure 2B). This model identifies each spatiotemporally homogeneous unit as a space–time atom (ST atom), and many such atoms of the same land type

Sem ID	Land type	Other
1	O	…
2	A	…
3	B	…

(A)

Time ID	Land type	Other
1	T1	…
2	T2	…
3	T3	…
4	T4	…

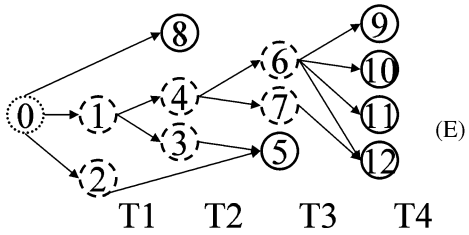
(B)

Space ID	Area	Perimeter	Other
P5	…	…	…
P8	…	…	…
P9	…	…	…
P10	…	…	…
P11	…	…	…
P12	…	…	…

(C)

Sem ID	Land type	Space ID list
O	T1	P1
A	T1	P2
O	T2	P4
A	T2	P2
B	T2	P3
O	T3	P6
A	T3	P7, P8
B	T3	P5
O	T4	P9
A	T4	P8, P11, P12
B	T4	P5, P10

(D)



(E)

Figure 3. The data in Figure 1 represented by the space–time three-domain model: (A) the semantics table, (B) the time table, (C) the space table, (D) the domain link table, and (E) the spatial graph.

constitute a spatiotemporal object. To retrieve the land change history of a parcel or cell, we have to overlay many ST atoms that occupy or at least share the location of the parcel or cell of interest. To retrieve the land history for Parcel 10, the model needs to identify the ST atoms based on their locations: first P10, which we know is type B at T4; then P6, which we know is O at T3. Similarly from P4 and P1, we can find their land type as O. Therefore its land change history is O → O → O → B. When measurements are taken at a precise resolution, P10 is represented as [T4,) B, implying P10 of type B occurs from T4 inclusively, and persists until the end (thus a space before the right parenthesis). The parentheses refer to the point not included, and square brackets the point included. When measurements are taken at a coarse resolution, P10 is represented as (T3,) B, suggesting that P10 of type B occurs any time after T3 (but before T4), and persists until the end.

Finally, the three-domain model (Yuan, 1999) uses a semantics table (Figure 3A) to stand for different land types (O, A, and B in our example), a time table (Figure 3B) for different times (T1–T4 in our example), a space table (Figure 3C) for the polygons or spatial objects at the last time (T4), a domain link table (Figure 3D) that connects the preceding three tables, and finally, a spatial graph that displays the land change history (Figure 3E). To retrieve the land change history of Parcel 10, the model first checks the space table, and finds the location (and other characteristics) of Parcel 10. Visiting the domain link table, we can find that its type is B. Then checking the spatial graph, we can find the parental parcels for Parcel 10 at T3, T2, and T1 are Parcels 6, 4, and 1. A revisit to the domain link table shows that O is the type for these parcels. Thus O → O → O → B is derived as the land change history for Parcel B. When measurements are taken at a precise resolution, the time table and domain link table become larger. The interpretations of the land change trajectories at both time resolutions, such as O → O → O → B for Parcel 10, are the same as those in the spatiotemporal object model.

Complexities in Space–Time Data

The preceding data models have different strengths and weaknesses (see Yuan 1999 for an excellent review), but all connect to views of space and time. The traditional Newtonian absolute view of space and time tends to regard space (and time sometimes) as a container or framework, in which various contents are included and processes are ongoing. The relative view of space and time tends to regard both space and time as qualities of a certain phenomenon, which are extended over both

space and time (Bian 2003). The space–time data in land change science, in the context of either view, usually have the following unique complexities (the term *complexity* in this article primarily refers to heterogeneity and uncertainty).

Spatial Complexities. The space has to be partitioned into discrete units (pixels or objects), despite the fact that land surface is continuous, that correspond to two fundamental conceptual models for representing space and spatial phenomena. The first is the raster or field data model that imposes a simplified data structure, and the space is represented as a collection of grid cells that simplifies analyses, which is the choice of the snapshot model and the model by Peuquet and Duan (1995), but it fails to provide "natural" units matching real-world entities such as lakes and hills. On the other hand, the object model allows objects to take their natural shape, represented as polygons, lines, and points, but requires an ontological framework for selecting the right objects. The space–time composites model, spatiotemporal object model, and the three-domain model fit better to this type of spatial representation when the time component is removed. Recent years have seen some hybrid data models that parallel the development of object-orientation programming in computer science since the 1990s.[1] Such hybrid models, including the object-oriented grid model (i.e., an aggregate of many cells used to represent a meaningful geographic feature) and the object-oriented field model (i.e., fields or patches treated as objects in computer science; Bian 2003), are still based on the most fundamental representations of space.

Whatever spatial model is chosen for a certain research goal, there has to be some simplification, and some trade-off has to be made between the ease of data collection and analysis and reduction of information loss or distortion in light of the question being investigated. In addition, decisions about the appropriate scale and resolution (cell size in the raster model), degree of fuzziness (indistinct boundary between, e.g., a forest polygon and a grass polygon in the field model), and many types of uncertainties (e.g., lack of adequate information; Peuquet 2001) make our choice of spatial model challenging. This choice may be even more challenging if temporal complexities of spatial data are also of concern, which relates to our next issue about temporal complexities.

Temporal Complexities. Land changes can occur at any time, and such changes may be measured either at precise or coarse temporal resolutions. The data in the snapshot model may have a relatively coarse time resolution due to logistical (e.g., data availability or storage) considerations despite the fact that in theory we can still choose the times (T1, T2, etc.) at precise resolutions. In this case, what we know is that the events happen during some interval determined by the data-collection times, rather than the specific times at which they occur. This may result in ties in land change occurrences—that is, multiple land changes recorded during the same time interval (e.g., P10 and P11 in Figure 1D) have the same (i.e., tied) development time regardless of their true development times—and little research has been devoted to this type of issue. The other three space–time models, although allowing for relatively precise time resolutions, still suffer from uncertainties that may arise when land changes have occurred before we collect our data (e.g., Parcel 2 in Figure 1A) or no events have occurred until the last time of our data collection (e.g., Parcel 9 in Figure 1D, which may be further split and developed into other types).

Implicit Dynamic Information. Many other characteristics, regardless of whether they are properties of the land parcels themselves (e.g., soil fertility) or of the surrounding environment (e.g., neighborhood population density), may change over time, and may directly or indirectly affect directions or rates of land change. To better understand land changes, such dynamic characteristics should be accommodated in our space–time data, whether they are saved, retrieved, and calculated as attributes of the cells or polygons in our data, or separately as different data layers (e.g., a soil layer, a population density layer). When investigating space–time data as a whole, a great deal of information related to land change dynamics may be revealed, but usually cannot be retrieved through regular GIS operations such as overlay or map queries. For instance, the numbers of parcels of type A and type B change over time, which may suggest different risks (for detail, see Appendix A) of being developed into land type A and B, or popularities of land type A and B, over both time and space. We may need metrics to quantify such risks aside from those we usually use, such as probability. If such risk differentiation is not random and can be quantified, such risks may connect to other factors such as the dynamic characteristics mentioned earlier. We may further investigate whether the effects of such dynamic characteristics on the risks of land development change over space or time.

Land-Unit Complexities. No matter what spatial units and spatial model we choose for analysis in accordance with our research objectives, land units seldom remain unchanged over time. Boundaries (e.g., from ownership boundaries to territories of different countries) are subject to changes over time. Furthermore, land parcels may be developed into multiple types (land type A and B in Figure 1), which raises the issue of competing risks of land development: Any undeveloped parcel has a risk of being developed into different types.

Traditional Models

Land change models, categorized in a variety of ways (e.g., Baker 1989; Lambin, Arounsevell, and Geist 2000; Agarwal et al. 2002), have been developed to represent human behavior, spatial patterns, and temporal dynamics of land changes. With the aid of remote sensing and GIS in providing, processing, and analyzing land change data, statistical models (we thus exclude other types of models such as cellular automata and agent-based models in our discussion) have been employed extensively when identifying localized deterministic relationships is impossible or unaffordable (in terms of time and resources) and when identifying exogenous causes is necessary. Such models, including Markov chain models, logistic-function models, and multivariate linear regression models, primarily focus on the spatial component in predicting or explaining land changes; little research has been done to explore complexities related to time (e.g., is the time interval appropriate?) and implicit dynamic information (e.g., does changing population density in the neighborhood affect the development of a certain land type differently over time?). Choosing and defending appropriate land units is a task for the researcher that we address later. Next we briefly review these three types of models.

Markov chain models have found numerous applications in land change studies (e.g., Baltzer 2000; Brown, Pijanowski, and Duh 2000), particularly when exogenous sociodemographic and biophysical factors are not easily accounted for due to problems such as data constraints and difficulties in modeling human decisions (Luijten 2003). Although valuable for their mathematical and operational simplicity, Markov models focus on predicting (rather than explaining) changes in the types of predetermined land units at the designated times (i.e., it does not address temporal uncertainty, such as whether the time intervals or span are chosen appropriately), assume stationarity in the transition matrix, and usually only deal with first-order processes (Baker 1989; Lambin 1997). Although remedies have been proposed to handle these limitations, such as regressing the transition probabilities against exogenous variables and using variable transition matrices over time (e.g., Baker 1989; Brown, Pijanowski, and Duh 2000), Markov modeling has limited application in land change analysis as it is only appropriate for short-term projection (Lambin 1997).

Logistic-function models and their variants (e.g., multinomial-logit models) are widely used in land change modeling studies, where the response variable (logit transformation of a variable indicating presence or absence of a certain land-use type) is expressed as the linear combination of an intercept term and explanatory variables. The coefficients thus obtained can be used to calculate the weights in algorithms to generate maps showing the probabilities of a category of land-use or land-cover change over the study area (Mertens and Lambin 2000). If more than two land-use or transition types (or other categories) are involved, multinomial-logit models are often used (e.g., Mertens and Lambin 2000; Müller and Zeller 2002). Even if land units and time intervals are defensibly chosen, logistic regression still suffers from the following limitations when dealing with the preceding space–time data (Figures 1 and 2): (1) It is difficult to choose the values of the dependent and independent variable if both land type and some variables have changing values over time; if the researcher chooses to use the latest values for both the dependent and independent variables or aggregate such time-changing values for the associated independent variables, the model might suffer from (2) a loss of information and failure to capture potential dynamic mechanisms; if the researcher chooses to establish a model for each time step, the model could suffer from (3) a reduction in degrees of freedom, especially at earlier times.

One variant of logistic-function models is called the trajectory model, where land parcels or pixels are classified into various trajectories depending on both the type and timing of development, such as the O → O → O → B trajectory for Parcel 10 (Figure 2A). The nominal trajectories are regressed against the explanatory variables (e.g., Mertens and Lambin 2000). Although this method can capture the dynamics in the dependent variable (i.e., land change types), it fails to capture the effects of those independent variables that take time-changing values. It also suffers from other problems such as reduction in degrees of freedom. We discuss trajectory models further later in this article.

Multivariate linear regression models almost invariably deal with cross-sectional data (the sections in this sense are equivalent to the land units, which are chosen and assumed appropriate) combined with GIS (e.g., Mertens et al. 2000). Strictly speaking, multivariate linear regression includes logistic-function regression models. Here we refer to multivariate linear models that do not use any transformations (e.g., logit or probit) for the response variables. Although some examples involve dividing the entire period into a few periods and analyzing the data within these periods separately (which may result in reduction in degrees of freedom; e.g., Mertens et al. 2000), very few applications handle temporal variations explicitly. Panel data analysis techniques, integrating cross-sectional and time-series data, provide algorithms for handling intersection and interperiod heterogeneity in model coefficients. For reasons of computational complexity (Hsiao 1986, 130) and limited software availability, however, researchers still tend to estimate models in which only the intercept varies (e.g., over time or individuals). An exception is the work by Seto and Kaufmann (2003), which estimated a few models of land-use transitions in the Pearl River Delta (China) and allowed for temporally variable coefficients for a set of sociodemographic factors.

These models are short of diagnostic metrics and methods to unveil past trends in key variables of interest and test hypotheses about them, and weak in representing, disclosing, or predicting temporal variations of some key variables, especially when spatial heterogeneity and uncertainty are also of concern (e.g., Serneels and Lambin 2001). The temporal aspects of land change cannot be ignored while studying variations in other dimensions such as human behavior and space, because they are often intertwined. It is very difficult (if not impossible) for current land change statistical models to answer questions such as "by a certain time, what proportion of land-use type i would be converted to type j $(i, j = 1, 2, \ldots, n, i \neq j)$" or "was the transition rate from land use type A to B the same as (or faster than) that from A to C? If not, what factors led to such differences?"

Methods

Given the described issues and challenges in land change analysis, we propose a framework for choosing land change units that addresses the complexities related to space, time, and land units. Further, we argue that the combination of this framework with survival analysis has great potential to address the complexities involved in analyzing land changes.

Framework to Choose Land Change Units

As demonstrated in the earlier example (Figure 1) and the four major space–time data models, land changes are subject to many complexities over time and space. In the context of the question under investigation and the space–time data (including the associated data model) we have, we may choose as our units of analysis the parcels at the most recent time (T4 in Figure 2A), which correspond to the most recent land composites in the space–time composite model, the ST atoms with open-ended times (e.g., (T2,) O for P8 in Figure 2B) in the spatiotemporal object model, or the spatial objects at the end of the spatial graph in the three-domain model (Figure 3E). For the snapshot model, we may choose collections of homogeneous cells at the latest time (e.g., P10 in Figure 1D), which correspond to aforementioned parcels. The advantages of this "most recent parcel or collection of cells" framework include the following aspects. First, land tends to become more fragmented and heterogeneous over time (although merging can dominate some landscapes; e.g., where agricultural firms are experiencing consolidation), thus choosing as land units the parcels at the latest time helps to retain the most updated land-use/land-cover information and to track historical events at earlier times. Second, the most recent parcels are the ones that affect future local land-related decisions and developments. Third, and last, data of such type may be most available. On the other hand, depending on research questions at hand, choosing individual pixels may work in some instances (e.g., for regression analysis).

Following the choice of land units, we can derive land change trajectories of these units—if such units are to be used in regression-related analyses, we should take a sample of units in consideration of spatial autocorrelation. A common way to reduce the effects of spatial autocorrelation is to take samples that are far apart from each other in space (e.g., the distances between the nearest pairs are not less than the range in the corresponding semivariogram); if individual pixels are the land units, the selected pixels should also belong to different collections or clusters of homogeneous pixels. Recent work has developed theoretical and empirical explorations in detecting and incorporating "any lingering spatial correlation," or "frailty," in the spatial residuals of survival models (interested readers see http://geography.sdsu.edu/People/

Pages/an/SA_app.pdf; Banerjee, Wall, and Carlin 2003 Banerjee and Dey 2005). We already showed how to derive the land trajectories earlier. One concern might be the lack of consistency in the shape, boundary, and size of land units, which can be well observed from the spatial graph (Figure 3E). When analyzing such trajectories, we have to focus on some aspects and overlook some less important aspects. A parcel may be treated as a point (i.e., no shape, no area, no boundary) when the shape, boundary, or size of the parcel is trivial for the question under investigation and can be ignored.

Survival Analysis

Survival analysis is a collection of statistical methods used to characterize the occurrence and timing of events that represent any qualitative changes in state during some time-series processes. It originated from the study of mortality or failure in many disciplines, where attention is paid to the survival times of individual patients or experimental animals (thus the name survival analysis; Allison 1995; Klein and Moeschberger 1997), the failure time of some equipment, or the duration of some status until some event (e.g., birth, divorce, arrest) happens (e.g., Murthy and Haywood 1979; Liestol, Andersen, and Andersen 1994; Du et al. 2002). Although going by different names (e.g., event history analysis in sociology, reliability or failure time analysis in engineering, and duration or transition analysis in economics), survival analysis attempts to answer questions about the same characteristic; that is, occurrence and timing of some events during some processes of individual entities.

Despite its popularity in other disciplines, survival analysis has been applied in only a few pioneering articles related to geographical land change in recent years. Vance and Geoghegan (2002) constructed a complementary log-log model to estimate the effects of a number of sociodemographic and biophysical factors on forest clearance processes in southern Mexico, answering questions such as "What factors significantly affected the household (*ejido*) forest clearance decisions?" Irwin and Bockstael (2002) examined negative externalities (repelling effects) that arise from existing residential subdivisions using the Cox proportional hazard model, where time-varying covariates were used to capture the ever-changing land-use status around the parcel under consideration. Coomes, Grimard, and Burt (2000) studied the survival rates of land plots among land-rich and land-poor households in an Amazonian community in Peru, and how fallow-period length could be explained by some exogenous variables using the Cox proportional hazard model, which is a popular form of survival analysis. In a broader sense, survival analysis has been used to study job mobility and residential relocation in urban areas of the United States (Clark and Withers 1999) and the spacing of settlements in Nebraska (Odland and Ellis 1992).

In these pioneering studies, survival analysis has undoubtedly shed important light on their topics, but they also raise a warning regarding its application in land change analysis: lack of independence between events over time. We address this issue in our discussion. As case studies, these examples have not provided systematic and theoretical explorations that permit conclusions about how survival analysis can help land change studies. As a result, this method is rarely mentioned in articles reviewing geographical land change models (for instance, Baker 1989; Lambin 1997; Agarwal et al. 2002), with the exception of Irwin and Geoghegan (2001) and Plantinga and Irwin (2006).

Before we introduce the strengths of survival analysis, we first introduce two critical concepts in survival analysis, the survival function, $S(t)$, and hazard function, $h(t)$, which are defined to be

$$S(t) = \Pr(T > t) = 1 - F(t) = \exp\left\{ -\int_0^t h(x)dx \right\} \tag{1}$$

and

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr\{t \le T \le t + \Delta t | T \ge t\}}{\Delta t} = -\frac{d}{dt} \log S(t) \tag{2}$$

They are the probability that an individual can survive beyond time t (i.e., the event does not occur until t) and the instantaneous ($\Delta t \to 0$) risk that an event will occur at time t given that the individual survives to time t, respectively. The hazard can be understood as an intrinsic property of any individual, and is conceptually different from probability (for details see http://geography.sdsu.edu/People/Pages/an/SA_app.pdf). This definition, however, does not preclude the possibility that we can still calculate overall hazards based on the aggregate data of all individuals over the entire time frame, where the time frame is usually discretized into several periods (if originally not discrete in the data), and the number of events in

each period is used in these calculations (for details, see Machin, Cheung, and Parmar 2006, 23–49). In this sense, the hazards can be understood as the average risks that a land parcel would be subject to over time. The other term in survival analysis, survival probability, calculated in a frequentist manner, offers a general indicator of what proportions of land parcels under investigation may remain undeveloped over time. Worthy of mention is that hazards may go up and down, whereas survival probabilities are always nonincreasing over time. In the section "Case Study," we show an example of such curves.

In land change science, analogous to a person's death or equipment's failure, a land parcel may be developed (or developed into one of different types). Therefore, the event for a land parcel is its *development* (or development into a certain type if multiple development types are under investigation) or land change, and the survival time is the time when the parcel remains undeveloped or the change has not exceeded the threshold of being regarded as a "change" or "development." Developments can reoccur: A parcel might be first developed into residential then into commercial space. For simplicity of analysis, we restrict our analysis to single developments (once a parcel gets developed, it remains in that land-use type without further change) despite the capacity of survival analysis to handle reoccurrences. Hereafter we use the word *individual(s)* to refer to the potential study participant(s), who may be patients in medical research, machines in an engineering test, or land parcels in a land change study.

Data Censoring in Land Changes

In traditional applications of survival analysis, the events under study are often known to have occurred in a time interval (t_1, t_2) within the time range, in theory, from $-\infty$ to $+\infty$. This gives the three basic types of intervals: the event may have occurred earlier than t_2 (i.e., within the interval $(-\infty, t_2)$ where $t_1 = -\infty$), between t_1 and t_2 or within (t_1, t_2), and later than t_1 (i.e., within the interval $(t_1, +\infty)$ where $t_2 = +\infty$). In survival analysis, the survival time is called left censored at time t_2, interval censored between t_1 and t_2, and right censored at time t_1, respectively. Figure 4 illustrates these censoring types, where Panel A is an example regarding deaths of patients in medical science, and Panel B shows development histories of three land parcels (Figure 1) measured at a *coarse* time resolution. In Panel A, all we know about Person 1 is that he dies before T3, so his survival time is left censored at T3; that is, (., T3). Similarly, Parcel P8 in Panel B is developed into type A before T3, and its survival time is also left censored at T3; that is, (., T3). Person 2 (Panel A) and Parcel P9 (Panel B) have survived the whole time frame; that is, their survival times are right-centered at T4, so (T4, .). Person 3 (Panel A) and Parcel P10 (Panel B) have their survival times interval-centered between T3 and T4, but they die of another reason (compared with Person 1) or get developed into another type (type B), which relates to the issue of competitive risks in survival analysis. Such types of censoring connect to the challenge related to time heterogeneity and uncertainty in land change analysis.[2] For other space–time data models where the precision of time measurements could be fine enough, right censoring still exists (i.e., Parcel 9 remain undeveloped until T4 in Figure 4B). Another topic related to temporal uncertainty is data truncation or late entry, a situation in which an individual is not at risk for some period(s) of time (Kalbfleisch and Prentice 2002, 23–24). We do not address it due to space limitations.

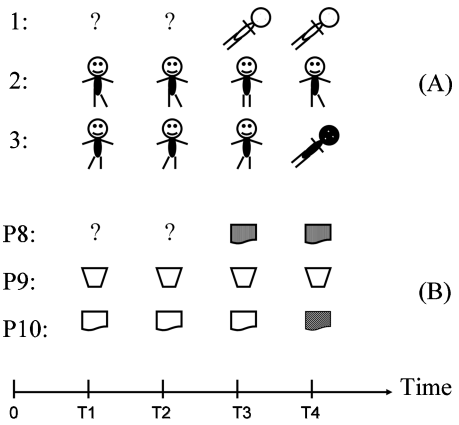


Figure 4. Illustration of the concepts of survival times, censoring types, and competitive risks in (A) traditional survival analysis and (B) land-change-related survival analysis.

Although it is self-evident that incorporation of censored data would benefit the models by retaining more, and probably more accurate, information than if these data were removed from the analysis, we provide a short statistical explanation about the importance of incorporating censored data explicitly. As shown in Equation 8 (Appendix B), survival analysis makes best use of the observed data by incorporating various types of events (i.e., uncensored, left-, right-, and interval-censored survival times) in estimating the coefficients. Ignoring these issues can result in substantially different, sometimes misleading, results. For more details, see Kalbfleisch and Prentice (2002, chap. 3).

Time-Dependent Variables

Some explanatory variables may take varying values (e.g., age of a patient) over time, which gives rise to the issue of time-dependent variables. Data in the land change analysis arena often exhibit these characteristics, which is the challenge identified as implicit dynamic information earlier. For instance, the neighborhood population density associated with a land parcel might change over time. If we regress land changes against a set of explanatory variables that take changing values, it is self-evident that such variables would affect the model estimates differently compared with a situation where these variables are either ignored or are forced to take the values of one time or the values averaged over time. Such issues occur frequently in land change science, but inadequate attention has been paid to them.

The Cox Hazard Model

The data as represented in the space–time composites, spatiotemporal object, and the three-domain data models could have relatively precise time measurements, thus interval censoring should not be a big concern, and the censoring types that need to be considered are right censoring and left censoring; that is, events or land developments have not occurred until the end of a time frame (Parcel 9 in Figure 1) or have occurred prior to our study time frame (Parcel 2 in Figure 1). In practice, it is common to regress the logarithm of the hazards against a linear combination of explanatory variables that are determined a priori (Klein and Moeschberger 1997, 282):

$$\text{Log } h_i(t) = \alpha(t) + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik}. \quad (3)$$

where $\alpha(t) = \log \lambda_0(t)$ (see http://geography.sdsu.edu/People/Pages/an/SA_app.pdf for $\lambda_0(t)$ and more details). Depending on whether the independent variables take changing values over time or we let any of the coefficients vary over time (i.e., insert an interaction term between time and any of the variables), we can estimate a proportional model (i.e., the hazards of any two parcels, although each may change over time, have a fixed proportion) or nonproportional model, respectively. As $h_i(t)$ is not directly observed, many software packages use survival time with censoring information as a substitute. The methods developed by Cox (1972; i.e., partial maximum likelihood estimation) are used to estimate each β without the need to know the value of $\alpha(t)$ or $h_i(t)$ at a certain time. The Cox model in Equation (3) can handle competitive risks by treating all the event and development types not being considered as being right censored (Appendix A). The Cox model needs a one parcel → one record format, where a record is created for one parcel or polygon, and time-dependent variables, if any, take values prior to the corresponding land changes. Nonproportionality (i.e., the hazard ratios between individuals change over time) may result from the existence of time-dependent variables and time-varying coefficients of some variables (http://geography.sdsu.edu/People/Pages/an/SA_app.pdf). Here we are more concerned with the latter, because time-varying coefficients represent changes in the relationship between the development hazard and a specific variable.

In general, the Cox model in Equation (3) has been considered very powerful and employed even when the time measurements are coarse in the literature of survival analysis (e.g., Allison 1995, 111–13; Machin, Cheung, and Parmar 2006, 139–40). Land change researchers can use this model in most situations, even for data with a coarse time resolution such as those in the snapshot model. Different algorithms are developed to handle (1) ties that arise from imprecise measurements of survival time with underlying ordering (most ties in land changes may belong to this type), and (2) ties that are discrete by nature; that is, two events actually occur at the same time (see Allison 1995, 127–37). The SAS phreg procedure can handle ties in survival time measurement (e.g., the EXACT and DISCRETE methods in this procedure are designed to handle these two situations, respectively). One exception is a situation in which interval censoring exists and cannot be reasonably ignored, our topic in the next section.

The Accelerated Failure Time Model

Compared to time measurements with precise resolution (Figure 5A), time measurements with coarse resolutions (Figure 5B-D) result in all three types of censoring. In theory, the Cox model should have the ability to handle interval-censored data, but this option is so far not available in SAS (SAS online documentation). Thus we introduce the accelerated failure time (AFT) models, focusing on the Weibull and exponential models (particularly the latter and its variant, the piecewise exponential model). For other alternative models in the AFT family such as the log-logistic and log-normal models, interested readers may see Machin,

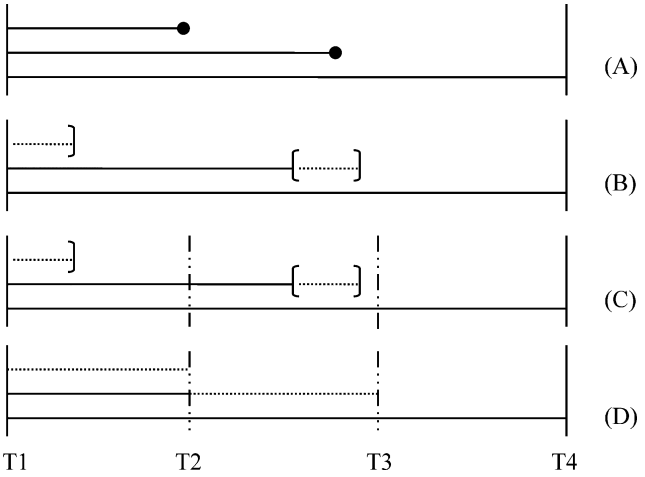


Figure 5. Illustration of precision of time measurements: (A) precise measurements, (B) coarse measurements without piecewise restructuring, (C) coarse measurements with piecewise restructuring, and (D) coarse measurements used in the logit and log-log model. The dashed horizontal bars in (B) and (C) stand for coarse time intervals within which land changes occur.

Cheung, and Parmar (2006, 108–115). The Weibull model takes the following form:

$$\text{Log } h_i(t) = \beta_0 + \alpha \log(t) + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} \quad (4)$$

By fitting the data, α can be estimated, and different magnitudes of this parameter suggest different shapes of the hazard function over time. Visit http://geography.sdsu.edu/People/Pages/an/SA_app.pdf for more details, where we describe why there is not a parcel-specific residual term as well. A special case of Equation (4) is to let $\alpha = 0$; that is, the hazard of each individual remains constant over time, suggesting that the differences in hazards over individuals are simply caused by the independent variables. This is the famous exponential model:

$$\text{Log } h_i(t) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} \quad (5)$$

Both Equations (4) and (5) can be estimated using the same lifereg procedure in SAS, which can handle various types of data censoring but cannot handle time-dependent variables directly. This problem, along with the implausibility of the assumption regarding constant hazards over time, leads to a variant of the exponential model, the piecewise exponential model, which may be especially suitable in land change analysis for reasons mentioned later.

The Piecewise Exponential Model

The piecewise exponential model belongs to the family of AFT models, but has some unique data format requirements and assumptions regarding the base hazards β_0. Rather than using the original data, the piecewise exponential model uses the restructured piecewise data, and has been widely employed (e.g., Bodian 1985; Starmer 1988; Karrison 1996; Boracchi, Biganzoli, and Marubini 2003). The piecewise data restructuring (Appendix C) simply breaks the entire time frame into several periods (three periods in Figures 5C–D), and one data record is generated for each period during which the parcel is either at risk or under development—each record of this type is a parcel period. For the parcel period when the parcel is at risk, let the survival time be right censored. Assuming the hazard is constant within each period but can vary across periods, then we build an exponential model for each parcel period (let j be the number of periods that the parcel is either at risk or under development):

$$\text{Log } h_i(t) = \beta_j + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} \quad (6)$$

where $j = 1, 2, \ldots, J$, and J is the number of periods when the parcel is either at risk or under development. Equation (6) expands on Equation (5) because (1) the survival time t in Equation (6) is less than or equal to the length of the period (e.g., the time between T1 and T2, T2 and T3, or T3 and T4; see Figure 5), whereas the survival time in Equation (5) could be greater than the length (e.g., the second horizontal bar in Figure 5B); (2) each period is allowed to have a different β_j, which may account for the changing hazard over periods; and (3) all the parcel periods are treated as independent observations. When the time precision continues to decline (e.g., from Figure 5C to 5D), the piecewise exponential model should still work, which is Model 2 in our case study (later).

Both time-dependent variables and all types of data censoring, especially when coupled with competing risks, can be taken into account in this model. There is no readily available commercial software that is designed to simultaneously handle these issues. We extend the regular piecewise exponential method (Shuster 1992; Allison 1995, 104–09; Cantor 2003, 13–15) that does not deal with competing risks and all types of censoring at the same time, and propose a three-step strategy: (1) Generate a piecewise dataset where each of the parcel-period observations incorporates all types of censoring and time-dependent

variables; (2) treat all the types that are not being considered in a particular model as right-censored observations, which allows for competing risks; and (3) model the relationships between the hazards and the explanatory variables using the lifereg procedure (for SAS code, see http://geography.sdsu.edu/People/Pages/an/SA_app.pdf).

Log-log Model

Very frequently, land change analyses have to deal with data with coarse time resolutions (data with many tied survival time measurements), such as data in many snapshot models. Let's assume there is a total of T periods, and each parcel may be developed or undeveloped in period t ($t = 1, 2, \ldots, T$); a probability P_{it} is defined to be the probability that an event (development) occurs to individual i at period t. In addition to the models already introduced, the following logit or complementary log-log model models could be used. As a first step, the data need to be restructured to be the piecewise format, where a new binary variable (e.g., dev_Status) should be defined, taking zero when it is at risk and one for the period during which it is developed (Appendix C). As we did in the piecewise exponential model, all parcel periods are treated as independent observations. As many researchers are familiar with the logit model, we introduce the complementary log-log model only:

$$\text{Log}\,[-\log(1 - P_{it})] = \beta_t + \beta_1 X_{it1} + \beta_2 X_{it2} + \ldots + \beta_k X_{itk} \quad (7)$$

This model can handle time-dependent variables; that is, when restructuring the data to the piecewise format, let the variable taking the corresponding value at time t, which is denoted by the subscript t. This model, however, is not good at dealing with competitive risks because the construction of the binary variable during data restructuring has to take either zero (undeveloped) or one (developed), and the parcels that are developed into other types not being considered may have to be coded as zero (undeveloped) or dropped from the model for the land change type being considered. The complementary log-log model is comparable to the Cox model in many aspects, such as interpretation of the coefficients (Allison 1995, 211–32).

Case Study

In this section we describe an application of the survival concepts and models in a land change case study. We first introduce the study site and data, then describe data interpretation and integration, and end up with data analyses that apply the models already described. Specifically, we first calculate survival probabilities and hazard curves, then construct a Cox model and a piecewise exponential model, followed by a complementary log-log model. Finally, we build a trajectory model with which researchers in land change science may be familiar, aiming to explore the degree of complementarity among these models.

The Study Site and Data

Land-use data were collected over eight exurban townships in southeastern Michigan: Flushing, Oregon, Pittsfield, Putnam, Ray, Scio, Washington, and Wood-

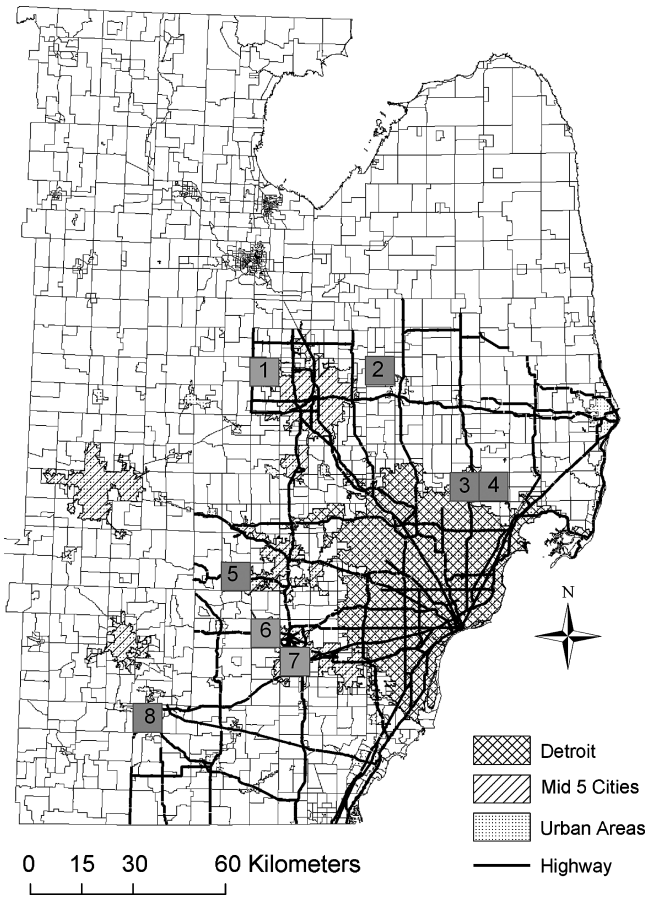


Figure 6. The location of the study site in southeastern Michigan. The townships in southeast Michigan include Flushing (1), Oregon (2), Washington (3), Ray (4), Putnam (5), Scio (6), Pittsfield (7), and Woodstock (8).

stock (Figure 6). Our land units, land parcels, were chosen by randomly sampling 4 percent of all the parcels in each township based on their most recent parcel maps (the year differs from township to township, ranging from 1997 to 2000), representing the "most recent time" approach to choosing land units in our framework; if multiple parcels were developed into a subdivision by a single developer, we merged them to create a subdivision polygon. The result was a total of 854 parcels[3] or polygons. This sample size is more than adequate to engage in statistical analyses, but is not so large that the increased computational time was counterproductive. Using five aerial photographs in each township, acquired in approximately ten-year intervals ranging from about the late 1950s to the late 1990s,[4] we visually interpreted the date of development and classified the development in each sampled parcel or polygon into one of five types based on its environmental and geographic characteristics: farm, rural lot, country subdivision, horticultural subdivision, and remnant subdivision (An et al. forthcoming; Brown et al. forthcoming), where farm is used as a land type from which all other four types (to be explained later) may be developed. Therefore any farm parcel is subject to competitive risks until it is developed. The approximately ten-year interval of land change measurements may be viewed as coarse in light of real estate developers who tend to develop their projects in an average of one and a half years (Vigmostad 2003, 103). In later modeling steps, we use all the models mentioned in the "Methods" section.

We then selected a set of geographic, biophysical, and sociodemographic explanatory variables (Table 1; An et al. forthcoming). The two response variables were the survival time (s_time, in decades) and the development type (type). The variable *type* was used to represent the type that a particular parcel or polygon belongs to at a given time: 0 for farm, 1 for rural lots, 2 for country subdivisions, 3 for horticultural subdivisions, and 4 for remnant subdivisions. The geographical variables included distances to the three levels of cities or urban areas, to the nearest lakes or streams, to county roads, and to the nearest highways (Figure 6). These three types of distances provide surrogates for a parcel's accessibility to work and urban facilities (e.g., shopping centers), to water features, and to the road systems (county roads and highways), respectively. As reported in the literature (e.g., Turner, Newton, and Dennis 1991; Sengupta and Osgood 2003), these geographic factors play important roles when people consider their residential locations. We also included biophysical variables to characterize soil quality (prime agricultural soil or not), tree cover percentage, and slope, which may affect the agricultural returns and thus the farmer's willingness to sell or the bid for sale; the forest canopy cover; and the landscape-view variability and aesthetic quality (e.g., Serneels and Lambin 2001; Müller and Zeller 2002). Two variables were used to represent the pressure from population growth (population density of the township) and socioeconomic status (household median income in the township), which affect the residential market and development decisions (e.g., Irwin and Bockstael 2001; Serneels and Lambin 2001). Q3

We collected the data for all these variables at each of the five time steps through GIS-based analysis (for

Table 1. Variables used in the models

Type	Variables	Description (unit in parentheses)
	s_time	The time that a particular parcel survives, represented as intervals (lower, upper) in some cases (decade)
	type	The land-use type that a parcel was or is at a time—used with s_time in survival models
Geographical	dist_ctyrd	Distance from the parcel or subdivision centroid to the nearest county road (m)
	dist_dtw	Distance from the parcel or subdivision centroid to Detroit (km)
	dist_5ct	Distance from the parcel or subdivision centroid to the nearest city among five midlevel cities[a] (km)
	dist_all	Distance from the parcel or subdivision centroid to the nearest urbanized area[b] (km)
	dist_wtr	Distance from the parcel or subdivision centroid to the nearest lake or stream (km)
	dist_hwy	Distance from the parcel or subdivision centroid to the nearest highway[c] (km)
Biophysical	ptcover	Percentage tree cover, represented as percentage of parcel area with identifiable tree cover
	soil	Soil quality, represented as 2 (prime soil) and 1 (nonprime soil)
	slope	Parcel slope (the ratio between the rise and the run)
Socioeconomic	pop_d	Number of people per square kilometer (people/km^2)
	income	Median household income at the township level over five decades ($1,000)

Notes: [a]Lansing, Jackson, Flint, South Lyon-Howell-Brighton, and Ann Arbor. [b]Defined by U.S. Census. [c]The straight (Euclidean) distance between the centroid and the nearest highway in kilometers using the Arc/Info command "near."

all the geographical and biophysical variables) and governmental archives (for all the sociodemographic variables). Our dataset, a space–time composites model, is characterized by (1) all the geographical and biophysical variables having only spatial heterogeneity with no temporal variations, and (2) all the socioeconomic and demographic variables having temporal heterogeneity, but with spatial variation represented only available at the township level. Due to our focus on methodological issues in relation to survival analysis in land change science as well as space limitations, we leave issues such as variable selection, theories behind such selections, and the degree of spatial and temporal heterogeneity in the independent variables to another related article (An et al. forthcoming).

Data Interpretation and Integration

We chose 1950 to be the time origin to assure that the time expansion is long enough to detect significant land changes given the availability of aerial photographs. For ease of illustration we excluded all the rural lots from our dataset and focused our analyses on farms and the three subdivision types because the development of rural lots is quite different from that of the three subdivisions in terms of both size and development style (An et al. forthcoming). As a result, we obtained a sample of 184 parcels or polygons. In this dataset, the survival time was measured at the decade level, and a variable "censor" took 0, 2, 3, and 4 to represent farm, country subdivisions, horticultural subdivisions, and remnant subdivisions, respectively. To facilitate trajectory modeling, we added another discrete variable "track" that describes seven trajectories. Although we had land change data for five decades, the trajectories only account for the timing of development with a binary distinction, earlier than (including) 1980 or later than 1980, to reduce the number of trajectories. This gives rise to variables that describe presence or absence of country subdivisions prior to 1980 or post-1980, horticultural subdivisions prior to 1980 or post-1980, and remnant subdivisions prior to 1980 or post-1980, for a total of six trajectories. The last (seventh) trajectory is no development at all through the fifth decade (prior to 1980 or post-1980). All the time-dependent variables took the values in the decade prior to the development.

We also generated a piecewise dataset (Appendix C). This dataset contained 488 usable parcel-period records. If the development occurred during the first period (thus a parcel had only one record), this record would have a survival time recorded as (., 1), indicating that the development occurred earlier than the end of the first period, with the start date unknown and represented with the dot "." (i.e., left censored). If the development did not occur up to the end of the fifth period (i.e., a parcel was not developed by the late 1990s or the end of our time frame), each of the first five records for the first five periods would have its survival time described as (1, .), implying total survival of the corresponding period with right censoring (Appendix A); the last record (i.e., record six) would have its survival time described as (0, .), which means it was still undeveloped after the end of our time frame (i.e., right censored). If the development occurred sometime in between Time 1 and Time 5, e.g., during the third period, the survival time for the first, second, and last record would be (1, .), (1, .), and (0, 1), suggesting the parcel or polygon survived the first two periods, but was developed between the start and end of the third period (interval censoring).

As in the one parcel → one record format, a variable "censor" took 0, 2, 3, and 4 to represent farm, country subdivisions, horticultural subdivisions, and remnant subdivisions in the piecewise dataset, respectively. All the time-dependent variables take their corresponding values in a manner that is consistent with the period in which a particular parcel period dwells. To facilitate the logit and complementary log-log modeling, we added a binary variable rSub, using one if the parcel was developed into remnant subdivision at the corresponding parcel period, and zero if not. As mentioned earlier, if a parcel was not developed until the end of our time frame, we have six parcel periods, and the last one has a (0, .) survival time. This last record was coded as no-data for the binary variable rSub and thus excluded from analysis in the logit or log-log models.

Data Analysis

Two major steps comprise our data analysis: analysis of temporal variability and uncertainty and estimates of several survival models. The analysis of temporal variability and uncertainty illustrates the overall temporal variations in the hazards and survival probabilities. Next, we compare and contrast the results from varying survival models. We primarily used the procedures in SAS (version 9.1) that were specifically designed for survival analysis, such as lifetest, lifereg, and phreg. Although SAS was employed as the analysis tool throughout this research, the concepts, methods, and results in this article are not software specific. Due to space limitations, we only report our models for remnant subdivisions for demonstration purpose. Some variables we

Table 2. Differences in results (remnant subdivisions only) caused by time-dependent variables, censoring, and different models

	Model No.					
					5. Trajectory	
	1. Cox	2. Piecewise exponential model[a]	3. Logit	4. Complementary log-log	12 (Prior 1980)	12 (Post 1980)
Intercept		−1.9182	−2.0150	−1.9710	0.0851	1.6614
dist_ctyrd	0.0012	0.0009	0.0011	0.0010		
dist_dtw	**−0.0282**	**−0.0207**	**−0.0202**	**−0.0201**	−0.0198	**−0.0567**
dist_5ct	**−0.0475**	**−0.0293**	−0.0260	−0.0253	**−0.0333**	**−0.0374**
dist_all	0.0922	0.0499	0.0466	0.0434	0.0459	**0.0256**
ptcover	0.0068	0.0114	0.0078	0.0076	**−0.0208**	**0.0113**
Soil	0.0754	0.0158	−0.0443	−0.0336		
Slope	**0.2219**	**0.2360**	**0.2733**	**0.2278**	**0.6294**	**0.5653**
pop_d	−0.0679	−0.0102	0.0014	0.0017		
Data format	1 parcel, 1 record		Piecewise data		**1 parcel, 1 record**	
Sample size	137	586	488	488	**184**	
−2 log L	193.70	216.67	217.94	218.15	**504.08**	
Generalized R^2	0.1459	0.0328	0.0368	0.0364	**0.4140**	

Notes: Bold numbers are significant coefficients at the $\alpha = 0.10$ level.
[a]The coefficients obtained from the lifereg (dist = exponential option) procedure were inverted in signs so that the coefficients in the entire table are consistent in meaning (Allison 1995, 62–68).

chose earlier (Table 1), such as income, distance to lakes and streams, and distance to highways, were found insignificant and thus excluded in the regression models presented (Table 2).

We first estimated a Cox model (Model 1) that handles time-dependent variables and right-censored survival times using the one parcel → one record dataset (see http://geography.sdsu.edu/People/Pages/an/SA_app.pdf for code). Our preliminary analysis showed that distance to the midlevel cities may have played a varying role over time in affecting residential land-use decisions (An et al. forthcoming), so we tested if any nonproportionality arose from this variable by adding an interaction term between this distance and time in the piecewise exponential model to be introduced next. If its coefficient was significant, the effects of the distance became larger over time when the sign is positive, or smaller over time when the sign is negative.

Next, we employed the piecewise dataset to integrate time-dependent variables, all types of censoring, and competing risks in a piecewise exponential model using the three-step strategy (Model 2). Following these two "classical" survival models, we estimated two models that are designed to cope with coarse time measurements, a logit model (Model 3) and a complementary log-log model (Model 4). Both models have time-dependent variables incorporated during the formation of the parcel periods in the piecewise data (Appendix C), but are weak in handling competitive risks and censoring. For a certain parcel, the last parcel period during which the parcel was developed into types other than remnant subdivision had to be dropped. Last, to shed light on whether and how the effects of some explanatory variables vary over time and test the effectiveness of the trajectory model we estimated a trajectory model using the last (seventh) trajectory (no development prior to 1980 and post-1980) as the reference level for the discrete variable track.

Results

The survival probabilities for the three types of subdivisions declined or remained nearly constant (country subdivisions during the third and fourth decades; Figure (7A), which conforms to the nonincreasing nature of survival probabilities. The hazard curves of three subdivisions differed in shape and timing. Country subdivisions showed a relatively flat trend, and horticultural and remnant subdivisions showed increasing hazards after the third and fourth decades, respectively (Figure 7B). Henceforth we report only the results in relation to remnant subdivisions due to space limitations.

The Cox model resulted in three significant independent variables (Table 2). The negative coefficients for distance to Detroit (dist_dtw) and five midlevel cities (dist_5ct) suggest that parcels near these cities

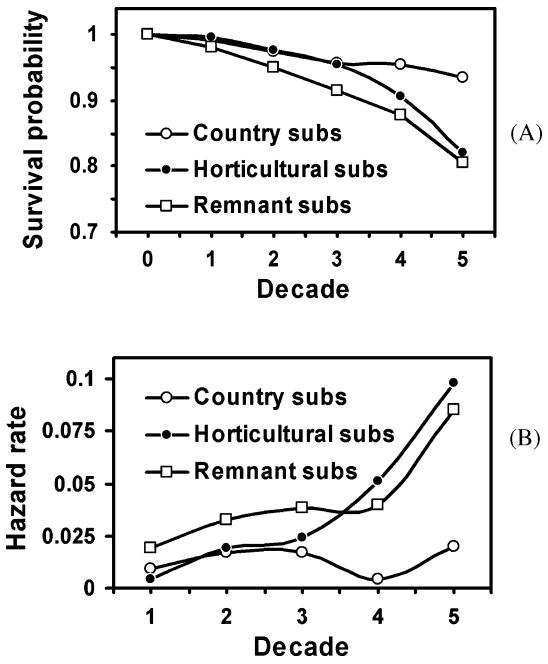


Figure 7. (A) The survival probabilities of three types of subdivisions, and (B) the hazard rates of three types of subdivisions over a span of fifty years.

have relatively lower hazards. Similarly, parcels associated with higher slopes had higher hazards to get developed into remnant subdivisions (Model 1, Table 2). In addition to these findings, the piecewise exponential model shows that percentage tree cover had a positive effect on the hazards (Model 2, Table 2). Although slightly different in the magnitudes of coefficients, both the logit and complementary log-log models had only two significant independent variables, suggesting parcels that were close to Detroit or in places with higher slopes should have higher hazards to be developed into remnant subdivisions (Models 3 and 4, Table 2). The trajectory model gave consistent and additional insight into the potential drivers (Model 5, Table 2). Distance to Detriot did not matter at earlier times, but it seems that as time moved on, places closer to Detroit had higher hazards of being developed into remnant subdivisions. Distance to the five midlevel cities had the opposite result: At earlier times places closer to the five cities had higher hazards of being developed, but this trend attenuated over time. Percentage tree cover had a negative relation with the hazard at earlier times, but not so at later times. Slope was positively related to the hazards regardless of time.

When the interaction between time and distance to the five midlevel cities (dist_5ct) was introduced, the overall fit of the model (Model 2) had a significant improvement: the –2 Log L decreased from 216.67 to 210.93 with the change in degrees of freedom equal to one, giving rise to a p value of 0.0166 in a test of the null hypothesis that this interaction term is insignificant (not shown in Table 2). The coefficients for dist_5ct and the interaction dist_5ct*time were –0.0815 and 0.0115, which can be combined as $(-0.0815 + 0.0115 \times \text{time}) \times \text{dist\_5ct}$, suggesting the parcels closer to the five cities had higher hazards of being developed into remnant subdivisions, but this trend dwindled over time.

Discussion and Conclusions

Due to our focus on methodology, we do not elaborate on our empirical findings (for details, see An et al. forthcoming). The power of detecting temporal trends of some land change events under study (Figure 7) is one of the most intriguing features of survival analysis; such graphs help to answer the questions raised earlier in this article: By a certain time, what proportion of land-use type i was or would be converted to type j ($i, j = 1, 2, \ldots, n, i \neq j$) (Figure 7A) or was the transition rate from land-use type A to B the same as (or faster than) that from A to C? If not, what factors led to such differences? (Figures 7B). Next we describe how the different models we used may complement each other, summarize the general strengths and weaknesses of survival analysis, and point to future directions.

Complementarity of the Models

All five models (Table 2) were consistent in estimating the effects of distance to Detroit and slope (dist_dtw), in terms of the sign, the magnitudes, and whether they were significant. The aggregated data corroborate such findings: Remnant subdivisions had shorter distances to Detroit compared to other land-use types except horticultural subdivisions, and the highest slope among all the land-use types (Table 3). Unlike

Table 3. Average distances to Detroit, the five midlevel cities, and small urban areas, soil, and slope of the four land-use types

	Farm	Country subdivisions	Horticultural subdivisions	Remnant subdivisions
dist_dtw (km)	74.36	69.85	59.36	63.38
dist_5ct (km)	28.31	26.98	19.44	23.30
dist_all (km)	12.61	12.96	12.00	13.14
Soil	1.22	1.42	1.19	1.25
Slope	1.02	0.99	1.14	2.23

the two models that do not consider censoring (the logit and complementary log-log models), the two classical survival models (Models 1 and 2 in Table 2) both identified a negative relation associated with the distance to the five midlevel cities, which can be supported in the following respects. First, this negative relationship was partially corroborated by the trajectory model with a negative coefficient prior to 1980 (i.e., −0.0333). Second, when the interaction between the distance to the five midlevel cities and time was added, the result makes more sense in the context of the immobility of land and nonincreasing supply of land: At earlier times, residents may choose to live closer to such cities due to considerations of the job and recreational opportunities (represented by the negative coefficient for dist_5ct). Over time new developments of remnant subdivisions may have to be located farther away from such cities (represented by the positive coefficient for the interaction) either due to the decreasing land supply in areas closer to the cities, the increasing demand from an increasing human population, or even a shift of preference to living "out of town" (Hansen et al. 2005). Finally, our aggregate data support that remnant subdivisions are associated with a shorter distance to midlevel cities (Table 3). Although we cannot guarantee this finding is true (can any statistical models do so?), we can claim that the two classical models performed better, which might partially arise from their ability to incorporate censored data.

A further comparison between the Cox model and the piecewise exponential model suggests that the latter may outperform the former in this study because it revealed a positive relationship between the hazard and percentage tree cover, which is an element in the definition of remnant subdivisions. Many factors may account for this difference, possibly including the fact that the piecewise exponential model has the capacity to handle left- and interval-censored data, which the Cox model cannot, at least in SAS for the time being. The trajectory model revealed some relationships (e.g., effects of slope) found in other models, and provided several new points that other models, at least at first glance, failed to provide. An example of such new points is the changing effects of distance to the five midlevel cities. In this sense, it is a very useful method in detecting temporal changes. It is relatively demanding in terms of data, however, and sometimes researchers might have to reduce the number of trajectories to increase the degrees of freedom, as we did. It also suffers from (1) relatively small sample size and loss of information while forming the trajectories; and (2) unevenness of data distribution; for example, the second trajectory (developed into country subdivisions prior to 1980) has only four observations, which may account for the negative coefficient for the variable percentage tree cover (–0.0208, inconsistent with estimates from other models and less defensible; Table 2). Given such considerations, we suggest that researchers be cautious in using the trajectory method and drawing conclusions from it when the sample is not large enough due to budget, time, or technical constraints. The trajectory method may be helpful, however, for exploratory or diagnostic purposes, such as an inspection of the changing effects of some variables over time.

Survival Analysis: Strengths, Caveats, and Future Directions

Land change science, or GIScience in a broader sense, has suffered from a lack of match between its space–time data models and those traditional statistical models as reviewed in the "Background" section. Among the four types of complexities (i.e., spatial complexities, temporal complexities, implicit dynamic information, and land-unit complexities) existing in land change data, such traditional statistical models are strong in handling spatial complexities; however, they cannot effectively address temporal complexities and implicit dynamic information, which are usually implicitly accommodated in many space-time datasets and very important in land change analysis and modeling. In the context of precise or coarse time measurements, we have shown that survival models, treating time in a relative view (i.e., a quality of the phenomenon under investigation), are excellent tools for addressing these two kinds of complexities, and to some degree, the land-unit complexities for the following reasons.

First, survival analysis models are powerful in dealing with time complexities existing in the three types of censored data. We demonstrated the effects of censored data in theory (e.g., Appendix B), and partially in the case study, where the two classical survival models that consider censored data (i.e., Models 1 and 2) outperform the three models without considering censored data to some degree (i.e., Models 3, 4, and 5). Second, survival analysis handles the challenge of implicit dynamic information by virtue of the very useful concept or metric of hazard and its capacity to incorporate time-dependent variables. The concept of hazard encapsulates the history of all the past and surrounding events, which may be a benefit of using the hazard concept in land change analysis. Hazard is an intrinsic property of the individual of interest, which may capture the risk

of a land unit being developed (or developed into a certain type) at any time, reflecting the popularity or the attraction or attrition of a certain land type in a location: High hazards mean high attraction and low attrition, and vice versa. This popularity can be decomposed into components that relate to some exogenous or endogenous variables, including time. Thus in analyses that incorporate feedback or agent–environment interactions, researchers can conveniently define a hazard function for a certain type of individual or agent and let it correlate with the behavior of other individuals or the changing environment. Therefore, exploring the theoretical hazard functions of some events may help to reveal the mechanisms of some processes and, particularly in land change science, help to develop theory and models that can "predict trajectories of land-use change as . . . for vegetation succession" (Hansen and Brown 2005). Development of such theory and models may be especially important for some irreversible events or events that take a long time to occur. Third, and last, survival analysis has an excellent capacity to tackle competitive risks, part of land-unit complexities, through treating land changes to types not being considered as right censored, although not much work has been devoted to this issue here due to space limitations. For the other part of the land-unit complexities, the changing boundaries of land units, survival analysis does not make it better or worse, and we propose a most recent parcel (or collection of cells) framework to address this issue.

More specifically, we associate the preceding survival models with the space–time data models that connect to varying time measurement precisions. For instance, the logit or log-log models may be more appropriate for space–time data represented in the snapshot model, whereas the Cox model and the AFT models (especially the piecewise exponential model) are excellent in handling data represented in the space–time composites model, the spatiotemporal object model, and the three-domain model. This association can be flexible, however, depending on many factors, such as the questions under investigation and how precise we perceive the temporal data to be. For instance, we applied the four survival models to the same data we had derived in southeastern Michigan (with some data restructuring targeted on different models), and all models gave largely similar results. It is more likely that the Cox and piecewise exponential models will outperform the logit and complementary log-log model in many instances for reasons we already discussed (e.g., handling censored data), even when the time measurements are coarse or of unknown precision. For data of coarse time precisions, the logit or the complementary log-log models may be worth trying.

Of particular interest is the piecewise exponential model in the context of the three-step strategy, not only because it has generated the richest set of very insightful empirical estimates (e.g., the positive coefficient of percentage tree cover), but also because it may have the potential to address the effects of changing shape, boundary, or size for particular land parcels and units. When the time frame is discretized to several periods and all the parcel periods are treated as independent observations, then the changes in shape, size, or boundary, if any, might not affect the model substantially. The key point may be that the period should be short enough such that the land parcels can be regarded as constant in their shape, size, or boundary, but more theoretical and empirical studies are needed in this regard. On the other hand, if survival analysis is not used for reasons like the changing identities or boundaries, people will have to use other methods such as the ordinary least squares (OLS) or regular logistic regression models without explicitly considering time-dependent variables or censored data. In OLS, for instance, researchers may still have to use the survival time (but with no consideration of censoring) of each parcel as the dependent variable, ignoring the fact that the parcel can change in shape, boundary, and size over time.

The piecewise exponential model has two additional merits. First, it incorporates competitive risks, all types of censoring, and some degree of time-dependent variables. Second, although assuming a constant hazard within each period, this model allows for changing base hazards over periods (i.e., the β_j may change over time in Equation 6), which may serve as an approximation of any varying hazards over time. In land change modeling, land supply usually declines or remains constant over time, which may contribute to an increasing hazard over time given other conditions in control. One approach to testing if β_j changes over time is to use the variable of decade as an independent variable in the piecewise exponential model, creating several dummy variables that each represent one decade. All such decadal dummies were insignificant in our case study (e.g., $p > 0.99$), suggesting that in our case study, the β_j do not change significantly over decades.

This article focuses on introduction of the concepts and models in survival analysis, and their potential applications in geographical land change science. Our choices of hazard functions are largely based on classical survival literature, and research may be needed to

connect such choices with economic theories in land change analysis (e.g., Irwin and Geoghegan 2001). Similarly, economic theories may help us to choose independent variables, which may better explain the economic returns and costs associated with different types of development. Another issue may be how to choose the origin of time measurements, which may affect the coefficient estimates to varying degrees. In survival analysis literature, researchers either choose a time origin that marks the onset of continuous exposure to risk of the event, the time of randomization to treatments in experimental studies, or the time from which the strongest variable(s) takes effect (Allison 1995, 22–25; Kalbfleisch and Prentice 2002, 12–13). In land change analysis, we may choose time of origin using the first criterion, the third criterion, or a combination of both. The second may apply when different land polices are used in two places as a comparative study. This is an area that needs further research in the application of survival analysis in land change science.

There has been a concern regarding the correlatedness between different events in land change studies, which may be connected to land immobility and scarcity. Put another way, the development of one parcel in one location at a certain time may reduce the supply of available land at other locations and in the future, which may affect the decisions of developers and homebuyers considering other parcels. This problem, however, is not more serious in survival analysis than in other types of models. For instance, in a logistic model that regresses the log odds of development (i.e., log $[p/(1-p)]$, where p is the probability of development) against a set of explanatory variables, how can the probability of development for a parcel at a time not be affected by a nearby parcel that was developed earlier? Survival analysis does not alleviate or worsen this problem; we use the explanatory variable population density (pop_d in Table 2) to account for this issue, because this density will presumably have a negative relationship with land availability, and any changes in the development hazard caused by an earlier development should result in a rise in the population density. In our piecewise exponential survival model (Model 2 in Table 2), population density was insignificant, showing the hazards of development might not have been substantially affected by changes in land supply as a result of population increase, which is understandable in an exurban setting with relatively abundant land supply.

Our work in this article extends survival analysis methods to the field of geographical land change studies, exploring topics such as how survival analysis concepts can be employed to address complexities in land change analysis, what insights survival analysis can bring to the field, and its strengths and weaknesses. To help researchers better understand the concepts and processes, we have included some details such as the data format and original code in the appendices and the Web site (http://geography.sdsu.edu/People/Pages/an/SA_app.pdf). We expect that the approach that integrates survival analysis with GIS and remote sensing data will help land change researchers to better understand the dynamic landscapes, and find more applications and theoretical effort in researching survival analysis in land change science.

Acknowledgments

We gratefully acknowledge financial support in the form of a grant from the National Science Foundation Biocomplexity in the Environment Program (BCS-0119804).

Notes

1. The term *object* in object orientation is different from that term in geography. In computer science, object orientation usually means the encapsulation of both structures and procedures, hierarchical structure of different types of objects, and inheritance (i.e., a child object can automatically possess, or inherit, the properties of its parental object; see Bian 2003 for more details).
2. Little research has been devoted to study the effects of time frame (i.e., over how long we should collect the time series data) and resolution (i.e., at what time interval we should collect the data) on results of land change analysis. This article does not directly focuses on these questions, but partially addresses how the uncertainty caused by censored data may affect analysis results.
3. If an individual parcel had been part of a larger parcel subdivided during the study period, we kept it in its most recent size and shape and traced back to determine earlier land-use types. If the parcel was part of a subdivision, it was merged with nearby parcels before tracing the resulting polygon back.
4. Each of the eight townships has a slightly different sequence of image intervals due to aerial photo availability.

References

Agarwal, C., G. M. Green, J. M. Grove, T. P. Evans, and C. M. Schweik. 2002. *A review and assessment of land-use change models: Dynamics of space, time, and human choice*. Newton Square, PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station, and Center for the Study of Institutions, Population, and Environment Change (CIPEC).

Allison, P. D. 1995. *Survival analysis using SAS®: A practical guide*. Cary, NC: SAS Institute.

An, L., D. G. Brown, J. Nassauer, and B. Low. Forthcoming. Timing, location, and determinants of residential development types in exurban southeastern Michigan.

Armstrong, M. P. 1988. Temporality in spatial databases. In *Proceedings of GIS/LIS'88, 2*, 880–89. Bethesda, MD: American Congress of Surveying and Mapping.

Baker, W. L. 1989. A review of models of landscape change. *Landscape Ecology* 2 (2): 111–33.

Baltzer, H. 2000. Markov chain models for vegetation dynamics. *Ecological Modeling* 126:139–54.

Banerjee, S., and D. K. Dey. 2005. Semiparametric proportional odds models for spatially correlated survival data. *Life Time Data Analysis* 11:175–91.

Banerjee, S., M. M. Wall, and B. P. Carlin. 2003. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics* 4 (1): 123–42.

Bian, L. 2003. The representation of the environment in the context of individual-based modeling. *Ecological Modeling* 159:279–96.

Bodian, C. A. 1985 A piecewise exponential model for comparing follow-up data with an external standard . *Biometrics* 41 (1): 333–33.

Boracchi, P., E. Biganzoli, and E. Marubini. 2003. Joint modelling of cause-specific hazard functions with cubic splines: An application to a large series of breast cancer patients. *Computational Statistics & Data Analysis* 42:243–62. Brown, D. G., K. M. Johnson, T. R. Loveland, and D. M. Theobald. 2005. Rural land use change in the conterminous U.S., 1950–2000. *Ecological Applications* 15 (6): 1851–63.

Brown, D. G., B. C. Pijanowski, and J. D. Duh. 2000. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management* 59:247–63.

Brown, D. G., D. T. Robinson, L. An, J. I. Nassauer, M. Zellner, W. Rand, R. Riolo, S. E. Page, and B. Low. Forthcoming. Exurbia from the bottom-up: Confronting empirical challenges to characterizing complex systems. *GeoForum*.

Cantor, A. B. 2003. *SAS® survival analysis techniques for medical eesearch*. 2nd ed. Cary, NC: SAS Institute.

Clark, W. A. V., and S. D. Withers. 1999. Changing jobs and changing houses: Mobility outcomes of employment transitions. *Journal of Regional Science* 39 (4): 653–73.

Coomes, O. T., F. Grimard, and G. J. Burt. 2000. Tropical forests and shifting cultivation: Secondary forest fallow dynamics among traditional farmers of the Peruvian Amazon. *Ecological Economics* 32:109–24.

Cox, D. R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* B34:187–220.

Du, W. B., K. S. Chia, R. Sankaplanarayanan, R. Sankila, A. Seow, and H. P. Lee. 2002. Population-based survival analysis of colorectal cancer patients in Singapore, 1968–1992. *International Journal of Cancer* 99 (3): 460–465.

Hansen, A. J., and D. G. Brown. 2005. Land-use change in rural America: Rates, drivers, and consequences. *Ecological Applications* 15 (6): 1849–50.

Hansen, A. J., R. L. Knight, J. M. Marzluff, S. Powell, K. Brown, P. H. Gude, and K. Jones. 2005. Effects of exurban development on biodiversity: Patterns, mechanisms, and research needs. *Ecological Applications* 15 (6): 1893–905.

Hsiao, C. 1986. *Analysis of panel data* (Economic Society Monographs No. 11). Cambridge, U.K.: Cambridge University Press.

Irwin, E., and N. Bockstael. 2002. Interacting agents, spatial externalities, and the endogenous evolution of residential land-use pattern. *Journal of Economic Geography* 2:31–54.

Irwin, E., and J. Geoghegan. 2001. Theory, data, methods: Developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment* 85 (1–3): 7–23.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The statistical analysis of failure time data*. 2nd ed. Hoboken, NJ: Wiley.

Karrison T. 1996. Confidence intervals for median survival times under a piecewise exponential model with proportional hazards covariate effects. *Statistics in Medicine* 15 (2): 171–82.

Klein, J. P., and M. L. Moeschberger. 1997. *Survival analysis: Techniques for censored and truncated data*. New York. Springer-Verlag.

Kleinbaum, D. G., and M. Klein. 2005. *Survival analysis: A self-learning text*. 2nd ed. New York. Springer-Verlag.

Lambin, E. F. 1997. Modelling and monitoring land-cover change processes in tropical regions. *Progress in Physical Geography* 21 (3): 375–93.

Lambin, E. F., M. D. Arounsevell, and H. J. Geist. 2000. Are agricultural land-use models able to predict changes in land-use intensification? *Agriculture, Ecosystems & Environment* 82 (1–3): 321–31.

Langran, G., and N. R. Chrisman. 1988. A framework for temporal geographic information. *Cartographica* 25:1–14.

Liestol, K., P. K. Andersen, and U. Andersen. 1994. Survival analysis and neural nets. *Statistics in Medicine* 13 (12): 1189–200.

Luijten, J. C. 2003. A systematic method for generating land-use patterns using stochastic rules and basic landscape characteristics: Results for a Colombian hillside watershed. *Agriculture, Ecosystems & Environment* 95 (2–3): 427–41.

Machin, D., Y. B. Cheung, and M. K. B. Parmar. 2006. *Survival analysis in practical approach*. 2nd ed. Chichester, U.K.: Wiley.

Mertens, B., and E. F. Lambin. 2000. Land-cover-change trajectories in southern Cameroon. *Annals of the Association of American Geographers* 90:467–94.

Mertens, B., W. D. Sunderlin, O. Ndoye, and E. F. Lambin. 2000. Impact of macroeconomic change on deforestation in south Cameroon: Integration of household survey and remotely-sensed data. *World Development* 28 (6): 983–99.

Müller, D., and M. Zeller. 2002. Land-use dynamics in the central highlands of Vietnam: A spatial model combining village survey data with satellite imagery interpretation. *Agricultural Economics* 27:333–54.

Murthy, V. K., and L. J. Haywood. 1979. Survival analysis by sex, age group and hemotype in sickle-cell disease. *Age* 2 (4): 132–32.

Odland, J., and M. Ellis. 1992. Variations in the spatial pattern of settlement locations: An analysis based on proportional hazards models. *Geographical Analysis* 24 (2): 97–109.

Peuquet, D. J. 2001. Making space for time: Issues in space–time data representation. *Geoinformatica* 5 (1): 11–32.

Peuquet, D. J., and N. Duan. 1995. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9 (1): 7–24.

Plantinga, A. J., and E. I. Irwin. 2006. Overview of empirical methods. In *Economics of rural land-use change*, ed. K. P. Bell, K. J. Boyle, and J. Rubin, 113–34. Aldershot, U.K.: Ashgate,.

Sengupta, S., and D. E. Osgood. 2003. The value of remoteness: A hedonic estimation of ranchette prices. *Ecological Economics* 44:91–103.

Serneels, S., and E. F. Lambin. 2001. Proximate causes of land-use change in Narok District, Kenya: A spatial statistical model. *Agriculture, Ecosystems & Environment* 85 (1–3): 65–81.

Seto, K. C., and R. K. Kaufmann. 2003. Modeling the drivers of urban land-use change in the Pearl River Delta, China: Integrating remote sensing with socio-demographic data. *Land Economics* 79 (1): 106–21.

Shuster, J. J. 1992. *Handbook of sample size guidelines for clinical trials*. Boca Raton, FL: CRC Press.

Starmer C. F. 1988. Characterizing activity-dependent processes with a piecewise exponential model. *Biometrics* 44 (2): 549–59.

Turner, R., C. M. Newton, and D. F. Dennis. 1991. Economic relationships between parcel characteristics and price in the market for Vermont forestland. *Forest Science* 37 (4): 1150–62.

Vance, C., and J. Geoghegan. 2002. Temporal and spatial modeling of tropical deforestation: A survival analysis linking satellite and household survey data. *Agricultural Economics* 27: 317–32.

Vigmostad, K. E. 2003. Michigan real estate developer perspectives on development, sustainability, and nature: An autoethnography. PhD dissertation.

Worboys, M. F. 1992. A model for spatio-temporal information. In *Proceedings: The 5th International Symposium in Spatial Data Handling, 2,* ed. P. Bresnahan, E. Corwin, and D. Cowen, 602–11. San Jose, CA: American Congress of Surveying and Mapping.

Yuan, M. 1999. Use of a three-domain representation to enhance GIS support for complex spatiotemporal queries. *Transactions in GIS* 3 (2): 137–59.

Appendix A: Risk and Risk Set

When an individual has not experienced the event(s), it is at risk; after the event occurs, the individual is not at risk. For coarse time measurements in land change analysis (e.g., Figure 5C), we consider an individual at risk during the time interval it has been developed, for example, (T2, T3). For instance, Person 3 (Figure 4A) or Parcel P10 (Figure 4B) are at risk before T4, and we are unsure whether Person 1 or Parcel P8 are at risk at any time before T3—but we know they are not at risk after T3. When considering competitive risks (e.g., a person may die from one of multiple different diseases, or a parcel may be developed into one of several different land-use types), an individual experiencing an event of the type not being considered is said to have a survival time right censored for the type being considered (Kleinbaum and Klein 2005, 400). For instance, Parcel P10 is developed into land type B, and Parcel P8 type A (Figure 4B). If we are considering land type A, then Parcel P10 is considered to be right censored at T3, which also means after that time it is no longer at risk of being developed into type A. When considering the event(s) at the aggregate level, those individuals that are at risk (or have not experienced the event) constitute the risk set.

Appendix B: How Censored Data Are Accounted for in the AFT Model

The AFT model can be expressed in matrix form for more generality: $\log(\mathrm{h}(\mathbf{t})) = \mathbf{X}\beta + \sigma\varepsilon$, where σ is an unknown scale parameter, and ε is a vector of errors assumed to come from a known distribution (e.g., the standard normal distribution). In estimating the vector β, the Newton–Raphson algorithm is usually used to find parameters that maximize the likelihood of the given data (called maximum likelihood estimation, or MLE) that have g,h,k, and m uncensored, right-censored, left-censored, and m interval-censored observations:

$$L = \prod_{i=1}^{g} f_i(t_i) \times \prod_{i=1}^{h} S_i(t_i) \times \prod_{i=1}^{k} F_i(t_i) \times \prod_{i=1}^{m} (F_i(u_i) - F_i(v_i)) \quad (8)$$

where $f_i(t_i)$, $S_i(t_i)$,and $F_i(u_i)$ are population density function (PDF), survival function, and cumulative distribution function (CDF) of the survival time t_i. (v_i and u_i are the lower and upper bounds of a time interval when the event occurs.) This shows how uncensored and all types of censored data are all combined to make the MLE. In SAS, the lifereg procedure is developed for this purpose. Instead of using hazards as the response variable, it uses the failure time t. If we decide to evaluate the Weibull model in Equation (3) in terms of failure time t (i.e., $\log(t)) = \mathbf{X}\beta^* + \alpha\log t$), there exists a nice relationship that links the coefficients of the

preceding two models: $\beta = -\beta^*/\sigma$ and $\alpha = 1/\sigma - 1$. As $\alpha = \sigma = 1$, so $\beta = -\beta^*$ for exponential models.

Appendix C: Piecewise Data Restructuring

A common practice in survival analysis is to conduct piecewise data restructuring for some type of analysis (Shuster 1992). Due to various reasons such as data availability or the researcher's intentional choice, the entire frame may be divided into several periods (three in Figures 5C and 5D), which are usually of equal length, but not necessarily so (Allison 1995, 109). For all the periods during which the parcel is at risk, we record the information related to all the dependent and independent variables, forming data of so-called parcel periods. In such parcel periods, a specific independent variable may take changing or constant values depending on whether or not it is a time-dependent variable, and the dependent variable may be the survival time within that specific period and the associated censoring status (Figure 5C), or a binary development status (Figure 5D), depending on the type of model that is to be constructed. In Figure 5C, the top parcel has one parcel period, where the survival time is left censored; the middle parcel has two parcel periods with the first right censored at the end of the period, and the second interval censored at the time interval illustrated by the brackets; the bottom parcel has three parcel periods with the each right censored at the end of the period. In Figure 5D, we only have binary developed and undeveloped information (we use a variable dev_Status to represent it) by the nature of the data, so the top parcel has one parcel period with dev_Status equal 1 (developed); the middle parcel has two parcel periods, and the dev_Status equals 0 (undeveloped) and 1 (developed) for them, respectively; the bottom parcel has three parcel periods, and dev_Status equals 0 for all of them. Because the bottom parcel in Figure 5D survives beyond T4, a fourth parcel period may be formed as right censored; that is, (T4, .). Such data on parcel periods are treated as independent observations later in survival analysis models. A common concern may be the lack of independence among the parcel periods from the same parcel, which has been shown not to be a problem in the survival analysis literature (for details, see Allison 1995, 108, 223–25).

Correspondence: Department of Geography, San Diego State University, San Diego, CA 92182–4493, e-mail: lan@mail.sdsu.edu.