

# Design and implementation of a Space-Time Intelligence System for disease surveillance

Geoffrey M. Jacquez, Dunrie A. Greiling, and Andrew M. Kaufmann

BioMedware, Inc. 516 N. State Street, Ann Arbor, MI 48104, USA  
(e-mail: jacquez@biomedware.com)

Received: 13 November 2003 / Accepted: 25 October 2004

**Abstract.** Modeling chronic and infectious diseases entails tracking and describing individuals and their attributes (such as disease status, date of diagnosis, risk factors and so on) as they move and change through space and time. Using Geographic Information Systems, researchers can model, visualize and query spatial data, but their ability to address time has been limited by the lack of temporal referencing in the underlying data structures. In this paper, we discuss issues in designing data structures, indexing, and queries for spatio-temporal data within the context of health surveillance. We describe a space-time object model that treats modeled individuals as a chain of linked observations comprised of an ID, space-time coordinate, and time-referenced attributes. Movement models for these modeled individuals are functions that may be simple (e.g. linear, using vector representation) or more complex. We present several spatial, temporal, spatio-temporal and epidemiological queries emergent from the data model. We demonstrate this approach in a representative application, a simulation of the spread of influenza in a hospital ward.

**Key words:** Geographical information systems, Time, Temporal modeling, Indexing

## 1 Introduction

Our research on space-time information systems (STIS) has two goals. The first goal is to develop a library for programming STIS software. The second is to apply this library to create a software application (STIS-Influenza) for modeling an outbreak of influenza in a simulated hospital ward. These goals

---

This research was supported by grant R44ES010220 from the National Institute of Environmental Health Sciences (NIEHS) and by grants R01CA092669 and R01CA96002 from the National Cancer Institute (NCI). The content of this paper does not necessarily represent the official views of the NIEHS or the NCI.

were selected because their accomplishment would allow us to evaluate some of the constructs of time geography that appear relevant in epidemiology, especially for use in health surveillance software. We simulated an outbreak of influenza to provide an experimental system in which all parameters are known and under the researcher's control. This allowed us to evaluate scalability and to determine whether the epidemiological queries regarding infection events provided correct output. Should STIS-Influenza prove feasible, we believed we would then have a sound foundation for addressing a third goal, extension of the STIS software to real-world applications. This paper concerns itself with the first two research goals. Accomplishment of the third goal is described in the articles by Greiling et al. (2005) and Meliker et al. (2005) in this special issue.

To motivate development of the STIS we begin by presenting some current applications of GIS in cohort identification, exposure assessment, disease mapping and public health surveillance, followed by an assessment of the advantages that improved handling of the time dimension might provide. We then consider the importance of space-time processes and temporal lags in disease surveillance, and how these critical factors are not well-represented in conventional GIS. Several previous approaches to making GIS time enabled are then discussed, several of which are modifications to existing spatial data models. We consider these in light of the requirements of space-time disease surveillance and conclude that modification of existing spatial data models is unlikely to lead to a solution suited to the representation and modeling of human diseases. This conclusion motivates our development of space-time data structures for disease surveillance and sets the stage in the methods section for a description of data models and queries that underpin our Space-Time Intelligence System. The results section describes the STIS-influenza application and the discussion section identifies future research directions.

### *1.1 Shortfalls of health-GIS, and requirements of health-STIS*

Applications of health-GIS include cohort identification, exposure assessment, disease mapping, and public health surveillance (see Barnes and Peck 1994; Clarke et al. 1996; Croner et al. 1996; Jacquez 1998 for further discussion on GIS in health). Health conditions explored include reproductive outcomes (Stallones et al. 1992; Rushton and Lolonis 1996), vector-borne diseases (Kitron et al. 1994; Glass et al. 1995; Richards 1993), infectious diseases (Gould 1997), and cancer (Xia, Carlin et al. 1997); using methods including disease clustering (Kulldorff et al. 1997; Lawson and Waller 1996; Jacquez et al. 1996), exposure assessment (Holm et al. 1995), and health service location analysis (Teutsch and Churchill 1994). While these applications demonstrate significant contributions of spatial-only GIS, we argue that the techniques employed are constrained by the limited ability of GIS to deal effectively with time, as described below.

Demographic data, when coupled with exposure prediction and GIS operations (such as buffering and Boolean operations across coverages) support identification of study subjects that otherwise is difficult to accomplish. Examples include Croner et al. (1996) who characterized residential demography in proximity to toxic waste sites using GIS, and

studies of leukemia and EMF exposure that selected study subjects based on proximity-based exposure models (Wartenberg et al. 1993; Kleinerman et al. 2000). But time-varying exposures, daily activity spaces, and residential mobility are poorly represented in traditional GIS, and unrealistic assumptions regarding these factors must be made when conventional GIS are used. These assumptions may include constant exposure through time, that human beings are sessile, and that causative exposures occur at place of residence. Time-enabled approaches are needed that relax these unrealistic assumptions and make it possible to identify study subjects by incorporating time-dependent exposure models and by readily accounting for the movement of individuals on daily, weekly, and annual time scales.

GIS has been used in exposure assessment to reconstruct individual exposures to health-related factors (Wartenberg 1992, Hjalmarsson et al. 1996, Ward et al. 2000, Bellander et al. 2001, Elgethun et al. 2003) and to assess public health threats near hazardous sites (ATSDR 1992, Holm et al. 1995). Risk assessment in GIS has historically focused on the hazard as the object of interest – such as the locations of industrial sites of high concentration in pollutants that are known to be human carcinogens – instead of the individual (Mark et al. 2000). In addition, the latency period between exposure and disease onset often is ignored (Marbury 1996). These weaknesses in large part are due to the poor ability of current GIS to handle multitemporal geographic information and the movement of individuals. Not surprisingly, few studies explicitly state the temporal ordering of exposure variables and their inter-relationships, even though the ordering and timing of such events is critically important through such mechanisms as synergism, DNA repair mechanisms, and multi-hit models of carcinogenesis (Ben-Shlomo 2002). Time-enabled approaches are needed that make explicit the timing and ordering of exposures and that provide mechanisms for modeling the movement of individuals as well as temporal change in the magnitude and locations of exposure sources.

The creation of thematic maps through map overlays and operations across coverages is one of the principle functions of GIS, and the ability to account for spatially structured covariates has enhanced the construction of disease maps (Waller and McMaster 1997). But disease maps produced by GIS are static, and GIS software has made little progress beyond the traditional map worldview (Goodchild 2000). Static maps can be strung together to visualize how spatial disease patterns change through time (e.g. try <http://zappa.nku.edu/~longa/geomed/modules/sv/lec/pop.html> for a map of the US population through time, and <http://zappa.nku.edu/~longa/geomed/rabies/test.html> for a visualization of a raccoon rabies epidemic in the Eastern US). But these can be cumbersome to create and are not interactive. Time enabled techniques of map generation are needed in which change on maps is a native function. This will make possible interactive movies that can be viewed at all intermediate time points and that support queries regarding relationships among health outcomes in space and through time.

Teutsch and Churchill (1994) define surveillance to include the detection of epidemics, the assessment of infectious disease potential, and the design and evaluation of health interventions. Because of its ability to identify and map environmental factors associated with disease vectors, GIS has been

useful for infectious and vector-borne disease surveillance. Examples include Lyme disease (Glass et al. 1995), malaria (Kitron et al. 1994), and onchocerciasis (Richards 1993), among others. GIS is particularly useful for tracking adverse health outcomes in proximity to known hazards (e.g. monitoring of reproductive outcomes in mothers living near hazardous waste sites; Stallones et al. 1992). But almost all disease processes involve space-time lags in which the health response occurs at space-time locations other than where the causative exposure(s) occurred (Schaerstrom 1996). For example, in the multi-hit model of carcinogenesis damage to DNA accumulates over many years, and it is only after a series of mutations that the unregulated cell replication known as cancer begins. For infectious diseases the duration of the lag between exposure to the pathogen and onset of symptoms may be short (e.g. a few days for influenza) or long (months or even years for slow viruses and prions). Humans are mobile and the locations where an individual's exposure occurred and where the disease is manifested are rarely coincident, even for latencies on the order of a few days. Influenza infection transmission sites, for example, include school, work and public places such as shops and restaurants. But symptoms arise at a later date and the location of case appearance assigned in most surveillance systems is the patient's home. The static view of conventional GIS does not readily handle the space-time lag typical of human disease, motivating Loytonen (1998) and others to call for a "higher-dimensional GIS" or STIS. Because an individual's exposure events, disease status, and mobility can be handled in a STIS, Jacquez (2000) argued they are better suited to disease surveillance and the generation and evaluation of epidemiologic hypotheses.

### *1.2 A summary of approaches to time in GIS*

Current GIS access spatial data rapidly because their underlying data structures are organized to do so. Examples of such data structures include R-trees, Quad-trees and others. The spatial queries these data structures support are based entirely on spatial relationships between geographic locations, and exclude the time dimension. This motivated several researchers to treat time as a direct extension of current GIS data models, extending the grid data model to represent snap-shots (time stamping) of geographic data at different time intervals. Time-stamping supports the direct comparison of spatial data from different time slices, but quickly becomes unwieldy as the number of time slices increases. Hazelton (1991) and Langran (1992) proposed amendment vectors to extend the vector data model to the time dimension, but this approach is problematic because it treats space-time relationships as being imbedded in a fixed space-time geometry (Peuquet 1994).

There are numerous data models that can be adapted for representing dynamic exposure histories (see Miller 2004 for a review of spatio-temporal data modeling). Space time models must incorporate a space-time observation consisting of an object ID (unique identifier of the individual used throughout all recordings), a location (a spatial descriptor such as a xy coordinate, a city, or a street address), and a time value or time stamp (such as minutes or years). Yearsley and Worboys (1995), proposed a space time

object model that integrates abstract spatial data types with a geometric layer to construct a higher-level topological data model, Raper and Livingstone (1995) used an object oriented approach to represent dynamic spatial processes as spatio-temporal aggregations of point objects, and Peuquet and Duan (1995) formulated an event-based spatio-temporal data model (ESTDM) that maintains spatio-temporal data as a sequence of temporal events associated with a spatial object. Peuquet and Duan's (1995) data model deals effectively with both absolute and relative space-time relationships. The relative view treats space-time as relationships between objects, a phenomenology that is intuitively suited to epidemiological queries, which seek information regarding associations between health events. Whereas current GIS data models store information regarding *what* and *where* events occur, Peuquet and Duan's data model is designed to store *what*, *where* and *when*. To fully realize space-time modeling capabilities they argued that space-time relationships are best modeled using an event-driven approach. Accordingly, their ESTDM supports *both* absolute and relative spatio-temporal queries. Other attempts to incorporate time by event processing include the Oogeomorph (Raper and Livingstone 1995), three domain (Yuan 1995) and triad (Peuquet and Qian 1996) models. However, ESTDM and Oogeomorph deal only with point data, and the others of this genre are not fully defined (Cheng and Molenaar 1998).

In general, these attempts to deal with time fall into two camps: Extensions to spatial data structures (e.g. time stamping, amendment vectors), and process-based approaches (e.g. ESTDM, Oogeomorph). Extensions to existing spatial data structures are inherently problematic because they treat time as an attribute, thereby preclude queries optimized for space-time data structures, and are therefore too slow to be of practical use. The object designs of process-based approaches are specific to scientific fields (e.g. the Oogeomorph approach is appropriate to geomorphology) and thus are not general solutions.

Hagerstrand (1970) conceptualized the *space time path* as an individual's continuous physical movement through space and time, and visually represented this as a 3-dimensional graph. Hornsby and Egenhofer (2002) recognized that space-time paths mediate individual-level exposure to pathogens and environmental toxins, and that practical application would require a mechanism for representing location uncertainty. A *space time prism* refers to the possible locations an individual could feasibly pass through in a specific time interval, given knowledge of their actual locations in the times bracketing that interval. The *potential path area* (Miller 2004) shows the locations the individual could occupy given these constraints, and represents places where exposure events might occur. These constructs enabled new research approaches in diverse fields such as student life (Huisman and Forer 1998), sports analysis (Moore et al. 2003), social systems (Kwan 2000), transportation (Miller 1991), and the analysis of disparities in gender accessibility in households (Kwan 2003). But despite these applications, in general Hagerstrand's time geographic approach has not been exploited in disease surveillance, primarily because enabling software tools are lacking. In this research we evaluated the feasibility of using a space-time approach for implementing disease surveillance software.

## 2 Methods

Disease surveillance applications require space-time data structures that provide spatial, temporal and spatio-temporal queries on individuals, their disease status, and the duration and temporal ordering of exposure opportunities. This is fundamentally different from simply adding time to existing GIS (the “GIS-Time” approach), and we therefore use the term “Space-Time Intelligence System” (“STIS”) to describe the approach implemented in this research. This section describes our approach to constructing space-time information systems: the data structures; including the object-based data model; movement and attribute change models; space-time indexing; and the simple and complex queries possible within a STIS. We then describe a prototype application in the results section.

### 2.1 Cognitive framework

We can think of a general instance of a `SpaceTimeObject` as possessing the following characteristics: ID, spatial location, start time, stop time (optional), movement model, attributes, attribute change model, pointers to the previous and next objects in the *object chain* (described below) and the label of the parent chain. A complete set of objects composes the object chain that represents the individual’s movement through time. Mark, Egenhofer et al. (2000) use the term “geospatial lifeline” to describe an individual’s movements and exposure throughout a defined time period. Hornsby and Egenhofer (2002) use the term “lifeline thread” to describe an ordered sequence of connected space-time samples. Here we employ the term “object chain” since this construct can apply with equal validity to non-living things (such as airplanes) as well as to people.

This cognitive framework extends the diad (what, where) used in conventional GIS to the triad (what, where, when) necessary for spatio-temporal modeling. Each `SpaceTimeObject` identifies a modeled entity (i.e. an individual on a hospital ward). The spatial location and time stamps define space-time locations that may be space-time points (e.g. latitude, longitude, altitude, date, movement model) or a more complex object such as a space-time polygon (e.g. polygon centroid, polygon boundary; date, movement model). The movement model defines how the object moves through space as a function of time, the simplest non-static movement model being a vector and velocity defining linear movement at a constant speed. In addition, *morphing* can occur when the shape of a complex object, such as a line or polygon, changes through time. Morphing can be gradual, in which case the change in the object’s boundary occurs over a defined time interval; or it can be abrupt. Morphing is a ready mechanism for handling changes in cadastral systems such as realignment of administrative and political boundaries (see Meliker et al. 2005 for an example). *Attributes* are observations on variables describing the modeled entity (e.g. influenza infection status such as susceptible or infectious) and the *attribute change model* describes how those attributes change through time.

This cognitive framework can be applied to a variety of space-time objects. Examples includes cadastral systems where the entities are areas (polygons) that might be defined by type of land use (e.g. agricultural) or ownership; and

population-based disease data where entities are areas (e.g. counties) and the attributes are cancer rates within those counties. The implementation of this framework in software is defined at the application level by defining a space-time location along with its object ID, and by then extending this object definition through inheritance. This provides a useful mechanism for designing custom objects that retain the full functionality of ancestral types. For purposes of exposition we now assume the modeled entities are individuals, and we will concern ourselves with modeling space-time location and disease status at the individual level.

In the influenza application we need to model the locations of nurses and patients, and think of these entities as point locations in space-time (Eq. 1).

$$\mathbf{u}_{i,t} = \{x_{i,t}, y_{i,t}\} \quad (1)$$

Here the subscript  $i$  denotes the  $i^{\text{th}}$  object (patient or nurse), and  $x_{it}, y_{it}$  are the spatial coordinates of the  $i^{\text{th}}$  object at time  $t$ . For the influenza application each individual (nurse or patient) has a space-time location so long as he or she is on the hospital ward.

## 2.2 Object chains

*Object chains* represent individuals moving through time and space as chains of the space-time locations from Eq. 1. We then define the object chain specifying the space-time coordinates for individual  $i$  as:

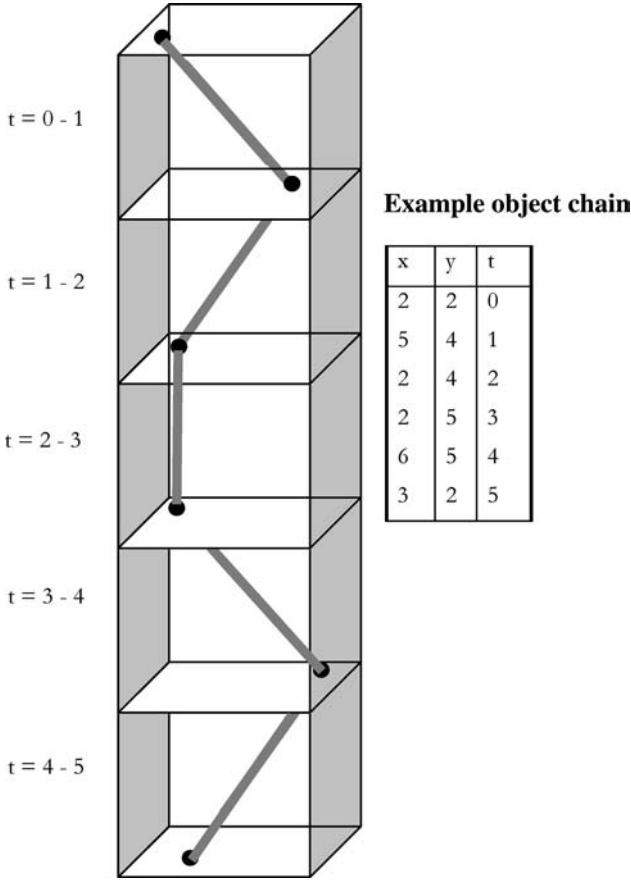
$$\mathbf{L}_i = (\mathbf{u}_{i0}, \mathbf{u}_{i1}, \dots, \mathbf{u}_{iT}) \quad (2)$$

This defines individual  $i$  being at location  $\mathbf{u}_{i0}$  at the beginning of the simulation (time 0), and moving to location  $\mathbf{u}_{i1}$  at time  $t = 1$ . At the end of the simulation ( $t = T$ ) individual  $i$  may be found at  $\mathbf{u}_{iT}$ . Object chains can have time-dependent attributes associated with them such as weight, height, disease status, size, etc. While observations on chains occur at a finite number of time points or observation times, these observations do not have to happen at the same time for all individuals under scrutiny.

## 2.3 Movement models

*Movement Models* define how the modeled entity gets from a geographic location at time  $t$  to another location at time  $t + 1$ . One of the simpler movement models is linear motion at a constant speed (Fig. 1). But movement models could also be non-linear, corresponding for example to the physics of cannon ball trajectories. Movement models can also be constructed that are constrained to adhere to networks such as rivers, roads and corridors. Logical operators could also be defined to describe changes in trajectory upon collisions with other objects and features. The objective of course is to define movement models that are descriptive of the behaviors found in the physical or biological system under scrutiny.

For the influenza simulation we explored two kinds of movement models. The first involves instantaneous displacement from the spatial coordinates for entity  $i$  at time  $t$  ( $\mathbf{u}_{it}$ ) to those at time  $t + 1$  ( $\mathbf{u}_{i(t+1)}$ ). We defined this instantaneous displacement as occurring at time  $t + 1$ . We viewed this as an



**Fig. 1.** Object chain and the linear movement model. Locations of nurses and patients in space-time are represented as object chains (Eq. 2) illustrated here by the space-time coordinates listed in the “Example Object Chain”. The Figure uses cubes to represent intervals between observation points that occur at  $t = 0, 1, 2, 3, 4, 5$ , and represents the object chain as a stack of such cubes. Movement between observation points is modeled using linear motion at a constant velocity between locations of an object at two time points

observational model in which the entity is assumed to reside at its known location up until that moment when it is observed elsewhere. We also explored network-based movement models that assumed constant velocity along the networks defined by the hospital corridors. Notice that both of these approaches preclude unrealistic results such as nurses moving through walls.

For the instantaneous displacement model one can summarize the nearest neighbor relationships among  $N$  modeled entities at some time  $t$  as:

$$\mathbf{n}_{k,t} = \begin{bmatrix} 0 & n_{1,2,k,t} & \cdot & \cdot & n_{1,N,k,t} \\ n_{2,1,k,t} & 0 & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & n_{N-1,N,k,t} \\ n_{N,1,k,t} & \cdot & \cdot & n_{N,N-1,k,t} & 0 \end{bmatrix} \quad (3)$$

Here the  $n_{i,j,k,t}$  are binary variables quantifying the  $k$ -nearest neighbor relationships among entities  $i$  and  $j$  at time  $t$  such that:

$$n_{i,j,k,t} = \begin{cases} 1 & \text{if and only if } j \text{ is a } k \text{ nearest neighbor of } i \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

By convention we define  $n_{i,i,k,t} = 0$  since we do not wish to count individuals as nearest neighbors of themselves. The matrix (Eq. 3) enumerates the  $k$  nearest neighbors for each of the  $N$  individuals in existence at time  $t$ . The entries of this matrix are 1 (indicating  $i$  and  $j$  are  $k$  nearest neighbors at time  $t$ ) or 0 (indicating they are not  $k$  nearest neighbors at time  $t$ ). The matrix is asymmetric about the 0 diagonal since nearest neighbor relationships are not necessarily reflexive. Since two individuals cannot occupy the same location, we assume at any time  $t$  that any individual  $i$  has  $k$  unique  $k$ -nearest neighbors. The row sums of the matrix thus equal  $k$  ( $n_{\bullet,2,k,t} = k$ ), although the column sums vary depending on the spatial distribution of the individuals at time  $t$ . The sum of all the elements in the matrix is  $Nk$ . There exists a  $1 \times T + 1$  vector of times denoting those instants in time when either (1) the system is observed and the locations of the entities are recorded, or (2) under the instantaneous movement model at least one entity changes geographic location. We can then consider the sequence of  $T$  nearest neighbor matrices defined by

$$\mathbf{n}_k^T = \{n_{k,t}; \quad t = 0..T\} \quad (5)$$

This defines the sequence of  $k$  nearest neighbor matrices for each unique temporal observation recorded in the data set, and thus quantifies how nearest neighbor relationships change through time. This demonstrates one way in which space-time queries (here the nearest neighbor relationship) can be used to quantify and construct spatial weight matrices through time. In the influenza example we used nearest neighbor relationship to identify exposure opportunities in which infection transmission events might occur between infectious and susceptible individuals.

## 2.4 Attribute change models

*Attribute change models* describe how the values of attributes change between observations times. The simplest attribute change model is a step function that updates an attribute's value when a new observation is made and that attribute value has changed. Continuous attribute change functions can also be defined that interpolate the values through time. These include simple linear interpolation between time points that ignore spatial and temporal dependencies. More complex interpolation functions that account for spatial and temporal covariance can be used to interpolate values through space and time. These include techniques from the field of geostatistics that provide a probabilistic framework for space-time interpolation by building on the joint spatial and temporal dependencies between observations (Kyriakidis and Journel 1999). For STIS-influenza we use a step function to model infection status whose states were comprised of susceptible, infected and infectious, infected but not infectious, and recovered and immune.

## 2.5 Influenza simulation

The *Influenza simulation* provided a realistic data set for which we knew the actual chain of infection as well as the mechanics of the infection transmission model. Several approaches are available for simulating infection transmission systems, including deterministic and stochastic approaches of compartmental analysis (see Jacquez 1999) and dispersion on networks (Bian 2004). We employed an individual-based model in which exposure opportunities arise between mobile susceptible and infectious individuals. Once susceptible and infectious individuals are sufficiently close to one another (within 1 m), infection occurs with a fixed probability per unit time. Transitions between disease states also occur with a fixed probability per unit time, with these probabilities selected to support epidemic spread within the allotted 15-day simulated study period. The ward contains forty-one mostly stationary patients and three nurses who make periodic rounds. Because they are mobile and visit patients, as well as one another, nurses mediate the bulk of infection events, although some patient to patient transmission occurs. In this simulation, the patients only leave their beds when they are discharged. The data set consists of over 30,000 records recording the location, time and health status of each of the ward's 3 nurses and 41 patients over 15 days. One infectious patient starts the outbreak, and it spreads within her ward, and then is spread to other wards by nurses. The simulation models 15 days, and monitors the ward at 1 second intervals. The dataset records observations whenever an object changes its space-time trajectory or attributes (disease state). We emphasize the purpose of this simulation is solely for evaluating the use of STIS in disease surveillance. Once the simulation was completed, we loaded the resulting data set describing the locations and disease status of the nurses and patients into the STIS-influenza prototype.

## 3 Results

Our research had two goals; the first goal was to develop a library for programming STIS software. The second was to apply this library to create a software application for modeling an outbreak of influenza in a simulated hospital ward. To accomplish these we created a programming library, STroodle, and then used it to develop the STIS-Influenza disease surveillance software.

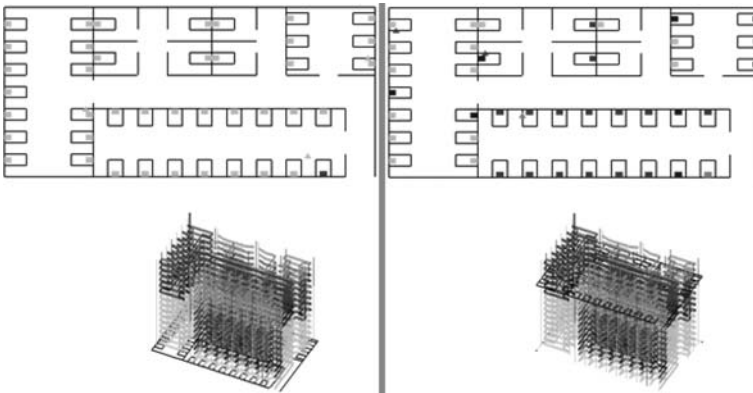
STroodle was written in C++ and implements the space-time object model; a space-time data structure based on a 3-dimensional R-tree indexed by a Hilbert space-filling curve; and a set of spatial, temporal, spatio-temporal and epidemiological queries directly relevant to disease surveillance.

STIS-Influenza gained access to STroodle's functionality through an Active-X control. The Active-X control used a single function call to access the data structures and space-time queries through STroodle's Application Programmers Interface (API). This allowed us to evaluate the feasibility of using STroodle to develop a broader range of custom software applications using a modular software design. In this approach the custom software application is in charge of application-specific features such as the graphical

user interface and higher-level functionality that is specific to different knowledge domains, while STroodle provides basic functionality such as the data model, data base, and space time queries.

The user interface for STIS-influenza visualized the dataset in 2-D (two spatial dimensions at one point in time) and 3-D (two spatial dimensions plus time). Both 2-D and 3-D views are fully animated so that the changing infection status and locations of the patients and nurses are displayed as they change through time. An animation controller with stop, step, play, fast forward and fast backward buttons controls the animation. The 3-D view represents nurse and patient object chains as lines in space-time, allowing visual analysis of disease spread and the movement of patients and nurses. The 2-D display is the projection of the intersection of the space-time paths at the time specified by the animation controller with the plane defined by the floor plan. Different symbols are used to display nurses and patients, with the location of individuals on the ward determined by the time shown on the animation controller. Each nurse visits a series of patients, and their paths are shown as the spiraling lines in the 3-D view in Fig 2.

At the beginning of the simulation, one individual is infected with the flu. By the end of the fifteen days, after infectious nurses have made their rounds, only six patients remain susceptible, the rest are infectious, infected or immune. The left side of Fig. 2 shows the ward at the beginning of the outbreak, when only one patient has the flu. The uninfected nurses, shown by gray triangles, are making their rounds. The right side of the figure shows the ward at the beginning of the spread to other rooms. Two nurses are infectious, while the nurse for ward A is now immune. At the later time, the floor plan of the ward has moved up through the data set on the right side of the figure to that point in time specified by the animation controller.



**Fig. 2.** Visualization of object chains. The left and right panels in the figure display 2-D (*top*) and 3-D (*bottom*) views at different stages of the outbreak. Patients are shown as squares and nurses as triangles. Patients and nurses have four possible states—susceptible, infected, infectious (all shown in *gray*), and immune (*black*). The 2-D view shows the hospital ward comprised of nine rooms visited by three nurses three times per day. The 3-D view represents individuals as object chains

## 4 Queries

While the 2-D and 3-D views provide a rapid visual assessment of disease spread, they do not quantify important determinants of infection transmission such as the amount of time susceptible individuals were in close proximity to infectious individuals. Questions of this kind may be addressed using the space-time queries in the STroodle library (Table 1). Spatial queries include most of those found in traditional GIS, including distance between objects, and nearest neighbor and adjacency measures. Time queries evaluate relationships such as temporal order, overlap of adjacent time periods, and containment of one interval within another. For instance, you can view all the objects that exist at a particular time (Time\_Slice) or find objects that exist within or overlap a particular time interval (Interval\_Containment, Interval\_Overlap respectively). The overlap in time of two objects can be found (Lifespan Intersection) as can the union of their life spans (Lifespan\_Union). For the influenza simulation the term “Lifespan” denotes the set of time intervals in which a given individual was on the hospital ward. The lifespan of an individual is returned by the Object\_Lifespan query. The disease status of an individual at a particular time or over an interval is returned by the Attribute\_at\_Time and Object\_Attributes queries.

Space time queries return both spatial and temporal information. For instance, Polygon\_Containment\_Intervals returns the time intervals when an object is contained within a polygon. Collide returns the times when two

**Table 1.** Queries in STroodle/STIS-Influenza. Refer to text

Spatial queries	Temporal queries	Epidemiological queries
Length	Lifespan	Event/diagnosis queries
Area	View time slice	define event
Position	Interval containment	find first event
Point path	Interval overlap	find events in interval
Nearest neighbor	Lifespan union	nearest events
<b>Adjacency</b>	Lifespan intersection	to focus event
Border polygons	Attributes at time	nearest event to certain time
Border length	<b>Spatio-temporal object queries</b>	waiting time between pairs of events
Distance	Object chain and subobject	waiting time and distances between pairs of events
Overlap	Threshold distance	between all events
Polygon containment	Minimum distance	Exposure time
<b>Network queries</b>	Polygon containment intervals	Find exposed
Distance	Distance over time	Epidemic curve
Adjacency	Collisions	
Nearest neighbor	Close calls	
Neighbors by node	Distance between objects over time	

specified objects collide, while `Inside_Threshold_Distance` returns the times when two specified objects approach within a threshold distance. `All_Collisions` tracks all collisions that occur in the modeled system, while `Brushes_with_Object` finds all objects that come within a threshold distance of a target object.

The above spatial, temporal and space-time queries have applications in knowledge domains other than disease surveillance. Using these as building blocks, we constructed epidemiological queries to quantify events known to be associated with infection transmission. Time lags can be quantified with a number of methods including the lag between two infection events (`Waiting_Time`) and lags between sequential infections (`Adjacent_Waiting_Times`). A list of waiting times and the spatial distances between temporally adjacent infection events is returned by `Adjacent_Waiting_Times`. `Adjacent_Waiting_Time_Distance_Pairs` returns the distances between infections adjacent in time, and `All_Waiting_Time_Distance_Pairs` tracks the waiting times and distances between all possible pairs of infection events. These are particularly useful for evaluating space-time dependencies that may arise under contagious processes, including the space-time regression envisioned by Mantel (1967) and other space-time interaction tests (e.g. Knox 1964, Jacquez 1996).

We also constructed queries to quantify which susceptible individuals were exposed to infectious individuals, and the duration of the exposure. `Find_Exposed_Objects` returns all individuals who were within a specified distance of an infectious individual for a minimum amount of time. `Exposure_Time` finds the amount of time an individual was exposed to infectious individuals during a specified time interval. Finally, `Epidemic_Curve` renders a graph of the number of susceptible, infectious, infected and immune individuals through time.

As a test case `STIS-Influenza` provided detailed temporal resolution that exceeds that expected under most scenarios. `STIS-Influenza` provides a “movie” that displays the locations and disease states of the ward’s nurses and patients at up to 1-second resolution over the 15-day course of the influenza outbreak. This is accomplished in both 2-D (spatial) and 3-D (space-time) views. Each “tic” of the clock requires a space-time query on each and every individual’s (patient or nurse) spatial location and disease state, and updating of the 2-D and 3-D views. The data set is based on a probabilistic simulation of influenza and records observations whenever an object changes its space-time trajectory or disease state. Hence each “tic” of the movie requires, for every patient and nurse, space-time interpolation to estimate geographic locations between the observations that bracket the desired time point. Hence the movie provides a reasonable test of the efficiency of the space-time indexing and queries. The space-time indexing approach prunes the search space at a point in time to a plane; while temporal queries over a spatial domain prune the search space to a small space-time volume. Hence the relationship between access time and dataset complexity is very favorable such that query access time is  $N \log N$  where  $N$  is the number of space-time objects. We found the practical limit on our investigations to be the available disk space for storing the data set rather than query access times.

## 5 Discussion

The term “Technological Determinism” captures the notion that, to a certain extent, scientific advances and the questions that may be addressed depend on the state of current technology. For example, the recognition that infections are caused by pathogens such as bacteria and viruses arose only after the development of the light microscope. While the limitations of the static map view of GIS are well recognized (e.g. Goodchild 2000), the extent to which the static view limits applications in disease surveillance is not entirely clear. This research has explored how the surveillance of infectious diseases might benefit from relaxation of the static map view. This is a question of some relevance, since disease surveillance systems constructed on traditional GIS are being deployed to monitor the health of populations and to provide rapid response in the face of bioterrorist threats.

To this end our significant findings were that the performance of the space-time data structures was acceptably fast, capable of retrieving spatial, temporal, and attribute data at speeds that support rapid 2-D and 3-D animation. We demonstrated that queries based on this space-time data structure can be modified at application development time to construct custom queries for epidemiology and exposure assessment, potentially opening new avenues to handling, visualizing and interrogating data. The ability to handle the movement of individuals within disease surveillance software is expected to lead to new approaches to exposure assessment and to the exploration of health-environment relationships, as demonstrated by recent results of Meliker et al. (2005).

Future directions include enhancing object movement models, attribute change models, and developing object-in-field techniques. In STIS-influenza movement models are incorporated into the space-time object model as a *movement vector* defining the distance and direction the object moved between movement stop and start times. Future applications should instead use a *movement function* for incorporating non-linear motion. This will include the ability to represent movement on networks such as rivers and roads.

Our data model assumes observations occur at discrete times at which the attributes of an object are quantified. Attribute change models model how the values of attributes change between observation times. In this research, we employed step functions that update an attribute’s value when a new observation is made on that attribute. Future applications should develop attribute change models based on space-time interpolation methods. These can build on an existing theory in the area of space-time interpolation, as described below after the discussion of object-in-field techniques.

Space-Time Information Systems appear capable of capturing the dynamics of a complex world. The selection of data, construction of data models, and the specification of methods depend critically on how this real world is represented. Both object-based and field-based representations (Mark 1993; Smith and Mark 1998; Goodchild 1992; Cova and Goodchild 2002) are relevant to STIS. *Objects* represent real-world entities that are coherent and indivisible, such as the patients and nurses. *Fields* represent measurements on variables whose values vary continuously through geographic space, such as electromagnetic fields. In this research, we implemented and tested a space-time object model. Future applications

should integrate this space-time object model with a space-time field representation, integrating attribute change models and space-time interpolation techniques to obtain a space-time field model capable of returning attribute values for any pixel at any point in time. Coupled with the space-time object model, this is expected to make possible *object-in-field techniques* for assessing an individual's exposure to environmental factors whose values vary through space and time.

## References

- ATSDR (1992) Public health assessment guidance manual. Chelsea, Michigan, Lewis Publishers
- Barnes S, Peck A (1994) Mapping the future of health care: GIS applications in health care analysis. *Geographic Information Systems* 4:31–33
- Bellander T et al. (2001) Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environmental Health Perspectives* 109:633–639
- Ben-Shlomo Y (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology* 31:285–293
- Bian L (2004) A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B* 31(3):381–395
- Cheng T, Molenaar M (1998) *A process-oriented spatial-temporal data model to support physical environmental modeling*. 8th International Symposium on Spatial Data Handling, Vancouver, British Columbia, International Geographical Union, Geographic Information Science Study Group
- Clarke KC, McLafferty SL et al. (1996) On epidemiology and geographic information systems: A review and discussion of future directions. *Emerging Infectious Diseases* 2:85–92
- Cova TJ, Goodchild MF (2002) Extending geographical representation to include fields of spatial objects. *International Journal of Geographical Information Science* 16:509–532
- Croner C, Sperling J et al. (1996) Geographic Information Systems (GIS): New perspectives in understanding human health and environmental relationships. *Statistics in Medicine* 15:1961–1977
- Elgethun K et al. (2003) Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environmental Health Perspectives* 111:115–122
- Glass GE, Schwartz BS et al. (1995) Environmental risk factors for Lyme disease identified with geographic information systems. *American Journal of Public Health* 85:944–948
- Goodchild M (1992) Geographical Data Modeling. *Computers and Geosciences* 18:401–408
- Goodchild M (2000) GIS and Transportation: Status and Challenges. *GeoInformatica* 4:127–139
- Gould P (1997) *Spreading HIV across America with an air passenger operator*. International Symposium on Computer Mapping in Epidemiology and Environmental Health, Tampa, Florida, April 1995, World Computer Graphics Foundation
- Greiling DA, Jacquez GM, Kaufmann AM, Rommel RG (2005) Space time visualization and analysis in the Cancer Atlas Viewer. *Journal of Geographical Systems* (Accepted)
- Hagerstrand T (1970) What about people in regional science? *Papers of the Regional Science Association* 24:7–21
- Hazelton NWJ (1991) Integrating time, dynamic modeling and GIS: Development of four-dimensional GIS. Ph.D. Thesis, Department of surveying and land information, The University of Melbourne
- Hjalmar U, Kulldorff M et al. (1996) Childhood leukemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine* 15:707–715
- Holm DM, Maslia ML et al. (1995) Geographic information systems: A critical resource in exposure assessment. SUPERFUND XVI Conference and Exhibition, Washington, DC

- Hornsby K, Egenhofer M (2002) Modeling moving objects over multiple granularities, Special issue on Spatial and Temporal Granularity. *Annals of Mathematics and Artificial Intelligence* 36:177–194
- Huisman O, Forer P (1998) *Computational agents and urban life spaces*: A preliminary realization of the time-geography of student lifestyles. Proceedings of the 3<sup>rd</sup> International Conference on GeoComputation. University of Bristol, UK, Sept. 1998
- Jacquez GM (1996) A k-nearest neighbor test for space-time interaction. *Statistics in Medicine* 15:1935–1949
- Jacquez GM (2000) Spatial Epidemiology: Nascent Science or a Failure of GIS?. *Journal of Geographical Systems* 2(1):91–97
- Jacquez GM (1998) *GIS as an Enabling Technology*. GIS and Health. T. G. a. M. Loytonen. London, Taylor and Francis 17–28
- Jacquez GM, Waller LA et al. (1996) The analysis of disease clusters Part I: State of the art. *Infection Control and Hospital Epidemiology* 17:319–327
- Jacquez JA (1999) Modeling with Compartments. BioMedware Press, Ann Arbor, MI
- Kitron U, Pener H et al. (1994) Geographic information system in malaria surveillance: mosquito breeding and imported cases in Israel, 1992. *American Journal of Tropical Medicine and Hygiene* 50:550–556
- Kleinerman RA, Kaune WT et al. (2000) Are children living near high-voltage power lines at increased risk of acute lymphoblastic leukemia? *American Journal of Epidemiology* 151(5):512–515
- Knox G (1964) The detection of space-time interactions. *Applied Statistics* 13:25–29
- Kulldorff M, Feuer EJ et al. (1997) Breast cancer clusters in northeastern United States: A geographical analysis. *American Journal of Epidemiology* 146:161–170
- Kwan MP et al (2003) Accessibility in space and time: A theme in spatially integrated social science. *Journal of Geographical Systems* 5:1–3
- Kyriakidis PC, Journel AG (1999) Geostatistical space-time models: a review. *Mathematical Geology* 31:651–684
- Langran G (1992) *Time in Geographic Information Systems*. London, Taylor and Francis
- Lawson AB, Waller LA (1996) A review of point pattern methods for spatial modelling of events around sources of pollution. *Environmetrics* 7:471–487
- Loytonen M (1998) GIS, Time Geography and Health. In: Gatrell T, Loytonen M (eds) GIS and Health. Taylor and Francis, London
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27:209–220
- Marbury M (1996) GIS: New tool or new toy? *Health and Environment Digest* 9:88–89
- Mark DM, Egenhofer MJ, Bian L, Hornsby KE, Rogerson PA, Vena J (2000) Spatiotemporal GIS analysis for environmental health: solutions using geospatial lifelines. Proceedings of the 2nd International Workshop on Geomedical Systems, Geomed '99
- Mark DM (1993) Toward a theoretical framework of geographic entity types. In: Frank AU, Campari I (eds) Spatial Information Theory: A Theoretical Basis for GIS. Springer-Verlag, Berlin, pp. 270–283
- Meliker J, Slotnick M, AvRuskin GA, Kaufmann A, Jacquez GM, Nriagu JO (2005) Improving exposure assessment in environmental epidemiology: applications of a Space-Time Information System. *Journal of Geographical Systems* (Accepted)
- Miller H (2004) A measurement theory for time geography. Under review, *Geographical Analysis*
- Miller H (2004) What about people in geographic information science?. In: Peter Fisher, David Unwin (ed) Re-Presenting Geographic Information Systems. John Wiley, in press
- Moore AB et al (2003) Proceedings of the 7th International Conference on GeoComputation University of Southampton, United Kingdom 8 – 10 September 2003
- Peuquet DJ (1994) It's about time: A conceptual framework for the representation of temporal dynamics in GIS. *Annals of the Association of American Geographers* 84(3):441–461
- Peuquet DJ, Duan N (1995) An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Systems* 9(1):7–24

- Peuquet DJ, Qian L (1996) *An integrated database design for temporal GIS*. 7th Spatial Data Handling Conference, Delft, The Netherlands
- Raper J, Livingstone D (1995) Development of a geomorphological spatial model using object-oriented design. *International Journal of Geographical Information Systems* 9(4):359–383
- Richards FO (1993) Uses of geographic information systems in control programs for onchocerciasis in Guatemala. *Bulletin of the Pan American Health Organization* 27:52–55
- Rushton G, Lolonis P (1996) Exploratory spatial analysis of birth defects in an urban population. *Statistics in Medicine* 15:717–726
- Schaerstrom A (1996) *Pathogenic Paths? A time geographical approach in medical geography*. Lund University Press, Lund, Sweden
- Smith B, Mark DM (1998) Ontology and geographic kinds. International Symposium on Spatial Data Handling. Vancouver, Canada
- Teutsch SM, Churchill RE (1994) *Principles and practice of public health surveillance*. Oxford University Press, Oxford
- Waller LA, McMaster RB (1997) Geographic Information Systems and public health surveillance. *Geographical Systems* 4:327–342
- Ward MH et al (2000) Identifying populations potentially exposed to agricultural pesticides using remote sensing and a geographic information system. *Environmental Health Perspectives* 108:5–12
- Wartenberg D (1992) Screening for lead exposure using a geographic information system. *Environmental Research* 59:310–317
- Wartenberg D, Greenberg M et al. (1993) Identification and characterization of populations living near high-voltage transmission lines: A pilot study. *Environmental Health Perspectives* 101:626–632
- Xia H, Carlin BP et al. (1997) Hierarchical models for mapping Ohio lung cancer rates. *Environmetrics*
- Yearsley CM, Worboys M (1995) *A deductive model of planar spatio-temporal objects*. In: Fisher P (ed) *Innovations in GIS 2*. Taylor and Francis, London pp. 43–51
- Yuan M (1995) Wildfire conceptual modeling for building GIS space-time models. In: Frank AU, Kuhn W, Haunold P (eds) *Temporal Data in Geographic Information Systems* pp. 47–56