Special Handling of Spatio–Temporal Data

Jonathan W. Lowe

Are you ready for the unique computing challenges of spatio–temporal data management?

How many years ago did the Federal Communications Commission (FCC) require wireless carriers to provide a method for locating mobile 911 calls? It was a 1996 mandate with an October 2001, deadline. After the FCC's announcement, the deadline slipped, the telecommunications industry slumped, and all but a few location-based services startups slammed into bankruptcy. Nevertheless, the technology has quietly advanced.

Today, though late to the party, GPS-enabled phones are readily available in the United States. They're inexpensive, Java-ready, and capable of rapidly generating vast amounts of spatio–temporal data. This last point is key. The rapid flow, large volume, and multidimensional nature of data generated by location-aware devices sharply distinguish them from traditional static spatial data. Simply put, such spatio–temporal data requires special handling.

GPS phones and an array of market-specific, location-aware devices have paved the way for fleet-tracking, people-tracking, even pet tracking (see www.oxloc.com). Already, companies are positioned either to interpret live GPS-phone data in support of real-time decisions, or to data-mine user behavior over extended time ranges. This column describes the challenges that industry pioneers must overcome in their management of high-volume, real-time, spatio–temporal data.

Monitoring movement

Three such pioneers, @Road (www.road.com), Gearworks (www.gearworks.com) and Profilium (www.profilium.com), are quite familiar with the challenges of spatio–temporal data management. In fact, their business cases are all about helping others manage and interpret the data.

@Road and Gearworks track moving objects, such as delivery trucks, by pulling data from the custom tracking units or GPS-enabled cell phones of their drivers. By collecting the latitude, longitude, and timestamp of each object's changing location, these two fleet-tracking companies can supply drivers with real-time dispatch information (see Figures 1a and 1b), confirm vehicle arrivals at scheduled stops, and issue summary reports, such as the fleet's ability to deliver their goods on time. The aim of each service is to improve return on investment, mainly through increased efficiency and more timely communication.

Profilium data-mines months of user movement patterns for advertisers who then extend offers to the people most likely to be near the right store at the right time. Their assumption is that, when offered a bargain, people are more likely to buy if the store is nearby. For instance, someone who usually eats lunch across the street from a Wal-Mart (www.walmartstores.com) store on Wednesdays is a good prospect for an electronic (such as a short messaging service) Wal-Mart coupon sent to his phone at lunchtime next Wednesday.

Loading high-volume data

The previous business cases are easy enough to visualize. It's managing the underlying data that raises complex computing challenges. One challenge is the sheer volume of the data. In @Road and Gearworks' cases, even a small pool of 50 GPS phones, or similar location-aware input devices, can generate many records in a relatively short time. Let's do the math. If a single GPS phone updates its position once each minute, that's 1,440 records per day per phone (60 per hour, 24 hours per day). That's just one phone; the 50-vehicle fleet sends in 72,000 total records per day (1,440 × 50). After two weeks of round-the-clock operation, this small fleet logs 1,008,000 spatio–temporal records (14 days at 72,000 records per day).

In practice, individual fleet-tracking companies' combined customer volumes can exceed 100,000 simultane-

Glossary

FCC: Federal Communications Commission

Lat/Long: Latitude and longitude

TIGER: Topologically Integrated Geographic Encoding and Referencing

Net Results columnist Jonathan W. Lowe covers the role of emerging technologies in the exchange of spatial information. Lowe is the owner of Local Knowledge Consulting (Berkeley, California), where he designs and implements spatial Web sites. He can be contacted at info@giswebsite.com.

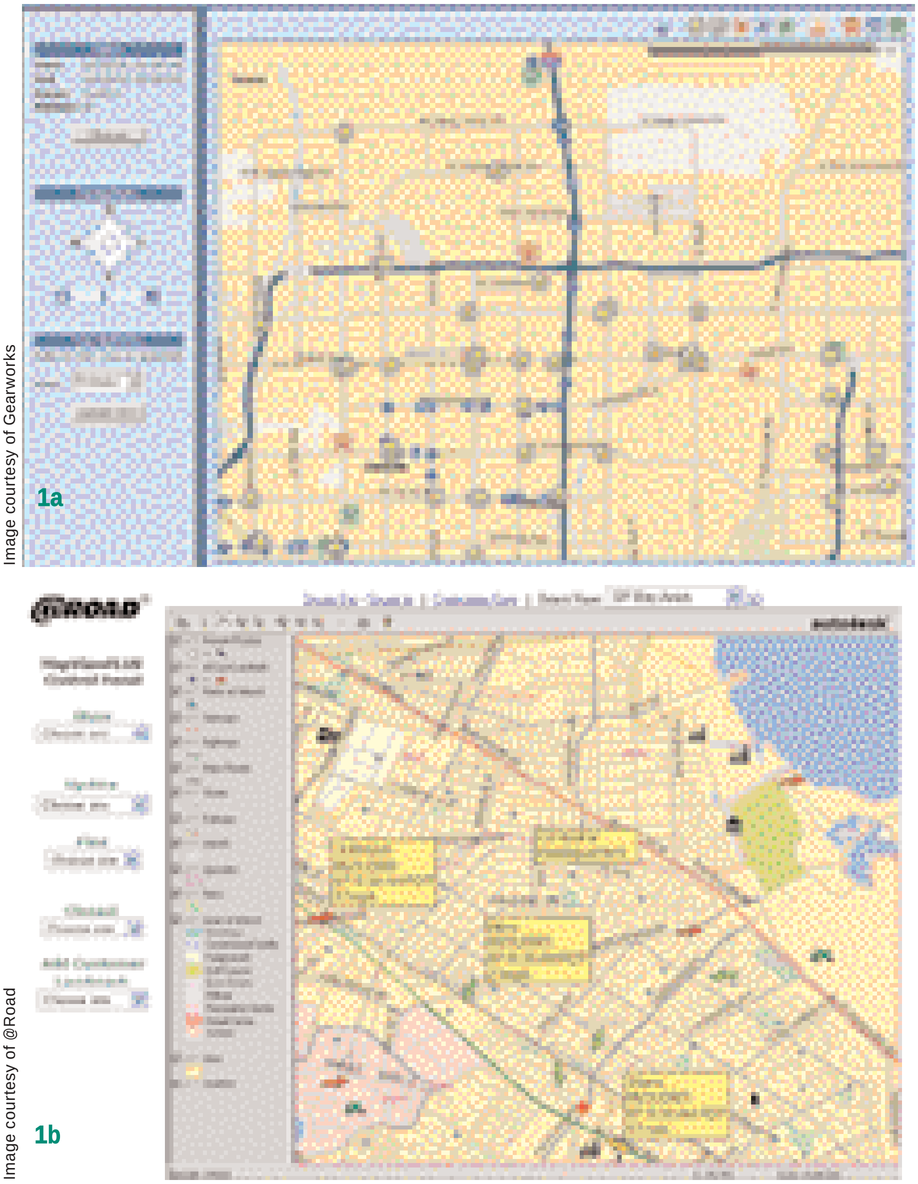
ously tracked vehicles. Even when operators lower the rate of positional updates per vehicle from every minute to every 15 minutes, a 100,000-vehicle fleet generates 9,600,000 records per day. (Wow! For perspective, the TIGER (Topologically Integrated Geographical Encoding and Referencing roads dataset for the entire United States contains approximately 53,000,000 records.

If the "fleet" is people rather than delivery vehicles, as in Profilium's case, the numbers change, but not necessarily for the better. Profilium doesn't need to know exactly where people are each moment — they just want the highlights. Their research indicates that people don't move around too much — most of the time they're stationary at home or in the office. Less ordinarily, we travel to eat, shop, play, or (far too rarely) go on vacation. Given this predictable behavior, Profilium expects most GPS phones to deliver 7–10 relevant positional records per day on average. That's far fewer updates than the fleet trackers generate, but with many more objects to track — telecommunications firms anticipate tracking millions of users per day.

High-volume datasets present potential load problems. The records must be consolidated, typically in databases. It takes time to load data into databases, however, and the loading routine must be fast enough to prevent the incoming mountain of records from piling up like cars waiting to get through a tollbooth. Individual database vendors have various products for high-performance loads. (A word of warning: if you're in the market for high-speed database loading tools, requesting a benchmark against your own sample data subset is one of the only reliable metrics. Results vary widely between datasets, machines, and load tools.) And loading the data is only the beginning — once captured, what does it actually mean? Furthermore, once the data make sense, how do the right people get the right alert messages in a timely fashion?

Image courtesy of Gearworks

Image courtesy of @Road



FIGURES 1a and 1b Gearworks (**1a**) and @Road (**1b**) use graphic browser-based maps to show real time locations of tracked vehicles.

Patterns and aggregation

Most spatio–temporal applications aggregate, summarize, or interpret data — turning an entire population or fleet's chaotic real-world motion into meaningful statistics or reports. To aggregate data intelligently, @Road, Gearworks, and Profilium all require a means for interpreting it. If they track vehicles that are delivering goods or services, their applications must be able to recognize whether a vehicle has arrived at its intended destination or is still in transit. For instance, if they already have points representing each known destination, how can they correctly match vehicles to destinations?

Initially, you might suggest, "Why not just match the known delivery point coordinates with vehicles' GPS coordinates?" In the real world, however, exact matches between GPS-phone points and known destination points are rare. Because of fluctuations in GPS accuracy, even a completely stationary vehicle's reported location may change from one GPS reading to the next. Also, a single destination point representing a large facility (such as one with multiple parking lots) is highly unlikely to intersect exactly with the GPS points of vehicles stopped there.

And even if a phone or vehicle's GPS unit is working properly, there's no guarantee that the wireless transmission of that data will successfully reach the central datastore. We've all had calls dropped before; the same thing happens with attempted transmissions of location data. Until a day when wireless carriers have perfect network coverage (don't hold your breath), the "dead zones" of wireless transmission will continue to cause gaps in vehicle-track histories.

To accurately associate vehicles with their stops despite this fuzziness of real-world data, fleet-tracking vendors have developed the ability to recognize clusters of point data hovering around their nearby known destinations. For instance, if a delivery truck is scheduled to stop at a Kmart (www.kmartcorp.com), then each of the truck's tracked points that fall within, say, a 30-meter radius of the Kmart are likely to represent the same arrival at that stop. Cluster recognition lets fleet trackers aggregate points representing the same stop at a single destination. This aggregation reduces the data volume to a more compact, but still meaningful subset. However, computing distances between points raises its own set of challenges.

Unprojected GPS data. Distance queries may be pretty trivial to compute for planimetric coordinates, but not for data in degrees of latitude and longitude (lat/long) — the native, unprojected format of GPS coordinates. Calculating accurate distances between two lat/long or geodetic points requires both a computationally expensive "transcendental" formula and a model of the Earth in it's true ellipsoidal shape rather than as a perfect sphere. (Most traditional desktop GIS programs approximate distances between unprojected data points by default, assuming that the Earth is exactly spherical.) Unless the spatio–temporal application can tolerate erroneous results or is geodetically enabled, vendors need to project the incoming GPS data. Unfortunately, time is precious in high-volume, real-time processing, and extra steps add time to loads and analysis. So, an ideal tool for managing GPS-generated spatio–temporal data would be capable of accurately analyzing native latitude and longitude coordinates without projecting them first.

Image courtesy of Gearworks

FIGURE 2 Gearworks transmits job details from the dispatching enterprise to individual mobile worker GPS phones.

Performant indexes. Again, with spatio–temporal data, fast analysis is as important as fast data loading. Whether the GPS points are projected or not, how will search routines avoid scanning multimillions of records each time a user queries for nearby points? Of course, the answer is to use indexes — presorted paths to similar records that return results quickly even when searching large record sets. Commercial geospatial tools (desktop applications and databases) almost all enable their users to index geometry, text, numbers, and dates. Let's take a conceptual look at how applications use temporal indexes to see how indispensable they are to performance, and to investigate a missing capability of most commercial software.

Temporal indexing. The timestamps with each GPS point provide a pivotal key when dividing and conquering big datasets. Over time, when collecting spatio–temporal data, the same region accumulates more and more points as trucks drive back and forth, day in, day out. Searching for trucks in a busy region, even if the points are spatially indexed, will become slower and slower as the dataset grows. However, each day of traffic is, on average, much like the next — in other words, time neatly divides large datasets into more manageable, balanced segments of roughly equal record-counts. Using a temporal index, the search engine can eliminate most of the millions of irrelevant records immediately when searching for vehicles in a particular time period, leaving a small subset for normal spatial scanning. Because the subsets are approximately equally sized, the response times for a query against any day will be just as fast as against any other day. Predictable performance is good.

Multidimensional indexing. So, temporal indexes are critical to spatio–temporal applications, but not all database indexing capabilities are created equal. Unfortunately, like accurate geodetic data handling, efficient spatio–temporal indexing is absent from most commercial products. Yes, index support for spatial, temporal, and standard data types is widely available, but users must index each type separately. Profilium discovered this early in their performance benchmarking tests. When they searched for a list of people who were near Joe's Pizza Parlor on Friday night, the search engine had to check both a spatial index (ignore points not near Joe's Pizza Parlor) and a separate temporal index (limit the search set to Friday night only). Possibly due to the novelty of spatio–temporal data in today's market, few products yet leverage the fact that space and time are related dimensions and can be indexed together rather than separately. Few, but not all.

Region-trees. In 1984, Antonin "Toni" Guttman published his revolutionary discovery of the r-tree index, in which the *r* stands for *region* and the *tree* refers to the cascading, branching roots (like an organizational chart) that the search engine follows when seeking a subset of data. Among the problems the paper solved was the quandary of sorting spatial objects. Text can be alphabetized (just as indexes at the backs of textbooks are) and numbers have inherent order, but how can we sort space? The answer, which I will humbly not attempt to translate here, is in Guttman's paper (see www.informatik.uni-trier.de/~ley/db/conf/sigmod/Guttman84.html), which spatio–temporal researchers still cite today, almost 20 years after its publication.

One reason r-trees are still hot topics is because they can handle not just two dimensions, such as the *x* and *y* coordinates of a bounding box, but a third dimension, such as altitude, and a fourth dimension, such as an instant or a time range. And all of these dimensions fit in a single index. In fact, r-trees maintain fast performance with more than 20 different dimensions!

It is therefore possible to create a single r-tree index of a geometry's location (*x*, y, *z*) and time range. I have to admit that this sank into my feeble brain rather slowly. Visualizing a multidimensional index as a three-, four-, or five-dimensional "cube" isn't something I do every day. But whether or not it bends your brain, the relevance to processing speed is that with a multidimensional r-tree, those queries for Joe's Pizza Parlor can scan a single index (rather than two separate ones) to narrow the search for place and time. It's hard to beat the speed of a single index scan.

Informix Geodetic. Guttman worked at Informix Software in the late 1990s before the IBM (www.ibm.com) acquisition. There, he collaborated with experts in surveying and spatial databases to combine multidimensional r-tree support with a tool called the Informix Geodetic Datablade, which analyzes lat/long coordinates in their native format exclusive of expensive transcendental functions, calculates distances in reference to Earth as an ellipsoid, and queries against a multidimensional spatio–temporal r-tree index. (Exhale!) The result was a custom-fitted tool for storage and high-performance analysis of GPS-generated data. The Oracle (www.oracle.com) 10g release also features r-tree and geodetic functionality, though it's not clear whether the two work together or with a built-in ellipsoid model.

Spatial permeates IT

By balancing database-centric and middleware-centric approaches, @Road, Gearworks, and Profilium have each built custom systems to deliver fast performance as their client bases and spatio–temporal datasets scale upward. Their proven return on investment in this demanding computing niche is prompting their customers to request support not just for fleet-tracking, but for overall enterprise workflow (a confirmation of Net Results review of the expansion of spatial business logic into enterprise IT; see *Geospatial Solutions*, October 2003, p. 45).

This trend surfaces not just technically, but in overall business strategy. For instance, @Road has rebranded their services from "Fleet Tracking" to "Mobile Resource Management," reflecting growth from specific spatial services to larger (and not necessarily spatial) enterprise integration. For instance, one of @Road's customers, O'Brien Concrete Plumbing of Phoenix, Arizona, was initially a standard fleet-tracking customer using browser-based graphic mapping tools to manage its fleet in real time. Now, though, O'Brien Concrete also links its field operations activities to its enterprise-level payroll databases using @Road's support for machine-to-machine communication. In real time, when a worker completes a job, @Road sends an alert (with confirmation of the actual number of hours worked, by activity and by worker) to the appropriate department's routing and payroll databases. On any given day, each worker might perform several services (charging different pay rates) at a single job site, which affects that day's payroll calculation. @Road's update to O'Brien Concrete's enterprise databases subsequently insures accurate and automatic payroll tallies.

Similarly, when Roto-Rooter (www.rotorooter.com) asked Gearworks to handle the dispatching of its fleet of plumbers, part of the solution involved integration with a job-based AS400 dispatch server. Dispatchers log incoming calls to a database in the AS400 and assign a worker to each call, in part by checking a Gearworks-generated graphic map of current vehicle locations. Job assignments then appear on workers' phone displays (see Figure 2). When the job is complete, Gearworks' system logs the statistics to the same AS400, which then also generates timesheet and payroll reports.

Both examples make sense given that a job site, the worker's time spent there, and the company's bottom line are all connected. In a fully integrated system, being able to track the first two (place and time) helps enterprises to accurately calculate the third.

Special delivery

If anything sweeping can be said about spatio–temporal datasets, it's "handle with care!" To my knowledge, there is no single commercial off-the-shelf product that solves all the challenges of managing spatio–temporal data. As with most real-time applications, the complexity of fleet-tracking or GPS-phone data-mining is daunting. Vendors must meet the challenges of capturing spatio–temporal data in real time, transmitting it to a central storage and processing point, making sense of it despite gaps and variations in carrier protocols, and then appropriately alerting end users with graphic maps, summary reports, or machine-to-machine updates. The few successful spatio–temporal pioneers that have already won customers are now extending their scalable architectures into previously non-spatial enterprise applications as well. Given the increasing availability of location-aware devices and the special handling spatio–temporal data requires, these innovative companies are likely to garner even more business as time goes by. 🌐

Origin–Destination Surveys and

Retail Market Analysis

François Des Rosiers and Marius Thériault

When choosing a business location or expanding their premises, a major challenge retailers face is correctly delineating trade areas and adequately assessing both the economic potential of their customers and the market penetration of competitors. As emphasized by Grant Ian Thrall and Michelle Fandre (2003), businesses need a clear picture of the socio–demographic profile of the population within trade areas in order to better identify demand characteristics. Lifestyle segmentation profiles (LSPs) provide a means of obtaining and analyzing these socio–demographic characteristics, and are among several methods for delineating trade areas (Thrall, 2002) — point maps, spatial statistical procedures and isochronous drive curves, are all currently used by analysts for this purpose.

To optimally define trade areas, however, retailers must be able to clearly identify each customer's residential location for deriving family, age, income status, and overall consumption potential. But by and large, detailed, individualized information about customers is only available through on-the-spot surveys, which can be quite expensive and are not likely to be accessible to outsiders because of the confidential nature of the information and the competitive advantage it affords. Barcode technology for tracking product sales also provides a means of obtaining data to size up current clientele tastes and shopping patterns while adjusting inventories to effectively meet demand. Combined with ZIP codes and GIS technology, barcoding has induced a quiet revolution in business marketing. But again, retailers' internal barcode-derived data and analyses are generally well-kept, undisclosed information.

Usually employed for transportation planning purposes, origin–destination surveys can also serve as a tool for calculating retail trade areas and spatially analyzing competition.

In this column, we demonstrate that origin–destination (O–D) surveys, long employed in several Canadian metropolitan areas for transportation planning purposes, can also be useful at determining retail trade areas and assessing local economic potential and spatial competition among stores. O–D surveys are typically telephone questionnaires that provide local and regional mass transit authorities with hard data about the commuting patterns of individuals and households so as to better predict travel demand as well as transportation infrastructure needs. We applied the O–D approach to a series of shopping centers in Quebec City.

Located 150 miles east of Montreal, Quebec City had a population of roughly 490,000 people as of January 1, 2002, as opposed to approximately 750,000 inhabitants in the Quebec metropolitan area (QMA). Retail activities in the region are particularly well developed, with shopping center complexes ranking among the largest in Eastern Canada. According to Quebec City's (formerly the Quebec Urban Community's) latest assessment roll, in 1997 Quebec City alone contained 82 shopping centers — 37 neighborhood, 40 community, three regional, and two super-regional — occupying roughly 28.8 million square feet and totaling 2,485 shops. The shopping center categories conform to the International Council of Shopping Centers' guidelines, which classifies centers based on the overall acreage, square footage, and relative importance of anchor tenants.

Glossary

CAI: Center Attraction Index

EPI: Economic Potential Index

GLA: Gross leasable area

LSP: Lifestyle segmentation profile

MTQ: Ministry of Transport of Quebec

O–D: Origin–destination

QMA: Quebec metropolitan area

RTC: Quebec City Transit Authority

SCI: Spatial Competition Index

François Des Rosiers is an economist and a professor of Urban and Real Estate Management at the Faculty of Business Administration, Laval University, Quebec City, Canada. Marius Thériault is a quantitative geographer specializing in GIS technologies and a professor in the Land Planning Department of the same university. Both belong to the Centre de Recherche en Aménagement et Développement, for which Thériault is the director.
Shop Talk presents in-depth analysis of business geographics topics.

For our analysis, we focused on two super-regional, three regional, and five community shopping centers. Although the latter category accounts for only 20.4 percent of all stores located in Quebec City's community centers, we considered all regional and super-regional shopping centers in the study and accounted for all major shopping destinations in the QMA. The overall sample included 1,220 shops, totaling 4.8 million square feet of gross leasable area (GLA).

Car ownership is well developed in Quebec City . . . the QMA is among the most road overequipped cities in North America (with a highway network totaling 21.7 kilometers per 100,000 inhabitants) . . . [and] urban sprawl is prevailing.

Trip sampling

The O–D survey we used for this study consists of a phone survey conducted jointly by the Ministry of Transport of Quebec (MTQ) and the Quebec City Transit Authority (RTC) from September 18 to December 17, 2001. The survey is based on a random sample of roughly 8 percent of the QMA population, with the survey area divided into 63 sampling zones. MTQ and RTC assigned each zone a specific sampling rate — from 5 to 15 percent of the local population — depending on the number of households, population density, and mass transit use. Overall, the survey interviewed 68,121 people living in 27,839 households who generated 174,243 daily trips during a typical weekday (Monday–Friday). To provide a complete picture of total commuting patterns, MTQ and RTC also geocoded (on a 1:20,000 map showing street networks and six-digit postal codes) the origin and destination points of each trip reported by respondents. They geocoded each point based on street address or assessment role (location of building), depending on which method provided the most accurate spatial references and best geographical scale (either building or city-block level).

Based on the geocoded data, MTQ and RTC derived expansion factors for each person and household using 2001 Canadian census socio–demographic attributes. This allowed them to extrapolate the sample data to the entire QMA population, giving statistically significant estimates of the total daily trips from each census tract or enumeration area. Through the extrapolation, MTQ and RTC concluded that the entire QMA population makes approximately 1.8 million daily trips.

In addition to origin and destination data, the survey collected information about the transportation mode and purpose for each trip. In the QMA, trips using private cars largely predominate over other transportation means with 73.3 percent driving or riding in cars, 13.2 percent using various bus systems, 11.4 percent walking, and 1.7 percent traveling by other (bike, rollerblade, and so forth) means. In terms of purpose, specific travel goals were related to work (42.6 percent); reaching primary or high schools, colleges, and universities (7.6 percent); shopping in large and small, mostly neighborhood-level, stores (13.1 percent); going to grocery stores (7.1 percent), leisure places (movie theatres, for example — 8.6 percent), healthcare centers (2.3 percent), and restaurants (4.0 percent); visiting friends and relatives (5.7 percent), and "other" trips (9.0 percent).

Putting the car before . . .

To avoid unnecessary complexities, we decided to confine our O–D-based retail trade-area analysis to car travel and taxi journeys. Considering that car ownership is well developed in Quebec City (less than 12.5 percent of households have no car, whereas 47.3 percent, 34.0 percent, and 6.2 percent of households own 1, 2, and more than 2 cars, respectively); that the QMA is among the most road overequipped cities in North America (with a highway network totaling 21.7 kilometers per 100,000 inhabitants); and that urban sprawl is prevailing (Quebec City has 490,000 persons sparsely distributed across 548.8 square kilometers) — such an assumption seemed most sensible.

Therefore, of the 174,243 trips reported during the 2001 O–D survey, we assessed only the 128,937 (74 percent) that were car and taxi trips. Once we excluded from the analysis return-to-home trips and those made for driving someone else, we were left with 63,799 car trips. And because our analysis was only interested in retail-oriented activities, we focused exclusively on trips to shopping and leisure centers as well as grocery stores and restaurants. In the 2001 O–D survey, these amounted to 20,900 reported car trips (32.8 percent of all travel using cars). Then, breaking out the data specific to the 10 shopping centers we sought to analyze, we identified 5,575 one-way car trips — corresponding to 56,595 trips by the entire QMA population after extrapolation.

Drive data. Because the O–D survey data did not include travel time and distance information (it did collect departure time but not arrival or travel times), we also found it helpful to supplement the survey with drive-time calculations. Consequently, we developed a computational procedure to estimate trip duration, whereby we used GIS to link the origin and destination points of each trip to the nearest street corner, selecting among 20,262 local street corners throughout the QMA. We completed this complex operation using a GIS-operated topological street network made of 29,035 road segments linked through a total of 20,906 nodes (local street corners and highway connectors). The network defines a transportation graph on which each link is either bidirectional or one-way and provides maximum speed as well as impedance (crossing time and turn penalties). We then implemented a simulation procedure within Caliper

Corporation's (www.caliper.com) TransCAD GIS software using the GISDK language (Thériault et al., 1999) to find the best route and compute total length (kilometers) and duration (minutes) for each reported trip.

Statistical market indicators

Knowing the number of customers traveling during week days from their respective residential locations in the QMA to each of the 10 selected shopping centers, we were able to delineate quite reliable trade areas for those major retail establishments. Moreover, by combining the O–D survey data, census information, and GIS-computed distances, we also calculated a series of statistical retail market indicators to assess sales potential, spatial competition among centers, and shopping center attraction.

To derive market indicators, we first integrated 1996 Canadian census data on household profile and income (2001 census income data was still not available when we completed this research) into a regional GIS operated with MapInfo (www.mapinfo.com) software. During this step, we reshuffled information by enumeration area and aggregated customer home locations according to a finer (500-meter radius), hexagonal grid containing 6,150 cells. (In Canada, enumeration areas are modified before each census to reflect growth of urbanized land. Using a hexagonal grid allows for temporal comparisons among censuses and O–D surveys. Moreover, it provides better mapping of densities, because enumeration areas in remote locations are too large and include agricultural and forested land.)

Economic potential. The first index we calculated was the Economic Potential Index (EPI), which measures the yearly sales potential for each shopping center as the product of the number of customers shopping at center i (C_i) and originating from grid cell j, multiplied by the annual personal income for cell j (R_j). A center's EPI is then expressed as a percentage of total EPIs for all centers. The equation for calculating EPI is

$$EPI_i = \sum_{j=1}^{6150} \left[C_{ij} * R_j \right] * 100 / \sum_{i=1}^{10} EPI_i$$

Spatial competition. We used a second retail market index, the Spatial Competition Index (SCI), to measure to what extent a given shopping center loses customers originating from any cell j to competitors. SCI is expressed as the squared complement of the rate of penetration pertaining to each center (C_{ij}/C_j), multiplied by the cross product of cell population (P_j) and personal income (R_j), which serve as weighting factors. In this case, the lower the index, the better the competitive position of a shopping center. Again, we translated the SCI for each center into a percentage of total SCIs for all centers according to

$$SCI_i = \sum_{j=1}^{6150} \left[\left(1 - C_{ij}/C_j\right)^2 * P_j * R_j \right] * 100 / \sum_{i=1}^{10} SCI_i$$

Attraction. Finally, for each center, we derived a Center Attraction Index (CAI), which is a direct adaptation of Reilly's (1929) model of retail gravitation. It measures the combined influence of shopping center size (S_i) based on GLA and the squared distance of customers to center i from any cell j (D_{ij}^2). A computed accessibility ratio is then weighted by the cell population (P_j) and the CAI expressed as a percentage of total CAIs. The CAI equation is written as

$$CAI_i = \sum_{j=1}^{6150} \left[\left(S_i / D_{ij}^2\right) * P_j \right] * 100 / \sum_{i=1}^{10} CAI_i$$

Table 1 shows the results of those computations. They confirm the prevalent sales potential of the two super-regional centers (C6 and C8), which have 43.1 percent of the overall GLA for all 10 shopping centers. The SCIs also suggest that the super-regional centers are relatively sheltered from excessive competition. The two regional centers (C5 and C10) also have less spatial competition. As for the attraction index, the high CAI assigned to the low-class regional center (C5) suggests that, in spite of its lower sales potential, it attracts a large clientele from nearby densely populated neighborhoods.

TABLE 1 Economic potential, spatial competition, and center attraction indices for Quebec City shopping centers

Center ID and type	EPI (percent)	SCI (percent)	CAI (percent)
C1 (community)	7.3	10.6	2.7
C2 (community)	5.4	11.3	4.4
C3 (community)	3.5	11.3	5.1
C4 (fashion)	3.9	11.3	3.2
C5 (regional)	15.8	8.9	30.6
C6 (super-regional)	21.8	7.5	15.0
C7 (community)	4.6	11.4	3.8
C8 (super-regional)	22.5	6.9	21.7
C9 (community)	5.3	11.1	3.7
C10 (regional)	10.0	9.8	9.8

Statistical insights . . . can be complemented and eventually qualified by visual representations of the phenomena under analysis and through the use of basic spatial analysis

Sales potential and dispersion ellipses

Although Table 1 provides useful statistical insights about Quebec City's retail dynamics, the information it conveys can be complemented and eventually qualified by visual representations of the phenomena under analysis and through the use of basic spatial analysis.

In Figure 1, for instance, we localized all 10 shopping centers using colored push-pins and plotted yearly sales potential for each center. Each residential neighborhood in the QMA was thus assigned a pie chart which identifies the volume of customers (shopping trips) by retail destination, multiplied by the average personal income in the neighborhood. Quite interestingly, the two largest, super-regional centers, C6

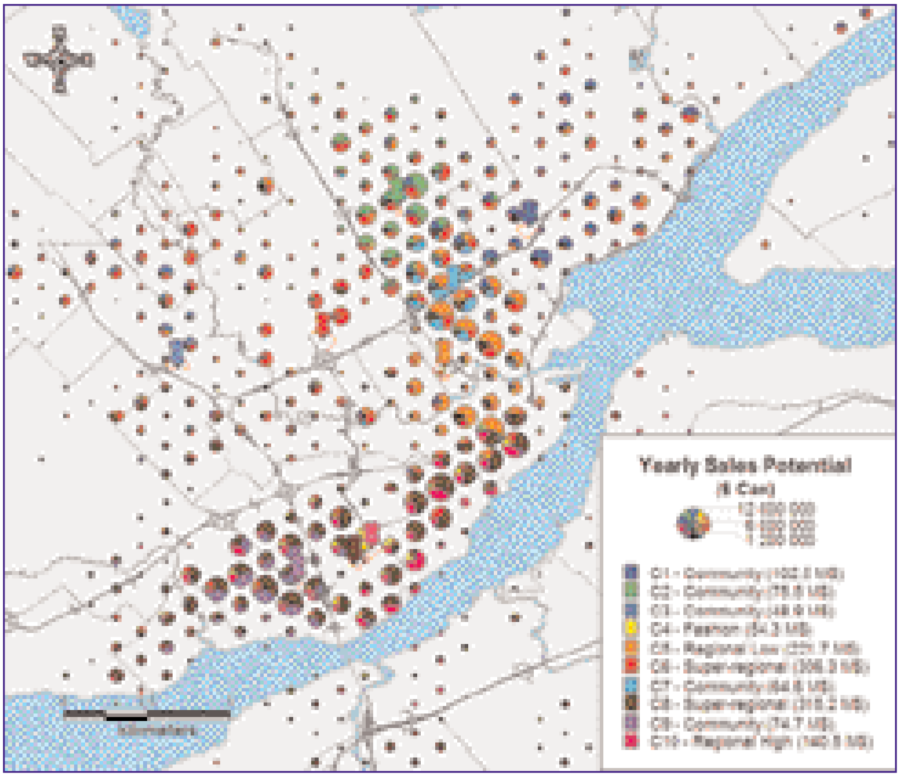


FIGURE 1 Using colored push-pin icons, the authors plotted yearly sales potential for the 10 shopping centers in their study. Each residential neighborhood in the QMA was then assigned a pie chart which identifies the volume of customers (shopping trips) by retail destination, multiplied by the average personal income in the neighborhood.

(red) and C8 (dark-brown), exhibit an almost identical sales potential — $306.3 and $315.2 million (Canadian), respectively. Interestingly, C6 seems to dominate the north QMA and to attract customers from residential neighborhoods scattered along an east–west axis, whereas C8 clearly benefits from a vast pool of high-income households concentrated along the Saint-Lawrence River axis (upper city) who also shop at nearby luxury regional centers, such as C4 (yellow) and C10 (magenta).

Although less fashionable as a shopping destination, C5 (orange) still enjoys an interesting sales potential ($221.7 million Canadian) from its well delineated trade area, mainly confined to the southeastern part of Quebec City. Finally, community centers C1 (dark-blue), C2 (green), C3 (light-blue), C7 (turquoise), and C9 (violet) all exhibit, as expected, more local trade areas, with sales potential varying from a low of $48.9 million (Canadian) for C3 to an high of $102.5 million (Canadian) for C1.

Turning to Figure 2, we created this centrographic analysis (Lefever, 1926; Raine, 1978) map to estimate with greater accuracy the extension of shopping center trade areas, with dispersion ellipses summarizing the distributional pattern of customers. The ellipses' areas do not encompass all of a center's customers, but rather they provide an "outlier-free" measure about how customers of a given center spread throughout the region. In addition, the ratio between the mutually orthogonal major and minor axes of the ellipses indicate the degree of anisotropy around a trade-area's gravity center. Moreover, actual distance and directional orientation from each shopping center to its trade-area gravity center provides clues about the effects of spatial competition.

The dispersion ellipse analysis enabled us to qualify previous findings about shopping centers' trade areas — with markedly different trade areas emerging from the analysis. It also highlighted the dynamics of spatial competition in the QMA retail market. Firstly, the centrographic analysis confirmed that the trade area relative to C6 (92.6 square kilometers) is actually confined to the northern portion of the territory and displays an east–west orientation. Secondly, and most interestingly, C8 (105.7 square kilometers) displays a trade area that extends in a north–south direction, drawing customers from the south shore of the river and expanding deep into its main competitor's trade area. C8 even includes C6 in its own dispersion ellipse. A similar pattern emerged for fashion and regional centers C4 and C10 (93.5 and 78.4 square kilometers, respectively), which offer specialized, high-value goods and services to a widespread clientele. The dispersion ellipse of center C5, in turn, is almost circular, thereby corroborating the more local impact on this shopping establishment.

As for community centers, their trade areas vary in size, shape, and direction depending on size, location in the region, and density of neighboring populations. Thus, for instance, centers C1 (58.8 square kilometers) and C3 (50.1 square kilometers) have elongated trade areas that reach customers far into the eastern and western parts of the QMA, whereas centers C2 (31.3 square kilometers) and C7 (32.9 square kilometers) exert a more limited influence resulting in circular trade areas that suggest an isotropic customer distribution. In addition, as can be seen from the large overlap between their respective ellipses, C1, C2, and C7 harshly compete for customers drawn from similar neighborhoods. Finally, C9 (61.3 square kilometers) displays an extended trade area over the southwest portion of the QMA.

Consequently, our centrographic analysis remained consistent with and validated the market indices shown in Table 1. For instance, the size, shape, and orientation of C8's dispersion ellipse adequately capture both the SCI obtained for the center — the lowest of all at 6.9 percent — and its 21.7 percent CAI, which stands well above that of center C6's 15.0 percent CAI. Moreover, the overall state of the spatial competition for retail trade

Continued on page 51

Reader Service Card

Want FREE information from advertisers featured in this issue?

Just fill out the card to the right, circling the number of the advertisers, products, or services you'd like to learn more about.

Find Reader Service Numbers at the bottom of each ad throughout the magazine or use the Advertisers Index as a handy tool to quickly pinpoint the circle number you seek.

Advertisers Index

Want the information NOW?

Point your Web browser to www.geospatial-online.com, click on the Reader Service link, and fill out an electronic Reader Service Card for a speedy reply.

Continued from page 49

at a given point in time is best illustrated by the distance between shopping centers and the gravity center of their respective dispersion ellipses. As shown by Figure 2, the major shopping complex consisting of C4, C8, and C10 extends its influence deep into the trade areas of shopping centers located north of the region. Furthermore, the combined dispersion ellipses of C4, C8, and C10 include C5 and C6, and, as a consequence, push the gravity centers of the C5 and C6 ellipses even further north. In contrast, apart from center C3, which undergoes a similar displacement westward, all community shopping centers — particularly those with a very local vocation (C2, C7, and C9) — seem less affected by this magnet-type repulsion phenomenon.

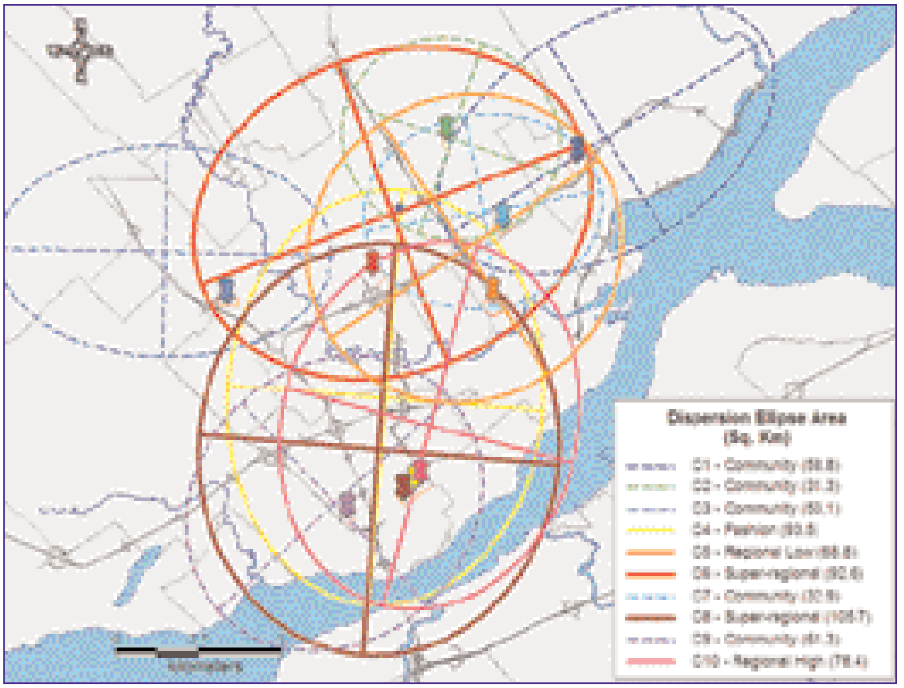


FIGURE 2 This centrographic analysis map shows dispersion ellipses summarizing the distributional pattern of customers for all 10 shopping centers. The ellipses provide an overall measure of customer distribution and the ratio between the mutually orthogonal major and minor axes of the ellipses indicate the degree of anisotropy around a trade-area's gravity center. The analysis also indicates the actual distance and directional orientation from each shopping center to its trade-area gravity center.

Working for the weekend

Overall, our study demonstrated that, as with LSPs and other GIS-driven analytical procedures, O–D surveys and corresponding spatial statistics can greatly enhance analysts' understanding of retail market dynamics and stakes. O–D surveys, then, are multifunction tools that serve both public (transportation departments, for example) and private planning purposes and interests. Plus, although we used centrographic analysis from a static perspective only, that approach could be applied at several points in time (from O–D surveys contacted at different times during a single year or across multiple years), thereby providing precious insights into retail market trends and evolving competition among retailers.

Lastly, it should be noted that we did experience one major flaw related to the O–D survey methodology with respect to the retail application discussed in this article. In the QMA, shops are open seven days a week and more than half of all shopping trips are actually made during weekends — a time period not covered by the 2001 O–D survey. Because people have more time to travel during weekends, we suspect that our study, therefore, somewhat underestimates the extent of shopping center trade areas. In addition, our analysis risks miscalculating the attraction impact of expensive goods (furniture, household appliances, home renovation goods, and so forth) — mostly shopped during weekends — relative to more frequently used and less expensive supplies (grocery, clothes, and so forth). Thus, retail analysts should remember that O–D survey methodologies are designed primarily for work-trip planning purposes. Obviously, this technical problem can be overcome by extending surveys to collect data about weekend travel.

References

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

Lefever, D.W. "Measuring Geographic Concentration by Means of The Standard Deviational Ellipse." *American Journal of Sociology* 32 (1926): p. 88.

Raine, J.W. "Summarizing Point Patterns with the Standard Deviational Ellipse." *Area* 10 (1978): p. 328.

Reilly, W.J. *Methods for the Study of Retail Relationships.* Austin: Austin Bureau of Business Research, University of Texas, 1929.

Thériault, M., M.H. Vandersmissen, M. Lee-Gosselin, and D. Leroux (1999). "Modelling Commuter Trip Length and Duration within GIS: Application to an O–D Survey." *Journal for Geographic Information and Decision Analysis* 3, no. 1 (1999): p. 41.

Thrall, G.I. *Business Geography and New Real Estate Market Analysis.* New York: Oxford Press, 2002.

Thrall, G.I. and M. Fandre. "Trade Areas and LSPs: A Map for Business Growth." *Geospatial Solutions* 13, no. 4 (2003): p. 48. ⊕

Grant Ian Thrall, Ph.D., is column editor of Shop Talk and First Impressions. He is a consultant; a professor at the University of Florida, Gainesville; and a member of the *Geospatial Solutions* Editorial Advisory Board. Thrall can be reached at thrall@ufl.edu