[DM-05-016] Linear Referencing

Abstract

Linear referencing is a term that encompasses a family of concepts and techniques for associating features with a spatial location along a network, rather than referencing those locations to a traditional spherical or planar coordinate system. Linear referencing is used when the location on the network, and the relationships to other locations on the network, are more significant than the location in 2D or 3D space. Linear referencing is commonly used in transportation applications, including roads, railways, and pipelines, although any network structure can be used as the basis for linearly referenced features. Several data models for storing linearly referenced data are available, and well-defined sets of procedures can be used to implement linear referencing for a particular application. As network analysis and network based statistical analysis become more prevalent across disciplines, linear referencing is likely to remain an important component of the data used for such analyses.

Keywords: dynamic segmentation, geocoding, linear referencing, network analysis, transportation, utilities

Author & citation

Curtin, K. and Turner, D. (2019). Linear Referencing. The Geographic Information Science & Technology Body of Knowledge (4th Quarter 2019 Edition), John P. Wilson (ed.).
DOI: 10.22224/gistbok/2019.4.3.

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Linear Referencing. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

and

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Linear Referencing Systems. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Definitions
2. What is Linear Referencing?
3. Implementing Linear Referencing
4. Applications of Linear Referencing
5. Analysis of Linear Referenced Data

1. Definitions

Linear Referencing System: a structure which is the basis for determining locations in a linear or network spatial domain. The system generally contains routes built from linear features, a definition of the origins of the routes, well-defined metrics along the routes, and the events that are located along the routes.

Routes: The foundational features on which linear referencing occurs. The set of routes represents the well-defined datum or spatial domain from which location references are generated.

Measures: Measures are values along the routes that are used to specify linearly referenced locations. Measures can be in standard metrics (e.g. feet, meters), or they can specify user defined costs/distances/impedance along the routes. Measures may indicate the percentage of the total route travelled. Measures can be defined by calculations in the GIS, or can be specified from outside data sources when more accurate distance or cost information is available from other sources.

Events: Events are the entities that are linearly referenced along routes. These can be the mile markers along highways, traffic accidents or potholes on streets, construction zones along railways, or pipe diameters along pipelines, among many more.

2. What is Linear Referencing?

Much, if not most, geographic analysis begins with the identification of a spatial reference on which locations for the phenomenon under study will be specified. Very frequently that spatial reference will either be based on a global ellipsoidal model of the earth, or on a planar representation of a portion of the earth’s surface. Both types of spatial reference depend on clear definitions (and broad acceptance) of the nature of the space on which locations are specified, an origin in that space, and a metric for measurements away from that origin. In the case of typical ellipsoidal systems, the space is defined by the parameters of the ellipsoid, the origin is most often the intersection of the equator and the prime meridian, and the metric for measuring offsets from the origin is angular measurements in degrees, minutes and seconds for latitude and longitude. Similarly, in planar spatial reference systems the space is a plane with limits as to its extent (e.g. a UTM zone), the origin is a well-accepted point within that extent (e.g. a survey marker or other significant well-known point), and the metric may be projected coordinates measured in feet or meters away from the origin. There are many hundreds or thousands of planar spatial reference systems that are designed for use for specific parts of the globe, or for different countries, governments, or agencies. Therefore, it is not unusual for geographic analysts to accept that locations can be expressed in many different ways based on the area under study or the nature of the study itself.

There are cases where the location in ellipsoidal or planar space is not significant for the analysis or application, but rather where the location on a network and the relationships between locations on a network are of primary interest. In these cases, a referencing system can be developed to specify the nature of the space (the network itself), the origin or origins to be used in the network, and the metrics for measuring offsets from an origin. Such a reference is termed a linear referencing system (LRS) and the concepts and techniques for using such a system to specify locations in the network space is more generally referred to as linear referencing.

The LRS with which most persons in the US will be familiar is that which defines the mile markers along US interstate highways (Federal Highway Administration, 2001; Federal Transit Administration, 2003). Those mile markers represent actual network distances from the origin of the particular numbered interstate being travelled. In that system the origins of the roads are (generally) specified to be the southern and western boundaries of the state through which the interstate runs, with measures increasing to the north and east, respectively. Each mile marker indicates the number of miles in true travel distance from the origin of the road to that marker. Using this system as an example consider that it can often be far more useful to specify a location by referencing a mile marker than it is to specify an ellipsoidal or planar spatial reference. More specifically, consider the case where a traffic accident or other emergency occurs on the highway. While a latitude/longitude or State Plane coordinate pair could certainly provide the location of the incident, those coordinates are not particularly useful to an emergency response crew attempting to reach the location. Conversely, the specification of a mile marker conveys not only the location itself, but also the distance to the location given that it can be quickly calculated by subtracting the incident mile marker from the response crew location mile marker, and also indicates the necessary direction of travel to reach the incident.

In the following sections the process for implementing linear referencing is outlined, the current and potential future applications for linear referencing are reviewed, and a discussion of analytical methods with linearly referenced data is provided.

3. Implementing Linear Referencing

There are several theoretical linear referencing data models – which work to identify the common elements and relationships inherent in linear referencing (Scarponcini, 2002) – and many of these were developed under the auspices of the National Cooperative Highway Research Program project 20–27 (Koncz & Adams, 2002; Vonderohe et al., 1998; Vonderohe, Chou, Sun, & Adams, 1997). In order to encourage broader use of linear referencing, this article follows a more practical approach to implementing linear referencing in a GIS as outlined by Curtin et al (2007).

Generally, linear referencing includes methods for associating events with a linear feature. Given a line feature with a starting and end point, a linear reference is given as an event at some interval from one of the end points (typically, the origin or starting point of the line segment). Thus, an event’s location is determined by its relative location on the underlying linear structure rather than a conventional coordinate location. The event may be singular or mark the start or end point of some continuous condition or feature on the line. For example, a line segment representing a street may have events defined at some distance interval from one of the end points that signify the location of light poles, the start and end points of a school zone speed limit, and the start and end points of a stretch of traffic guardrails.

Implementing linear referencing requires a series of steps that successively lay the foundation for the referencing system, define how the references will be captured, and finally locate events on the linear features so that they can be visualized and analyzed. Just as with the selection of a more traditional ellipsoidal or planar spatial reference systems, the choices made in defining a linear referencing system require a recognition of the needs

of the application and analysis to be eventually performed once the locations are determined. The analyst should – in advance – determine the application (e.g. vehicle locations on streets, inspection locations along pipelines), the nature of the underlying network (e.g. single street centerlines or dual carriageways), and the topological rules that apply to that network. While linear referencing can be used with any application where locations are network-based, it is critical to recognize that, for example, road networks and river networks have fundamentally different structures, and the nature of the network will influence the definition of the linear referencing system.

With an understanding of the application and the underlying network, the next (and critical) step in implementing linear referencing is to clearly identify the linear entities on which the measurements will be made. The underlying network elements in their original form may not be the same entities on which the measurements are made. Take for example the case of a street centerline road network based on the U.S. Census Bureau's TIGER Line data model. This type of data model enforces a split of linear features wherever there is a crossing or intersection of two features. Therefore, a single street – say Main Street – will be represented by a series of connected features in the database, representing each of the lengths of the street between intersections. In a typical US city there may be many dozens of street segments that make up a single named street. However, it is the entirety of the street that is needed to make measurements of how far along Main Street a particular event is located. The individual street segments need to be aggregated into a single feature, often termed a route, and it is the set of routes that correspond to the datum (or earth model) in traditional spatial referencing systems. The set of routes is the spatial domain on which it is agreed that locations will be specified. In Figure 1 features with Feature ID numbers (FIDs) 3 and 4 make up Route 1, and features with FIDs 1 and 2 are aggregated to form Route 2.

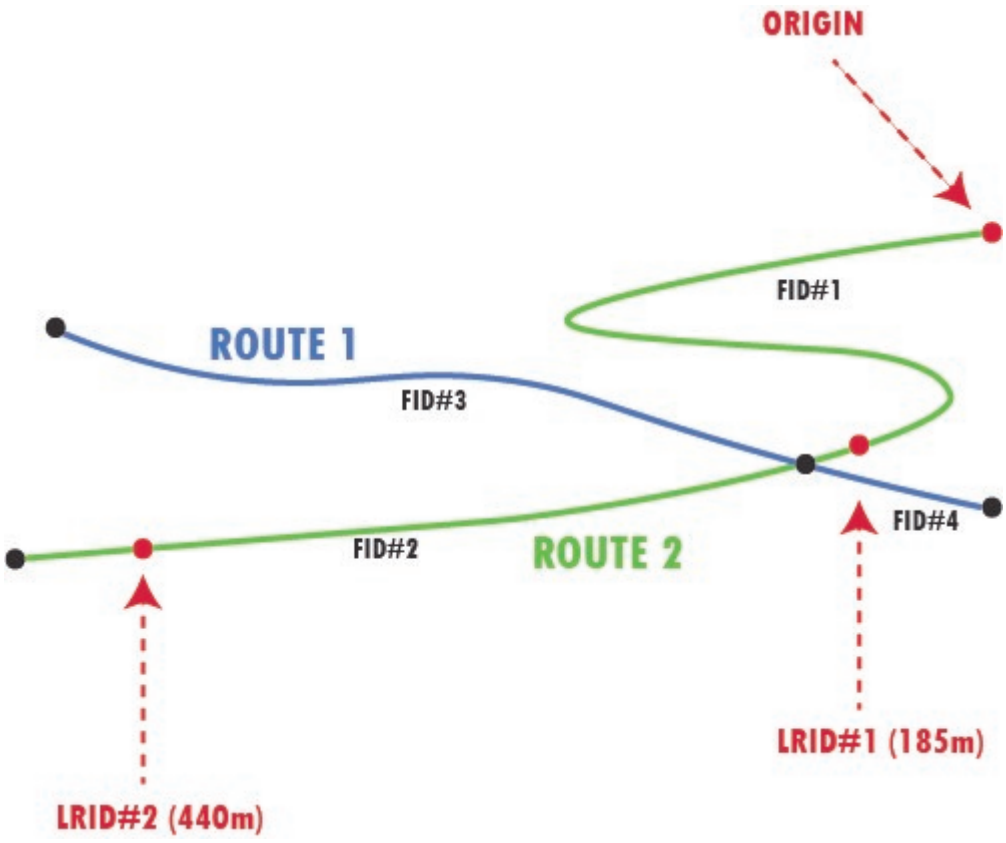


Figure 1. Routes are aggregations of linear features. Each route has an origin. Measures are

associated with routes and located in relation to the origin. Source: authors.

In contrast to conventional spatial referencing systems which generally have a single origin (or at least a single origin for each zone), each route in a linear referencing system has its own origin. As part of the process of defining the routes, the analyst must also decide where the origin of each route will be. This decision will likely be based in part on the knowledge of the application and original network gained in the first step as described above. For example, a river network may have a single route representing the main branch of the river, and additional routes for tributaries. The zero point for these routes could be the headwaters of each branch, with measures increasing with distance downstream. As another example consider the addressing system of the city of Chicago, IL. The origin of this system is at the intersection of State St. and Madison St. All addresses on East-West streets east of State will have “East” or “E.” prefixes and address numbers increase moving to the east. Conversely addresses on East-West streets west of State St. have “West” or “W.” prefixes and address numbers increase moving to the west. The similar situation exists for North-South streets that lie to the north or south of Madison St. If one wishes to define linear referencing routes for Chicago that mirror the addressing system, then the routes will have origins either on State St. or Madison St. and the measures will increase along the cardinal directions away from these origins. Figure 2 shows these routes for East (red), North (green) and South (blue).

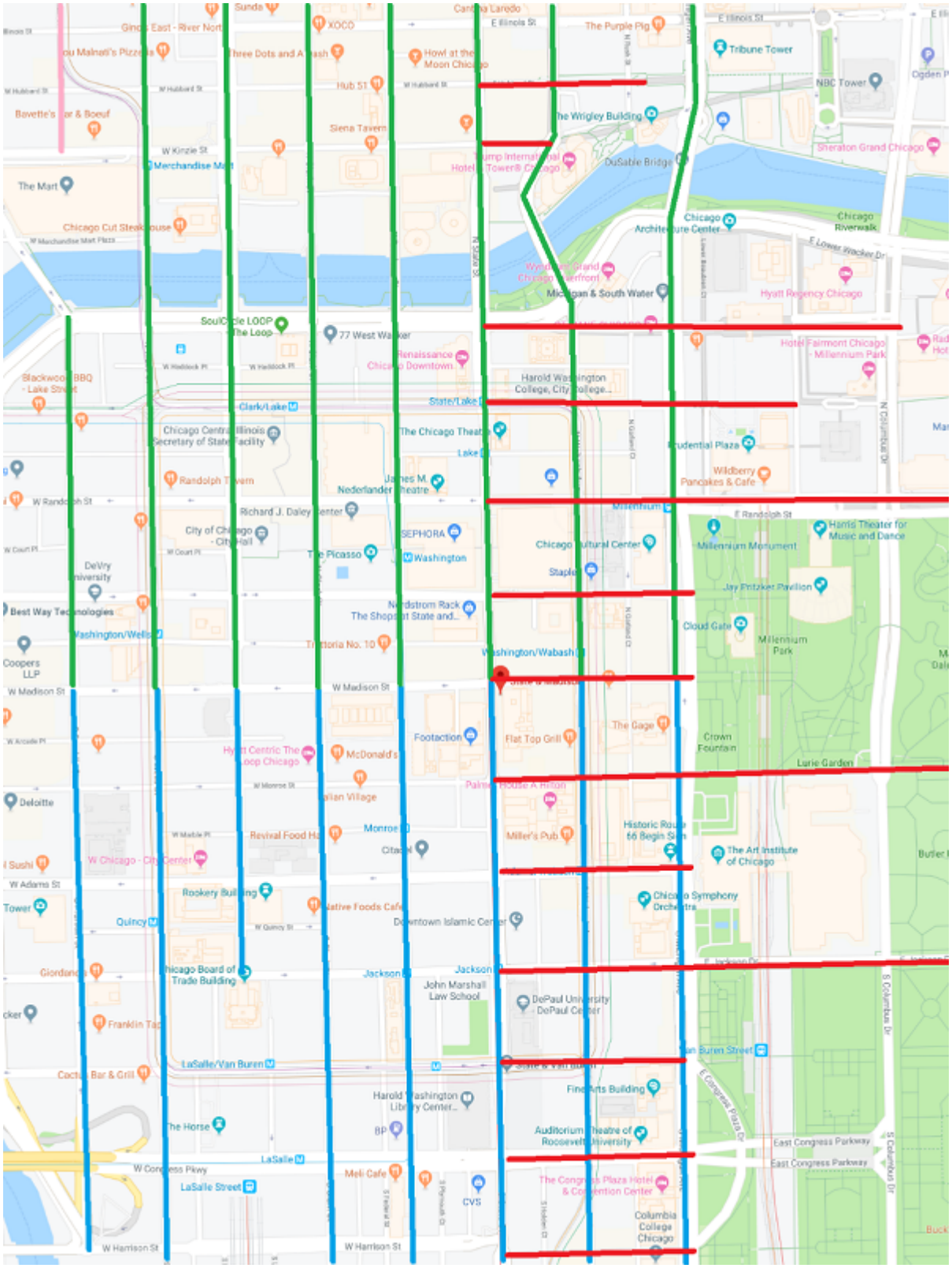


Figure 2: Routes designed to follow the street addressing system of Chicago. Source: authors.

With a set of routes and origins determined, the next step is one that distinguishes linear referencing significantly from traditional spatial referencing systems. The analyst must assign measure values along the routes from their origins. This requires first determining

the units in which the measures will be expressed, then assigning the measures based on trusted source data. Most commonly the measures along a route represent distance from the origin, and this distance can be expressed in any common measurement system (e.g. feet, meters, miles) and these values can be converted among standard metrics with little difficulty. However, the measure can also represent any cost or impedance of traversing the network (e.g. time, effort, fuel consumption, cumulative risk). The measures may also represent a percent of the distance from the origin of the route to its end.

The association of trusted source measure data with the routes is a characteristic unique to linear referencing among spatial references. Even a sophisticated GIS analyst might presume that the distances calculated within the GIS (distances or lengths of lines are routinely calculated from the planar coordinates of line vertices) would be used to populate the measures along the routes, and in fact this is often the case. However, those calculated distances are, in some cases, not sufficiently accurate for the application. There are two reasons for this: first the planar calculations are made on straight line approximations of the real-world feature, and second the planar measurements do not consider the additional distance due to changes in elevation. Both of these factors introduce error into the calculation that could be avoided by using measurements captured in another way, perhaps even by direct measurement in the real-world. In fact, multiple instantiations of the routes with different measures from different sources can be maintained to reflect competing measurements, or changes in measurements over time, all without editing the underlying geometry of the network features.

The fourth step in linear referencing is one that is often assumed to be the first – that is, to define how events will be associated with the network. In fact, event data are those that exploit the routes and measures previously defined. An event could be a traffic accident at a distance along a road, or a length of railroad track to repair. Events can be either point or linear features. When they are point features they simply are referenced with a single measure along a specified route. When they are linear features they have a starting and ending measure associated with a route. Many different sets of events can be associated with routes. For example, on a particular set of street routes one event table could hold the locations of street signs, another could store construction zone locations, and a third could identify school zones (Figure 3).



Figure 3: Many types of point and linear events referenced to a route.Source: authors.

Once the linear referencing system is defined by routes and their measures, and events are linearly referenced to those routes, what remains is to visualize and analyze those events to extract actionable information. The ability to more flexibly represent events cartographically has long been seen as an advantage of linear referencing. Perhaps the most widely known example of such graphics are the maps of metro systems that can show multiple types of events (different train routes) that overlap each other. The use of offsets to avoid visual conflict and promote human comprehension is a hallmark of linear referencing (Figure 4).

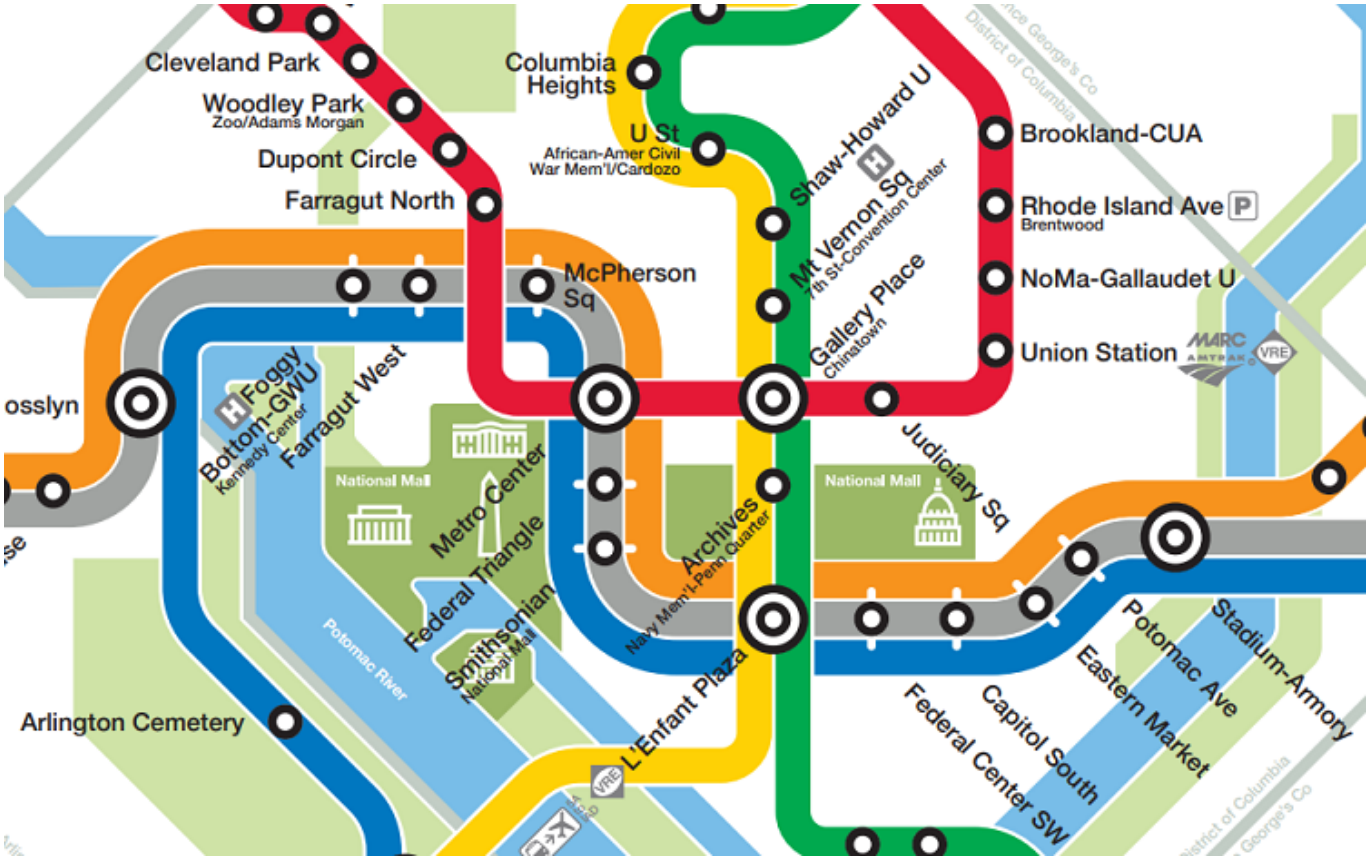


Figure 4: An example of a map that uses offsets to display multiple linear events that occur along the same route. Subway maps often stress the readability of linear features and events to make it easier for viewers to interpret them clearly and quickly.

Finally, and perhaps most importantly, the linear referencing of events permits the spatial analysis of network-based phenomena. The current and potential analytical methods that can exploit linearly referenced events are explored further in the following sections. As with any process that requires the design and use of spatial databases, the process should be revised and maintained as new data becomes available, and as new applications are addressed.

4. Applications of Linear Referencing

Linear referencing can be applied to any phenomenon that operates on a linear or network spatial domain. As described above, perhaps the most well-known use of linear referencing is the use of mile markers on highways. Other common applications for linear referencing also fit broadly into the area of transportation – in the sense of the movement or communication of some good or information. This includes applications along rivers where the locations of infrastructure (e.g. bridges or flow monitors) are maintained. Other transportation applications involve the location of traffic incidents along roadways (Pande et al., 2017), locating vehicles during their operations (Zhou et al., 2017), maintaining the locations of needed repairs (e.g. potholes), or building an inventory of the locations of transportation assets (e.g. street signs) (Federal Highway Administration, 2001). Pipelines built for the transit of liquids and gasses can also benefit from linear referencing where gauges, valves, and other critical equipment are located along the pipeline network (Kukalo

& Grivtsov, 2015). Other utility applications such as cable television and internet networks, electricity transmission, and water and sewer systems can similarly benefit from locating network elements with linear referencing.

Linear referencing is also related to the concept of addressing on city streets, where routes and the origins of those routes are well-defined, and one can deduce a location based on an address if the direction and metric are known. In the Chicago example given in Figure 2, if one is aware that there are 8 blocks to a mile in Chicago, then an address near 4400 N. Racine Avenue must be 5.5 miles north of Madison St. Similarly, the distance between points can be determined based on this referencing system; if a person is at 1955 W. Addison St. and wishes to visit 1060 W. Addison, they know they need to travel 9 blocks east which is 1 and 1/8th mile. The use of an addressing system for linear referencing should not be confused with the concept of address geocoding, where knowledge of the addressing system is used to interpolate x,y coordinates in planar space.

Given the significant advances in network-based analysis in the recent past it is likely that there are many areas that have not yet begun to be explored with the potential of linear referencing. Consider that social network analysis is in its research infancy, and measures of distance between actors in a social network is a fundamental characteristic of those relationships. Imagine applications to the nervous or cardiovascular systems of the human body, where injuries or defects are located in proximity to crucial points on those systems. These applications and many more (e.g. information networks) represent the future of linear referencing.

5. Analysis of Linear Referenced Data

One of the obstacles in the adoption of linear referencing historically is the paucity of statistical analysis methods for points on a network. The analysis of stochastic point processes on Euclidean planes is well established, but for many applications a network geographical space is the more appropriate and less biased context. For example, planar analysis may indicate clustering of events around nodes (such as accidents around intersections) when, in fact, the events occur randomly along network edges. Indeed, this may apply to any analysis where route distance on a discrete network is more appropriate than Euclidean distance.

Many spatial statistics have been developed for network analysis that correspond to similar measures developed for planar space (Okabe & Sugihara, 2012; Okabe, Yomono, & Kitamura, 1995). These include measures for determining whether a network-based point distribution is even/dispersed, random, or aggregated/clustered, as well as the relative effects of different categories of network edges (i.e. different street types) and points (i.e. proximity to different facilities). For example, Okabe et al. (2009) describe kernel density estimation methods for points on a network (Figure 5).

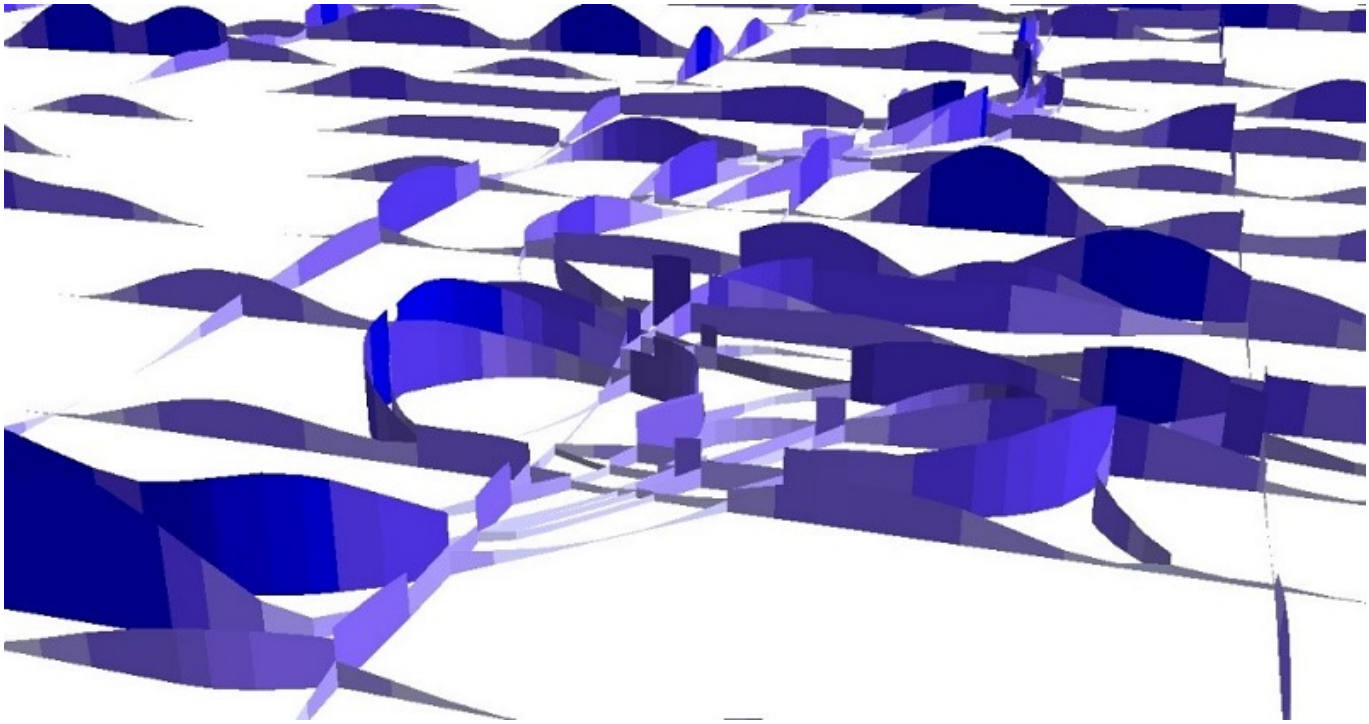


Figure 5: Point densities on the Manhattan street network around the Holland Tunnel entrance created using SANET network spatial analysis software (http://sanet.csis.u-tokyo.ac.jp/). Note how estimations from edge to edge can be discontinuous. Source: authors.

The value of incorporating linear referencing in GIS software – particularly for transportation applications – has been well-known for some time (Nyerges, 1990), and over time tools to implement linear referencing have been increasingly incorporated into the software (Goodman, 2001). However, these tools are largely focused on the implementation of linear referencing rather than on the analysis of linearly referenced events. The richness of analytical methods including overlay operations, buffering and other proximity tools, 3D analysis, and even robust cartographic tools are still being developed for linearly referenced data (Okabe & Sugihara, 2012).

In summary, the existence of linear referencing is a recognition that linear features and networks are a foundational spatial domain, and there are some phenomena that only exist – and are best modeled – in that domain. Given the persistence of networks in the context of GIS, the utility and complexity of linear referencing and related analytic methods are likely to increase in both number and complexity.

References

Curtin, K. M., Nicoara, G., & Arifin, R. R. (2007). A Comprehensive Process for Linear Referencing. URISA Journal, 19(2), 41–50.

Federal Highway Administration (FHWA). (2001). Implementation of GIS Based Highway Safety Analysis: Bridging the Gap (No. FHWA-RD-01-039). Technical Report, U.S. Department of Transportation.

Federal Transit Administration. (2005). Best Practices for Using Geographic Data in Transit: A Location Referencing Guidebook (No. FTA-NJ-26-7044-2003.1). U.S. Department of Transportation.

Goodman, J. E. (2001). Maps in the Fast Lane—Linear Referencing and Dynamic Segmentation (Vol. 2004). Directions Magazine.

Koncz, N. A. & Adams, T. M. (2002). A data model for multi-dimensional transportation applications. International Journal of Geographical Information Science, 16(6), 551–569.

Kukalo, I. A., & Grivtsov, S. N. (2015). Linear Referencing of Moving Object Geo-Coordinates to the Linear Part of the Main Oil Pipelines. Bulletin of the Tomsk Polytechnic University-Geo Assets Engineering, 326(11), 31–43.

Nyerges, T. L. (1990). Locational Referencing and Highway Segmentation in a Geographic Information System. Institute for Transportation Engineers. ITE Journal, March, 27–31.

Okabe, A., & Sugihara, K. (2012). Spatial Analysis Along Networks: Statistical and Computational Methods. John Wiley and Sons.

Okabe, A., Satoh, T., & Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. International Journal of Geographical Information Science, 23(1), 7-32.

Okabe, A., Yomono, H., & Kitamura, M. (1995). Statistical Analysis of the Distribution of Points on a Network. Geographical Analysis, 27(2), 152–175.

Pande, A., Chand, S., Saxena, N., Dixit, V., Loy, J., Wolshon, B., & Kent, J. D. (2017). A preliminary investigation of the relationships between historical crash and naturalistic driving. Accident Analysis and Prevention, 101, 107–116.

Scarponcini, P. (2002). Generalized model for linear referencing in transportation. Geoinformatica, 6(1), 35–55.

Vonderohe, A., Adams, T., Chou, C., Bacon, M., Sun, F., & Smith, R. L. (1998). Development of System and Application Architectures for Geographic Information Systems in Transportation (Research Results Digest No. 221). National Cooperative Highway Research Program, Transportation Research Board.

Vonderohe, A., Chou, C., Sun, F., & Adams, T. (1997). A generic data model for linear referencing systems (No. Research Results Digest Number 218). National Cooperative Highway Research Program, Transportation Research Board.

Zhou, Y., Zhang, Y., Ge, Y., Xue, Z., Fu, Y., Guo, D., … Li, J. (2017). An efficient data processing framework for mining the massive trajectory of moving objects. Computers, Environment and Urban Systems, 61, 129–140.