Benchmarking Spatial Big Data

Shashi Shekhar[1], Michael R. Evans[1], Viswanath Gunturi[1], KwangSoo Yang[1], and Daniel Cintra Cugler[2]

[1] Computer Science & Eng. Faculty, University of Minnesota
200 Union Street S.E. #4192, Minneapolis, MN 55455, USA
[2] Institute of Computing, University of Campinas, Campinas, SP, Brazil
{shekhar,mevans,gunturi,ksyang}@cs.umn.edu,
danielcugler@ic.unicamp.br

Abstract. Increasingly, location-aware datasets are of a size, variety, and update rate that exceeds the capability of spatial computing technologies. This paper addresses the emerging challenges posed by such datasets, which we call Spatial Big Data (SBD). SBD examples include trajectories of cell-phones and GPS devices, vehicle engine measurements, temporally detailed road maps, etc. SBD has the potential to transform society via a number of new technologies including next-generation routing services. However, the envisaged SBD-based services pose several significant challenges for current spatial computing techniques. SBD magnifies the impact of partial information and ambiguity of traditional routing queries specified by a start location and an end location. In addition, SBD challenges the assumption that a single algorithm utilizing a specific dataset is appropriate for all situations. The tremendous diversity of SBD sources substantially increases the diversity of solution methods. Newer algorithms may emerge as new SBD becomes available, creating the need for a flexible architecture to rapidly integrate new datasets and associated algorithms. To quantify the performance of these new algorithms, new benchmarks are needed that focus on these spatial big datasets to ensure proper comparisons across techniques.

Keywords: Benchmarking, Spatial Big Data.

1 Introduction

Spatial computing is a set of ideas and technologies that facilitate understanding the geo-physical world, knowing and communicating relations to places in that world, and navigating through those places. The transformational potential of mobility services is already evident. From Google Maps [17] to consumer Global Positioning System (GPS) devices, society has benefited immensely from spatial computing. Scientists use GPS to track endangered species to better understand behavior, and farmers use GPS for precision agriculture to increase crop yields while reducing costs. Google Earth is being used in classrooms to teach children about their neighborhoods and the world in a fun and interactive way. We've reached the point where a hiker in Yellowstone, a biker in Minneapolis, and a

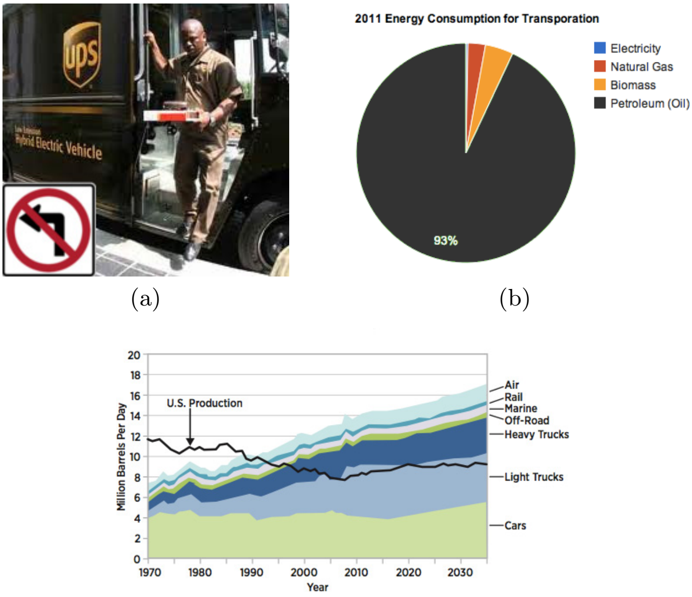
T. Rabl et al. (Eds.): WBDB 2012, LNCS 8163, pp. 81–93, 2014.

taxi driver in Manhattan know precisely where they are, their nearby points of interest, and how to reach their destinations using mobility services [47].

Increasingly, however, the size, variety, and update rate of mobility datasets exceed the capacity of commonly used spatial computing and spatial database technologies to learn, manage, and process the data with reasonable effort. Such data is known as **Spatial Big Data** (SBD). We believe that harnessing SBD represents the next generation of routing services. Examples of emerging SBD datasets include temporally detailed (TD) roadmaps that provide speeds every minute for every road-segment; GPS trace data from cell-phones, and engine measurements of fuel consumption, greenhouse gas (GHG) emissions, etc. SBD has transformative potential. For example, a 2011 McKinsey Global Institute report estimates savings of "about $600 billion annually by 2020" in terms of fuel and time saved [26,29] by helping vehicles avoid congestion and reduce idling at red lights or left turns. Preliminary evidence for the transformative potential includes the experience of UPS, which saves millions of gallons of fuel by simply avoiding left turns (Figure 1(a)) and associated engine idling when selecting routes [26]. Immense savings in fuel-cost and GHG emission are possible if other fleet owners and consumers avoided left-turns and other hot spots of idling, low fuel-efficiency, and congestion. Ideas advanced in this paper may facilitate 'eco-routing' to help identify routes that reduce fuel consumption and GHG emissions, as compared to traditional route services reducing distance travelled or travel-time. It has the potential to significantly reduce US consumption of petroleum, the dominant source of energy for transportation (Figure 1(b)). It may even reduce the gap between domestic petroleum consumption and production (Figure 1(c)), helping bring the nation closer to the goal of energy independence.

A domain-specific benchmark (such as a Spatial Big Data benchmark) should address four key criteria: relevance, portability, scalability and simplicity [18]. Related work in spatial database benchmarking [36, 49] presents workloads for traditional geographic information systems (GIS) related spatial computing, such as raster and vector datasets. Raster data is used in remote sensing (e.g., Google Earth) whereas vector data represents points, lines and polygons, each with their own library of necessary operators. A key-missing component of these related benchmarks is graph-based datasets, useful for applications such as routing and urban navigation. In addition, the rise of spatio-temporal datasets also requires new workloads, as road networks now come with traffic speeds measured every minute of every day.

This paper makes the following contributions: an up-to-date workload for spatial computing, including four types of data (raster, vector, network and spatio-temporal); a set of summary metrics reminiscent of the historical TPS (transactions per second) metrics [18] for Spatial Big Data (SBD-R, -V, -N, -ST) and requirements, both functional (specific behavior) and non-functional (overall operation of a system), for future spatial computing benchmarks.



(a) (b)

(c)

Fig. 1. (a) UPS avoids left-turns to save fuel [26]. (b) Petroleum is dominant energy source for US Transportation [54]. (c) Gap between US petroleum consumption and production is large and growing [5, 10]. (Best in color).

2 Traditional Spatial Big Data

The data inputs of spatial computing are more complex than the inputs of classical computing, as they include extended objects, such as: points, lines, and polygons in vector representation and field data in regular or irregular tessellation, such as raster data. The data inputs have two distinct types of attributes: non-spatial attributes and spatial attributes. Non-spatial attributes are used to characterize non-spatial features of objects such as name, population and unemployment rate for a city. They are the same as the attributes used in the data inputs of classical data mining. Spatial attributes are used to define the spatial location of extent of spatial objects [42]. The spatial attributes of a spatial object most often include information related to spatial locations, for example, longitude, latitude, and elevation, defined in a spatial reference frame, as well as a shape. There are four basic models to represent spatial data: raster (grid), vector (object), network (graph) and spatio-temporal:

Raster: In its simplest form, a raster consists of a matrix of cells (or pixels) organized into rows and columns (or a grid) where each cell contains a value representing information, such as temperature. A set of operations called Map Algebra was introduced [52] to manipulate representations of continuous variables defined over a common domain. These operations were categorized into three categories: local, focal and zonal; each based on the geographic size of the operation. For example, an elevation raster dataset can be queried with a zonal (large region) operation to derive slope. Raster datasets can be digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps.

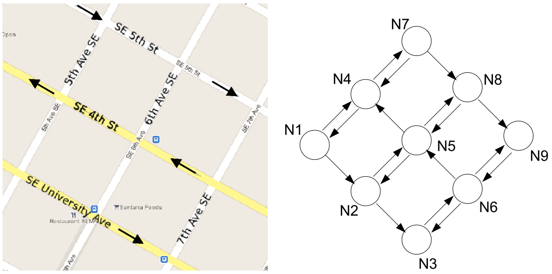
Vector: Geographic features in a vector format can be represented by points, lines, or polygons (areas). Vector data over a space is a framework to formalize specific relationships among a set of objects. In Table 1, a relationship between spatial and non-spatial data is described using spatial relations performed on vector datasets. These relations are separated into a number of classifications: topological to describe relationships regardless of projection, directional to describe orientation and metric to describe distances between objects.

Table 1. Common relationships among spatial and non-spatial data

Non-spatial Relation	Spatial Relation
Arithmetic	Set-oriented: union, intersection, membership, ...
Ordering	Topological: meet, within, overlap, ...
Instance-of	Directional: North, Left, Above, ...
Subclass-of	Metric: distance, area, perimeter, ...
Part-of	Dynamic: update, create, destroy, ...
Membership-of	Shape-based and visibility

Networks: Traditional spatial computing utilizes digital road maps [19, 31, 33, 46]. Figure 2(a) shows a physical road map and Figure 2(b) shows its digital, i.e., graph-based, representation. Road intersections are often modeled as vertices and the road segments connecting adjacent intersections are represented as edges in the graph. For example, the intersection of 'SE 5th Ave' and 'SE University Ave' is modeled as node N1. The segment of 'SE 5th Ave' between 'SE University Ave' and 'SE 4th Street' is represented by the edge N1-N4. The directions on the edges indicate the permitted traffic directions on the road segments. Digital roadmaps also include additional attributes for road-intersections (e.g., turn restrictions) and road-segments (e.g., centerlines, road-classification, speed-limit, historic speed, historic travel time, address-ranges, etc.) Figure 2(c) shows a tabular representation of the digital road map. Additional attributes are shown in the node and edge tables respectively. For example, the entry for edge E1 (N1-N2) in the edges table shows its speed and distance. Such datasets include roughly 100 million (10^8) edges for the roads in the U.S.A. [31].

Route determination services [28, 45], abbreviated as routing services, include the following two services: best-route determination and route comparison [41]. The first deals with determination of a best route given a start location, end location, optional waypoints, and a preference function. Here, choice of preference function could be: fastest, shortest, easiest, pedestrian, public transportation, avoid locations/areas, avoid highways, avoid toll ways, avoid U-turns, and avoid ferries. Route finding is often based on classic shortest path algorithms such as Dijktra's [24], A* [9], hierarchical [20, 21, 43, 44], materialization [38, 40, 43], and other algorithms for static graphs [4, 7, 8, 13, 14, 34, 39]. Shortest path finding is



Nodes

NID
N1
N2
N3
N4
N5
N6
N7
N8
N9

Edges

EID	From	To	Speed	Distance
E1	N1	N2	35mph	0.075mi
E2	N1	N4	30mph	0.075mi
E3	N2	N3	35mph	0.078mi
E4	N2	N5	30mph	0.078mi
E5	N3	N6	30mph	0.077mi
E6	N4	N1	30mph	0.075mi
E7	N4	N7	30mph	0.078mi
E8	N5	N2	30mph	0.078mi
...	...	...	...	...

(a) (b) (c)

Fig. 2. Current representations of road maps as directed graphs with scalar travel time values. (a) Example Road Map [17]. (b) Graph Representation. (c) Tabular Representation of digital road maps.

often of interest to tourists as well as drivers in unfamiliar areas. In contrast, commuters often know a set of alternative routes between their home and work. They often use an alternate service to compare their favorite routes using real-time traffic information, e.g., scheduled maintenance and current congestion. Both services return route summary information along with auxiliary details such as route maneuver and advisory information, route geometry, route maps, and turn-by-turn instructions in an audio-visual presentation media.

OpenLS [28] presents a system (see Figure 3) that incorporates a wide-spectrum of spatial technologies, ultimately reporting to a location-aware client. The location utility performs as a geocoder by determining a geographic position, given a place name, street address or postal code. The directory service provides users with access to the nearest, or a specific place, product or service. The presentation layer renders geographic information for display. The route determination component provides routing information between locations.

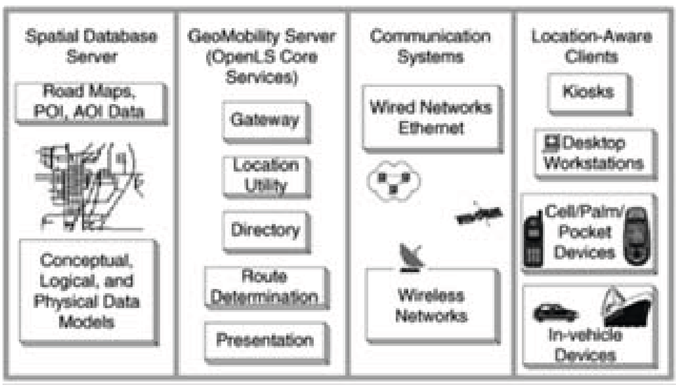


Fig. 3. OpenLS Architecture [28]

3 Emerging Spatial Big Data

Spatio-temporal datasets are significantly more detailed than traditional digital roadmaps in terms of attributes and time resolution. In this subsection we describe three representative sources of SDB that may be harnessed in next generation routing services.

Spatio-Temporal Engine Measurement Data: Many modern fleet vehicles include rich instrumentation such as GPS receivers, sensors to periodically measure sub-system properties, and auxiliary computing, storage and communication devices to log and transfer accumulated datasets [22, 23, 27, 30, 50, 51]. Engine measurement datasets may be used to study the impacts of the environment (e.g., elevation changes, weather), vehicles (e.g., weight, engine size, energy-source), traffic management systems (e.g., traffic light timing policies), and driver behaviors (e.g., gentle acceleration/braking) on fuel savings and GHG emissions.

These datasets may include a time-series of attributes such as vehicle location, fuel levels, vehicle speed, odometer values, engine speed in revolutions per minute (RPM), engine load, emissions of greenhouse gases (e.g., CO_2 and NOX), etc. Fuel efficiency can be estimated from fuel levels and distance traveled as well as engine idling from engine RPM. These attributes may be compared with geographic contexts such as elevation changes and traffic signal patterns to improve understanding of fuel efficiency and GHG emission.

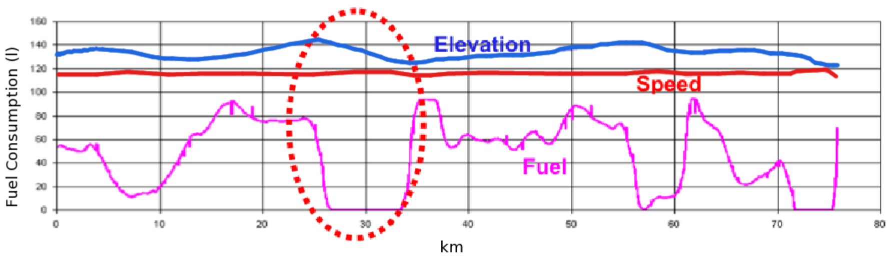


Fig. 4. Engine measurement data improve understanding of fuel consumption [6]. (Best in color).

For example, Figure 4 shows heavy truck fuel consumption as a function of elevation from a recent study at Oak Ridge National Laboratory [6]. Notice how fuel consumption changes drastically with elevation slope changes. Fleet owners have studied such datasets to fine-tune routes to reduce unnecessary idling [1,2]. It is tantalizing to explore the potential of this dataset to help consumers gain similar fuel savings and GHG emission reduction. However, these datasets can grow big. For example, measurements of 10 engine variables, once a minute, over the 100 million US vehicles in existence [12, 48], may have 10^{14} data-items per year.

GPS Trace Data: A different type of data, GPS trajectories, is becoming available for a larger collection of vehicles due to rapid proliferation of cellphones, in-vehicle navigation devices, and other GPS data logging devices [15,58] such as those distributed by insurance companies [57]. Such GPS traces allow indirect estimation of fuel efficiency and GHG emissions via estimation of vehicle-speed, idling and congestion. They also make it possible to make personalized route suggestions to users to reduce fuel consumption and GHG emissions. For example, Figure 5 shows 3 months of GPS trace data from a commuter with each point representing a GPS record taken at 1 minute intervals, 24 hours a day, 7 days a week. As can be seen, 3 alternative commute routes are identified between home and work from this dataset. These routes may be compared for idling, which are represented by darker (red) circles. Assuming the availability of a model to estimate fuel consumption from speed profile, one may even rank alternative routes for fuel efficiency. In recent years, consumer GPS products [15, 53] are evaluating the potential of this approach.

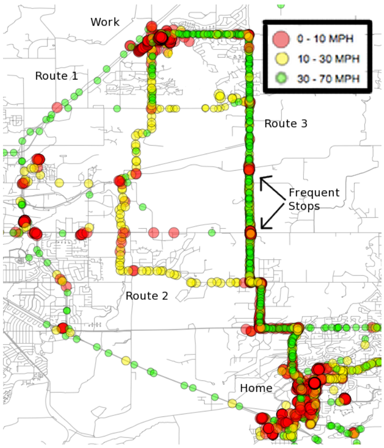
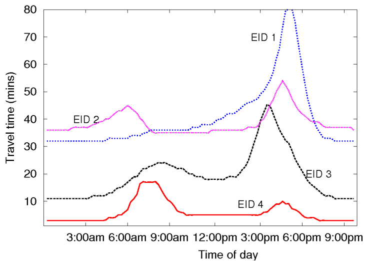
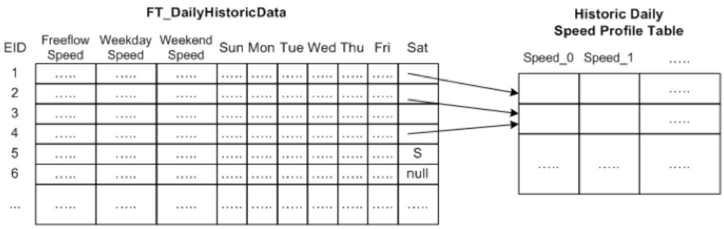


Fig. 5. A commuter's GPS tracks over three months reveal preferred routes. (Best in color).

Historical Speed Profiles: Traditionally, digital road maps consisted of centerlines and topologies of the road networks [16, 46]. These maps were used by navigation devices and web applications such as Google Maps [17] to suggest routes to users. New datasets from companies such as NAVTEQ [31] use probe vehicles and highway sensors (e.g., loop detectors) to compile travel time information across road segments for all times of the day and week at fine temporal resolutions (seconds or minutes). This data is applied to a profile model, and patterns in the road speeds are identified throughout the day. The profiles have data for every five minutes, which can then be applied to the road segment, building up an accurate picture of speeds based on historical data. Such temporally detailed (TD) roadmaps contain much more speed information than traditional roadmaps. Traditional roadmaps (Figure 2(a)) have only one scalar value of speed for any given road segment. In contrast, TD roadmaps may list speed/travel time for a road segment for thousands of time points (Figure 6) in a typical week. This allows a commuter to compare alternate start-times in addition to alternative routes. It may even allow comparison of (start-time, route) combinations to select distinct preferred routes and distinct start-times. For example, route ranking may differ across rush hour and non-rush hour and in general across different start times. However, TD roadmaps are big and their size may exceed 10^{13} items per year for the 100 million road-segments in the US when associated with per-minute values for speed or travel-time. Thus, industry is using speed-profiles, a lossy compression based on the idea of a typical day of a week, as illustrated in Figure 6(a), where each (road-segment, day of the week) pair is associated with a time-series of speed values for each hour of the day.



(a)



(b)

Fig. 6. Spatial Big Data on Historical Speed Profiles. (a) Travel time along four road segments over a day. (b) Schema for Daily Historic Speed Data. (Best in color).

In the near future, values for the travel time of a given edge and start time will be a distribution instead of scalar. For example, analysis of GPS tracks may show that travel-time for a road-segment is not unique, even for a given start-time of a typical week. Instead, it may consist of different values (e.g., 1, 2, 3 units), with associated frequencies (e.g., 10, 30, 20). Emergence of such SBD may allow comparison of routes, start-times and (route, start-time) combinations for statistical distribution criteria such as mean and variance. We also envision richer temporal detail on many preference functions such as fuel cost. Other emerging datasets include those related to pot-holes [35], crime reports [37], and social media reports of events on road networks [56].

4 Metrics For Spatial Big Data Benchmarks

Metrics for spatial big data can be categorized via a classification used in software engineering into functional (specific behaviors) and non-functional requirements (overall operation of a system). In this section, we will describe each and provide examples.

Metrics for Functional Spatial Big Data Requirements: Spatial computing traditionally operates on one of the four data types listed in Table 2:

Table 2. Data Types in Spatial Computing

Data Type	Representation	Operations	Potential Metric
Raster (field)	Geo-Matrix	Map algebra operations on Local, Focal, Zonal regions	Map algebra operations per second
Vector (object)	Points, Lines, Polygons	Intersection Model, Nearest Neighbor, Point Query, Range Query, etc.	Nearest Neighbors per second, Range-query (screen paint) per second
Network	Graphs (nodes, edges)	Shortest Path, Max Flow, Evacuation, etc.	Shortest-Paths per second
Spatio-Temporal	Trajectories, Temporal Networks	Time-dependent shortest path, GPS tracking, logging, etc.	Mobile device interactions per second

SBD-R: Raster datasets are frequently used for remote sensing applications, where large-scale map algebra and matrix operations are used. A helpful performance metric (e.g., map algebra operations per second) would measure how quickly representative operations of this type could be performed on a variety of dataset scales (e.g. terabyte, petabyte, exabyte, etc.).

SBD-V: Processing vector datasets in spatial database systems has historically been computationally expensive, with many key features (e.g., nearest neighbor queries) not being provided with the system. As newer systems are developed with these features, performance metrics measuring how quickly range queries and nearest neighbor queries can be computed are needed. Representative metrics include: nearest neighbors per second and range-queries per second.

SBD-N: Mapping services such as Google Maps has demonstrated the popularity of network-based datasets for use-cases such as personal transportation routing. It is not hard to imagine Google has a measure for how many shortest-paths per second it can calculate as it is serving the world driving directions, but universal and public benchmarks in this field will allow comparison between current spatial database systems. Representative metrics include: shortest-paths per second and evacuation planning.

SBD-ST: Spatio-temporal datasets are becoming more and more commonplace with the rise of location-based services and metrics for database systems rating their ability to handle some of these more common complex queries are crucial. For example, many applications currently request the location of a user, and potentially also monitor nearby points of interest to report back to the user.

So a metric that described the number of mobile device interactions (e.g., tracking, local context, location trigger, etc.) per second would be extremely useful for a variety of end-user applications. Representative metrics include: mobile device interactions per second, GPS logs per second, etc

Metrics for Non-Functional Spatial Big Data Requirements: Many Spatial Big Data use-cases (e.g., emergency services like E911, disaster response, etc.) typically require fault tolerance, where it should be resilient against natural calamities such as earthquakes, hurricanes, etc. Such requirements necessitate "triple-continental redundancy" [3], where the data is replicated on servers spread across multiple continents. This requirement poses several challenges for current cloud-based storage technologies due to performance issues inherent with wide-area replication and access. A potential metric for disaster resilience is a resilience footprint (e.g., 100 mile resilient, 1,000 mile, 10,000 mile), which may indicate the geo-spatial footprint of the disaster (e.g., fire, flood, tornado, earthquake, hurricane) that will not disrupt service.

Privacy of geographic information is an important and timely challenge due to personal information in GPS tracks, Check-in's, tweets, etc. While location information (GPS in phones and cars) can provide great value to users and industry, streams of such data also introduce spooky privacy concerns of stalking and geo-slavery [11]. For example, Ushahidi is a non-profit tech company providing technology for citizen-based reporting used in many countries with controlling regimes where privacy and protecting the reporter is extremely important [55]. Computer science efforts at obfuscating location information to date have largely yielded negative results. Thus, many individuals hesitate to indulge in mobile commerce and citizen reporting due to concern about privacy of their locations, trajectories and other spatio-temporal personal information [25]. It may be premature to provide specific metrics. However, Spatial Big Data benchmarks and metrics are needed to address many questions such as the following: "whether people reasonably expect that their movements will be recorded and aggregated..."? [32]. How do we quantify location privacy in relation to its spatio-temporal precision of measurement? How can users easily understand and set privacy constraints on location information? How does quality of location-based service change with variations in obfuscation level? Crucial to widespread adoption will be comforting the public, where a easy-to-understand metric describing the loss of privacy given information surrendered (e.g., adversary information gain per piece data submitted) will help people understand and compare various services against their privacy concerns.

5 Conclusion

Increasingly, location-aware datasets are of a size, variety, and update rate that exceed the capability of spatial computing technologies. This paper addresses the emerging challenges posed by such datasets, which we call Spatial Big Data (SBD), specifically as they apply to mobility services (e.g., transportation and routing). SBD examples include trajectories of cell-phones and GPS devices, vehicle engine measurements, temporally detailed (TD) road maps, etc. SBD has the potential to transform society.

Current benchmarks for spatial computing remain limited to small data sizes and only a portion of current popular data types. New benchmarks need to be built around Spatial Big Datasets, incorporating all four data types (raster, vector, network, spatio-temporal), while covering a wide variety of use-cases from emergency management, location-based services, advanced routing services, etc. New performance metrics, both functional (e.g., mobile interactions per second) and non-functional (e.g., disaster resilience footprint), will facilitate comparison between new systems being created and promoted by various spatial computing vendors.

Acknowledgments. We would like to thank Eric Horvitz (Microsoft), the Computing Community Consortium (CCC), Hillol Kargupta (UMBC), Erik Hoel (ESRI), Oak Ridge National Labs (ORNL) and the US-DoD for their helpful comments and support. This work was supported by NSF (0713214, DGE-0504195), US-DoD and FAPESP (grants 2011/19284-3, 2012/11395-3).

References

1. American Transportation Research Institute (ATRI). Fpm congestion monitoring at 250 freight significant highway location: Final results of the 2010 performance assessment (2010), http://goo.gl/3cAjr
2. American Transportation Research Institute (ATRI). Atri and fhwa release bottleneck analysis of 100 freight significant highway locations (2010), http://goo.gl/C0NuD
3. Bauer, E., Adams, R., Eustace, D.: Beyond Redundancy: How Geographic Redundancy Can Improve Service Availability and Reliability of Computer-based Systems. Wiley-IEEE Press (2011)
4. Booth, J., Sistla, P., Wolfson, O., Cruz, I.: A data model for trip planning in multimodal transportation systems. In: Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, pp. 994–1005. ACM (2009)
5. Brown, A.: Transportation Energy Futures: Addressing Key Gaps and Providing Tools for Decision Makers. Technical report, National Renewable Energy Laboratory (2011)
6. Capps, G., Franzese, O., Knee, B., Lascurain, M., Otaduy, P.: Class-8 heavy truck duty cycle project final report. ORNL/TM-2008/122 (2008)
7. Chan, E.P.F., Zhang, J.: Efficient evaluation of static and dynamic optimal route queries. In: Mamoulis, N., Seidl, T., Pedersen, T.B., Torp, K., Assent, I. (eds.) SSTD 2009. LNCS, vol. 5644, pp. 386–391. Springer, Heidelberg (2009)
8. Chang, T.: Best routes selection in international intermodal networks. Computers & Operations Research 35(9), 2877–2891 (2008)
9. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms. MIT Press (2001)
10. Davis, S.C., Diegel, S.W., Boundy, R.G.: Transportation energy data book: Edition 28. Technical report, Oak Ridge National Laboratory (2010)
11. Dobson, J., Fisher, P.: Geoslavery. IEEE Technology and Society Magazine 22(1), 47–52 (2003)

12. Federal Highway Administration. Highway Statistics. HM-63, HM-64 (2008)
13. Frigioni, D., Ioffreda, M., Nanni, U., Pasqualone, G.: Experimental analysis of dynamic algorithms for the single. ACM Journal of Experimental Algorithmics (JEA) 3, 5 (1998)
14. Frigioni, D., Marchetti-Spaccamela, A., Nanni, U.: Semidynamic algorithms for maintaining single-source shortest path trees. Algorithmica 22(3), 250–274 (1998)
15. Garmin, http://www.garmin.com/us/
16. George, B., Shekhar, S.: Road maps, digital. In: Encyclopedia of GIS, pp. 967–972. Springer (2008)
17. Google Maps, http://maps.google.com
18. Gray, J.: Benchmark handbook: for database and transaction processing systems, 2nd edn. Morgan Kaufmann Publishers Inc. (1993)
19. Hoel, E.G., Heng, W.-L., Honeycutt, D.: High performance multimodal networks. In: Medeiros, C.B., Egenhofer, M., Bertino, E. (eds.) SSTD 2005. LNCS, vol. 3633, pp. 308–327. Springer, Heidelberg (2005)
20. Jagadeesh, G., Srikanthan, T., Quek, K.: Heuristic techniques for accelerating hierarchical routing on road networks. IEEE Transactions on Intelligent Transportation Systems 3(4), 301–309 (2002)
21. Jing, N., Huang, Y.-W., Rundensteiner, E.A.: Hierarchical optimization of optimal path finding for transportation applications. In: Proceedings of the Fifth International Conference on Information and Knowledge Management (CIKM), pp. 261–268. ACM (1996)
22. Kargupta, H., Gama, J., Fan, W.: The next generation of transportation systems, greenhouse emissions, and data mining. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1209–1212. ACM (2010)
23. Kargupta, H., Puttagunta, V., Klein, M., Sarkar, K.: On-board vehicle data stream monitoring using minefleet and fast resource constrained monitoring of correlation matrices. New Generation Computing 25(1), 5–32 (2006)
24. Kleinberg, J., Tardos, E.: Algorithm Design. Pearson Education (2009)
25. Krumm, J.: A survey of computational location privacy. Personal and Ubiquitous Computing 13(6), 391–399 (2009)
26. Lovell, J.: Left-hand-turn elimination, December 9. New York Times (2007), http://goo.gl/3bkPb
27. Lynx GIS, http://www.lynxgis.com/
28. Mabrouk, M., Bychowski, T., Niedzwiadek, H., Bishr, Y., Gaillet, J., Crisp, N., Wilbrink, W., Horhammer, M., Roy, G., Margoulis, S.: Opengis location services (openls): Core services. OGC Implementation Specification 5, 016 (2005)
29. Manyika, J., et al.: Big data: The next frontier for innovation, competition and productivity. McKinsey Global Institute (May 2011)
30. MasterNaut. Green Solutions, http://www.masternaut.co.uk/carbon-calculator/
31. NAVTEQ, www.navteq.com
32. New York Times. Justices Say GPS Tracker Violated Privacy Rights (2011), http://www.nytimes.com/2012/01/24/us/police-use-of-gps-is-ruled-unconstitutional.html
33. OpenStreetMap, http://www.openstreetmap.org/
34. Potamias, M., Bonchi, F., Castillo, C., Gionis, A.: Fast shortest path distance estimation in large networks. In: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, pp. 867–876 (2009)

35. Pothole Info. Citizen pothole reporting via phone apps take off, but can street maintenance departments keep up? (2011), http://goo.gl/cGl3B
36. Ray, S., Simion, B., Brown, A.D.: Jackpine: A benchmark to evaluate spatial database performance. In: 2011 IEEE 27th International Conference on Data Engineering (ICDE), pp. 1139–1150. IEEE (2011)
37. SafeRoadMaps. Envisioning Safer Roads, http://saferoadmaps.org/
38. Samet, H., Sankaranarayanan, J., Alborzi, H.: Scalable network distance browsing in spatial databases. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, SIGMOD 2008, pp. 43–54 (2008)
39. Sanders, P., Schultes, D.: Engineering fast route planning algorithms. In: Demetrescu, C. (ed.) WEA 2007. LNCS, vol. 4525, pp. 23–36. Springer, Heidelberg (2007)
40. Sankaranarayanan, J., Samet, H.: Query processing using distance oracles for spatial networks. IEEE Transactions on Knowledge and Data Engineering 22(8), 1158–1175 (2010)
41. Schiller, J., Voisard, A.: Location-based services. Morgan Kaufmann (2004)
42. Shekhar, S., Evans, M.R., Kang, J.M., Mohan, P.: Identifying patterns in spatial information: A survey of methods. Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery 1(3), 193–214 (2011)
43. Shekhar, S., Fetterer, A., Goyal, B.: Materialization trade-offs in hierarchical shortest path algorithms. In: Scholl, M., Voisard, A. (eds.) SSD 1997. LNCS, vol. 1262, pp. 94–111. Springer, Heidelberg (1997)
44. Shekhar, S., Kohli, A., Coyle, M.: Path computation algorithms for advanced traveller information system (atis). In: Proceedings of the Ninth International Conference on Data Engineering, Vienna, Austria, April 19-23, pp. 31–39. IEEE Computer Society (1993)
45. Shekhar, S., Vatsavai, R.R., Ma, X., Yoo, J.S.: Navigation systems: A spatial database perspective. In: Location-Based Services, pp. 41–82. Morgan Kaufmann (2004)
46. Shekhar, S., Xiong, H.: Encyclopedia of GIS. Springer Publishing Company, Incorporated (2007)
47. Shrank, D., Lomax, T., Eisele, B.: The 2011 urban mobility report. Texas Transportation Institute (2011)
48. Sperling, D., Gordon, D.: Two billion cars. Oxford University Press (2009)
49. Stonebraker, M., Frew, J., Gardels, K., Meredith, J.: The sequoia 2000 storage benchmark. ACM SIGMOD Record 22, 2–11 (1993)
50. TeleNav, http://www.telenav.com/
51. TeloGIS, http://www.telogis.com/
52. Tomlin, C.D.: Geographic information systems and cartographic modeling. Prentice Hall (1990)
53. TomTom. TomTom GPS Navigation (2011), http://www.tomtom.com/
54. U.S. Energy Information Adminstration. Monthly Energy Review (June 2011), http://www.eia.gov/totalenergy/data/monthly/
55. Ushahidi, http://www.ushahidi.com
56. Waze Mobile, http://www.waze.com/
57. Wikipedia. Usage-based insurance — wikipedia, the free encyclopedia (2011), http://goo.gl/NqJE5 (accessed December 15, 2011)
58. Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L.: Discovering personal gazetteers: an interactive clustering approach. In: Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems, pp. 266–273. ACM (2004)