Comparing Shortest Paths Lengths of Free and Proprietary Data for Effective Pedestrian Routing in Street Networks
Revised version submitted: March 5, 2012
Word Count: 5586 words plus 6 figures/tables at 250 words each = 7086 words

Dennis Zielstra
University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314
Phone: (954) 577-6392
E-mail: dzielstra@ufl.edu

Hartwig H. Hochmair
University of Florida, Geomatics Program
3205 College Avenue
Ft. Lauderdale, FL-33314
Phone: (954) 577-6317
Fax: (954) 475-4125
E-mail: hhhochmair@ufl.edu

ABSTRACT

Ubiquitous mobile devices, such as smartphones, led to an increased popularity of pedestrian related routing applications over the past few years. Since pedestrians typically aim to minimize their walking distance, especially within non-recreational and multi-modal trips, pedestrian routing systems will only then be fully utilized if they are able to find the correct shortest path and thus help to avoid unnecessary detours. GPS-based car navigation systems have become standard already several years ago, which has lead to the availability of accurate street network data for car-based routing applications. However, pedestrian routing applications should consider pedestrian related network segments besides those utilized by motorized traffic, including footpaths, or pedestrian bridges. This paper performs a shortest path analysis of pedestrian routes for cities in Germany and the US. More specifically, we compare for a set of 1000 randomly generated origin-destination pairs the lengths of pedestrian routes that are computed based on different freely available network sources, such as OpenStreetMap (OSM) and TIGER/Line data, and proprietary datasets, such as TomTom, NAVTEQ, and ATKIS. The results show that freely available data sources such as OpenStreetMap provide a relatively comprehensive option for cities where commercial pedestrian data sets are not yet available.

INTRODUCTION

When walking from a trip origin to a destination, a pedestrian typically tries to minimize a variety of costs (disutilities), including distance, travel time, obstacles, turns, danger, slope, complex intersections, or uncertainties along the way. There exists a large body of research that reflects and analyses the different criteria a pedestrian may consider when choosing a route from among a variety of path alternatives (24, 9, 14, 17, 7, 21), how the importance of these criteria changes with the trip purpose (2), and how criteria are applied in combination with other transportation modes, i.e., within a multi-modal trip (20, 26, 4).

In multi-modal trips, travelers aim to minimize the distance and time of the walking portion of their trip, where the observed average walking distances are affected by demographic variables (income, education), characteristics of the pedestrian environment, such as topography and safety of intersections, or population density, among others (20, 27). A high percentage of pedestrian trips within the US is below one or 1/2 mile long and some increase in walking popularity has been reported (13). Walking is also the dominant mode to access public transit for short access distances and therefore correct modeling and consideration of the most sophisticated datasets for pedestrians are important. The results of an Onboard Transit Survey by the Atlanta Regional Commission clearly showed that most of the respondents were walking to the transit station (72.4%) with a distance less than 1/8 mile (53%) (8). Another study from the Netherlands showed that for rail transport walking is the dominant mode of access up to about 1.2 km (0.75mi) at the home end, and up to 2.2 km (1.4 mi) at the activity end (22).

The relatively new development of mobile devices with GPS and Wi-Fi positioning capabilities allows now to run routing applications and other location-based services on these small devices (1, 19). To be useful for pedestrian navigation these services need to be built on graph network structures with the required spatial granularity and detail. This is because compared to car drivers, pedestrians use different connections and they are not bound to lanes or to any other car navigation restriction (6). Shortest (or fastest) path is one of the most prominent route selection criteria in pedestrian navigation (9). It is oftentimes the default search mode in routing applications, besides alternative routes that are computed based on other optimization criteria (15). It is therefore of relevance to determine which commonly available street network data provide a useful basis for pedestrian routing. In this paper we analyze a variety of freely available and proprietary network datasets regarding shortest path lengths for two US cities (Miami and San Francisco) and two German cities (Berlin and Munich). The availability of pedestrian segments and their effect on shortest route generation will be analyzed for OSM, ATKIS (German government data), TIGER/Line, TomTom (formerly known as TeleAtlas) and NAVTEQ street datasets.

Since the increase of complementary geospatial data sets available, the integration of multiple network datasets has been a major research focus among transportation engineers, geographers, and computer scientists. The datasets often originate from different sources and vary in spatial accuracy, level of generalization, and attributes, which requires different approaches on the combination (conflation) of datasets (25, 3, 5) to facilitate their combined use in transportation applications. However, this paper does not consider conflation techniques and will focus on the analysis of the individual potential of widely used data sources on shortest path generation for pedestrians.

The remainder of the paper is structured as follows: The following section gives an introduction to the commercial and free street datasets used in this study and reviews previous findings on geodata comparison tests in the US and Germany. The next section provides

information about the applied analysis methods, followed by the results on the comparison of differences in path lengths found between different datasets. This is followed by a discussion of the results and aspects of future work.

5   **VOLUNTEERED GEOGRAPHIC INFORMATION AND PROPRIETARY STREET DATASETS**

Previous Work
The high demand for freely available spatial data within recent years has boosted the availability
10   of Volunteered Geographic Information (VGI) (10) on the Internet. The development of Web 2.0, the Global Positioning System (GPS) and its integration into cell phones, photo cameras, and other mobile devices, allows the Web community to interact with each other, provide information to central sites, and thus become a significant source of geographic information. The OpenStreetMap (OSM) project (available at openstreetmap.org), has the goal to create a detailed
15   map of the world based on VGI in vector data format. The information is collected by many participants, collated on a central database and distributed in multiple digital formats through the World Wide Web. It provides free downloads of spatial data from a large selection of themes, including roads, transit lines, tourist sites, or land use layers. Whereas most commercial providers of street data focus primarily on car navigation, OSM data also include pedestrian and
20   bicycle paths, which makes them a potential rich data source used for off-street routing. Within recent years the collection of VGI around the world created a fast growing selection of free street data. In connection with this development the quality of geodata and especially VGI has become a major research area. The quality of OSM data for England, Germany and parts of Florida, in particular the aspects of completeness, positional accuracy, and attribute accuracy, have been
25   analyzed in recent years (12, 29, 28).

For Germany, a comparison of coverage between TomTom and OSM data conducted for several periods during 2009 revealed the largest discrepancies for car navigation related data with total lengths being 43% smaller for OSM compared to TomTom datasets. This indicates that TomTom at that time specialized in car related data. For pedestrian navigation related data
30   the total length differences reduced from 27% to almost 10% over time, providing further evidence for the assumption that OSM specializes on smaller streets and alleys. The level of completeness of OSM data reveals some spatial pattern, which is a clear decrease in the completeness of the OSM dataset away from the metropolitan areas towards the surrounding rural areas. However, within more densely populated areas, OSM data offer more overall data
35   than the commercial provider (29).

Additional proprietary datasets for Germany can be found at the German Federal Office of Cartography and Geodesy and their corresponding regional surveyor's offices. ATKIS (Amtliches Topographisch-Kartographisches Informationssystem) includes several different formats and datasets for entire Germany; however charges will apply which depend on the size
40   of the area and amount of requested datasets. ATKIS data has not been included in the before mentioned analyses, thus there is no additional information available about comparison results. For our pedestrian specific analysis one ATKIS dataset has been added for the city of Berlin in this paper.

The results of the comparison of coverage between NAVTEQ, TomTom and OSM in the
45   United States, in particular the state of Florida, showed a different pattern. As opposed to the OSM coverage pattern observed for England and Germany, OSM data coverage is generally

higher in rural areas than for urban areas, when compared to commercial datasets. The good coverage results for OSM in the rural areas are not a result of user contribution but primarily based on the TIGER/Line import, since TIGER/Line contains more data especially in agricultural areas (small ways) than the commercial data providers. Despite the lower coverage
5   of OSM in urban areas compared to commercial datasets, comparison between OSM and TIGER data reveal that for some urban areas OSM data are actively collected by the Web community (29).

Since the computation of shortest paths for pedestrian routing is based on the underlying street network, it is of interest to compare the suitability of various data sets for shortest path
10   computations. It needs to be noted that we did not check any of the datasets for actual ground truth, since test areas are too large to do so. However, the study provides a relative comparison of shortest path lengths for different datasets as a proxy for data completeness. Most of the tested datasets are currently used in routing applications and available to users and decision makers.

15   **CASE STUDY**
For this study we computed shortest path distances of sample routes for four cities based on different street datasets. Different datasets were available for different cities (Error! Reference source not found.). From the commercial datasets, NAVTEQ Discover Cities data were available for Miami, while we had access to TomTom data for both US and German cities. In terms of free
20   data OSM network data were available for all cities, while TIGER/Line is a US data standard and therefore only available for the two US cities. Its counterpart for Germany, the ATKIS dataset, was provided to us for Berlin.

While the test area of each German city was determined through official city boundaries included in the TomTom dataset, for the US cities, county boundaries for Miami-Dade County
25   and San Francisco County were used instead. Test area sizes varied from 121 km^2 for San Francisco to 5040 km^2 for Miami and from 892 km^2 for Berlin to 310 km^2 for Munich. Three of the four tested cities (Miami, Munich and Berlin) show a wide spread street pattern which is scattered through parks and rivers. San Francisco shows a more compact street pattern because of the limited city area size. Two major parks are located in the western part of the city. The
30   public transportation network in cities can also influence the construction of pedestrian friendly infrastructure and pedestrian travel behavior in many ways. While the German cities and San Francisco show a sophisticated public transportation network, Miami is lacking this kind of system when moving away from the CBD area.

Extraction of Pedestrian Related Data
35   To be able to compute shortest routes for pedestrians, all network datasets needed to be filtered, which means that segments that were only accessible to cars, such as interstate highways, were removed. Due to different attributes and road classification schemes found in different datasets, the extraction methods had to be adapted. For example, for NAVTEQ data a query over street attributes formulated as [Pedestrian Access] = 'Y' was used to extract network segments that are
40   either restricted to pedestrians only, or that can be accessed both by cars and pedestrians. For TIGER/Line, street segments that prohibit pedestrian access were eliminated by the street name. OSM allows filtering the street network through specific tags that indicate accessibility for pedestrians. OpenRouteService, a routing application that runs on OSM data (available at openrouteservice.org), provides a wiki with a comprehensive list of street tags that can be used

for this purpose. TomTom and ATKIS provide a documented list of street attributes that can be used to filter pedestrian accessible network segments.

Shortest Route Generation

5   Before each dataset could be tested for shortest path lengths, random points were generated for each urban area. These were then used as potential start and end points of shortest path routes. The potential position of each random point was restricted by a predefined buffer. This buffer was generated by creating five meter buffers around each street network (generated from the different data sources) and intersecting these buffers. This way it was guaranteed that with each
10   dataset a given start or endpoint would be snapped to the correct street segment.
        Next, a random set of 1000 origin-destination point pairs was extracted from the random points in a way that it resulted in network distances that approximate a typical distance decay function for walking trips (18). This way the random route set reflects typical pedestrian walking distances. A sample size of 1000 routes was chosen since it was found that they cover many parts
15   of the networks, both less dense areas such as parks, as well as built areas. FIGURE 1 shows a map with a sample of 1000 pedestrian routes generated for Berlin. A larger number of routes would result in overlapping tracks and violate independence of observations which was presumed for further statistical analysis. Some origin-destination point combinations caused a situation of non-connectivity, e.g. start point on island and end point on mainland, where no
20   shortest route could be generated. Anytime this error occurred, the paired points and corresponding routes were removed, which reduced the total number of origin-destination point pairs to 996 for Miami, 993 for San Francisco, 978 for Munich and 977 for Berlin. ESRI's ArcGIS 10 Geographic Information System was used for all analysis procedures in this paper.
        If noticeable differences in shortest path length can be found for a given point pair and
25   two different underlying datasets, this can be attributed to two types of error, i.e., error of omission and error of commission (FIGURE 2). There are several ways to identify these errors, such as use of aerial photographs or street view images. An error of omission occurs if the dataset is missing a segment that is crucial for computing a shortest route, while this segment exists in the real world. FIGURE 2a gives an example for an error of omission found in the
30   Berlin network. TomTom, shown in dark blue color, is missing a segment in the street network, in this case a pedestrian bridge leading across a major highway. This causes a detour for the pedestrian of more than 1000 meters (0.6 miles) compared to the correct shortest path. As can be seen, OSM, shown in green, contains the pedestrian bridge in the dataset and thus provides the (correct) shortest route in this area. However, not each additional segment found in a network
35   dataset when compared to the remaining datasets can also be found in reality, and it would therefore falsely reflect a network shortcut. Such a situation leads to an error of commission, as shown in FIGURE 2b for Miami. TIGER/Line data in cyan indicate that there are several pedestrian paths running in north-south direction. However, a current satellite image in the background proves that these paths are nonexistent. TomTom, shown in red, provides a more
40   realistic dataset for this specific area. If data affected by this error of commission were not removed before the distance analysis, results would incorrectly show that TIGER/Line provides shorter shortest paths than other datasets in this area.

      While errors of commission were rarely found in any of the datasets ─ in our case only Miami
45   and Berlin showed any ─ they still needed to be identified during the analysis to give realistic and representative results. The focus of this paper was on additional shortcuts provided by any of

the data sets where errors of commission would be misleading and favor those datasets with errors of commission. Each time one of the datasets showed an error of commission during the analysis and in comparison to routes found for other datasets, the erroneous dataset was corrected, the network re-built, and the test repeated. Errors of omission, however, allow a realistic assessment of a shortest pedestrian route capability between different networks. They were not corrected before further analysis.

It must be noted that some online route planners, such as Google Maps, provide walking directions. However, these route planners use proprietary network datasets which cannot be accessed and downloaded for analysis. This restriction makes it impossible to compute buffers around street segments (which are necessary for start and end point generation). Further it complicates the analysis of errors of omission and commission. Pedestrian routes from online route planners were therefore not considered in this analysis, although it may be considered for future work.

RESULTS

This section starts with an overview of the total length differences between different networks. Total lengths are based on approximately 1000 calculated shortest routes for pedestrians found in the four urban area networks. Differences are then analyzed for their statistical significance.

The bar-charts in FIGURE show the total length of all calculated shortest routes based on different datasets used in each city. For the US cities differences between datasets were small. In Miami the largest difference between two data providers was observed between OSM and NAVTEQ with 33 km or 3% with an average route length of 1019 m for OSM and 1050 m for NAVTEQ. NAVTEQ showed the largest overall length value in this city, indicating that for the specific routes that were tested NAVTEQ provided the fewest shortcuts for pedestrians. For San Francisco OSM and TomTom showed the largest difference between two data providers with 16 km or 1.8% and an average route length of 864 m for OSM and 880 m for TomTom.

In German cities, the various datasets reveal larger differences in total shortest paths lengths than for US cities. While ATKIS, the purchasable federal dataset for Germany, provides significantly shorter pedestrian routes than the commercial TomTom dataset for Berlin, OSM provides significantly shorter paths than the other two datasets. Differences between OSM and TomTom are as large as 112 km or 11 % in Berlin with an average route length of 899 m for OSM and 1013 m for TomTom. A similar pattern could be observed for Munich, where only OSM and TomTom were available. Differences between OSM and TomTom are 76 km or 8% with an average route length of 893 m for OSM and 971 m for TomTom.

Previous research results showed large differences between OSM and TomTom in German cities with regard to the amount of pedestrian related data in the network, e.g. footpath shortcuts or pedestrian bridges (28). This leads to shorter pedestrian routes, as shown in FIGURE c and FIGURE 3d and explains the observed difference in path lengths. For US cities previous analysis revealed only relatively small differences in pedestrian related segments (28), which causes smaller differences in shortest path lengths. However, the results in our analysis showed that OSM, although not as much as in Germany, still provides the most complete free data source for pedestrian routes in both US cities as shown in FIGURE 3a and FIGURE 3b.

Statistical Analysis

A series of nonparametric Wilcoxon matched-pair signed rank tests were conducted to determine where the differences in median distances of shortest paths between different datasets were

significant. A Student's t-test could not be used because route lengths were not normally distributed. FIGURE a shows a Normal quantile plot (Q-Q plot) for one of the sample datasets which clearly indicates that the data is not normally distributed. Also, a logarithmic transformation on the route lengths did not lead to a normal distribution, but to a bi-modal
5   distribution (FIGURE 4b), which makes the use of a t-test inappropriate.

Since more than two datasets were used and compared for each city, a Bonferroni correction was applied to the test. The Bonferroni correction is a common method to handle problems caused by multiple comparisons. Applying a test multiple times will cause the probability of rejecting a true null hypothesis (i.e., a Type I error) to be no longer equal to α (a significance level associated
10   with the individual test). To retain an overall Type I error probability of α, a critical value can be found by dividing α by the number of tests carried out (Bonferroni correction).
FIGURE 5 gives an overview of the statistical results. By applying the Bonferroni correction α reduces to 0.05/3=0.017 in the case of Berlin since three combinations were tested. The α value has been left unchanged for Munich (only one comparison), and been adapted to 0.008 for
15   Miami (six combinations) and 0.017 for San Francisco (three combinations). The significance values in FIGURE 5 relate to individual tests. If a shown significance level is below the Bonferroni-corrected α-value, as specified before for each city, the difference in median path distance can be considered significant at a 5% level of significance. A Plus (+) sign indicates that the street network listed on the row header leads to significantly longer shortest paths than the
20   street network denoted on the column header, while a Minus (−) sign indicates the opposite. For example, in San Francisco, the OSM based network results in significantly shorter routes than the TIGER/Line network, which is indicated by the Minus sign in the first row.

The statistical results indicate that in the case of Miami, where little OSM data has been
25   collected after the TIGER/Line import, none of the median distances are significantly different from one another. San Francisco, however, shows how active contribution to OSM can help to improve connectivity. More specifically statistical test results show that the median distances are significantly smaller for OSM than for TIGER/Line, which can be attributed to voluntary data collection efforts for OSM after the TIGER/Line import into the OSM database. However, length
30   differences between OSM and TomTom in San Francisco were not significant at a 5 percent level of significance.
        FIGURE c and FIGURE d indicate larger differences in total shortest path lengths between commercial and freely available datasets for German cities than for the US. For both German cities that were tested in our analysis the statistical results show significantly shorter
35   distances for OSM in comparison to the TomTom and ATKIS data sets.

DISCUSSION AND FUTURE WORK
The results of our analysis show that the length of shortest paths generated for pedestrians is affected by the completeness of the underlying dataset.
        Further it can be observed that some urban street networks do not provide a significant
40   number of shortcuts (e.g. pedestrian bridges) that facilitate shortest routes for pedestrians compared to the road network. In these areas, even if all existing pedestrian shortcuts were integrated in the network dataset, shortest path lengths would not differ much from shortest path lengths generated from a road network without pedestrian specific shortcuts. As opposed to this, especially for pedestrian oriented cities with plazas, pedestrian paths and alleys, such as Berlin
45   and Munich, the choice of the dataset greatly affects the length of shortest paths. OSM data

provide a free and relatively comprehensive alternative for cities where commercial pedestrian datasets are not yet available, especially if there are active participants contributing data, such as in Germany.

A combination of both, freely available and commercial datasets would possibly provide the best coverage. However, it is important to note that a combined dataset including OSM and other commercial datasets cannot be considered for implementation due to current licensing issues. OSM can be used freely under the terms of the Creative Commons Attribution-Share Alike 2.0 license which includes copying, distributing, adapting and transmitting the data. Commercial use of the data is not prohibited either, however, the user must attribute the work in the manner specified by OSM and, most importantly, the resulting work must be under the same or similar license to the one used in OSM.

The number of sample cities used in this study with commercial pedestrian street data is too small to make definite statements about shortest path differences between OSM and commercial pedestrian datasets. However, for all cities analyzed in the US and Germany, OSM data resulted in shorter shortest paths for pedestrians commercial data sets. This observation indicates that the OSM community, also in the US, is active and helps to develop a comprehensive network of pedestrian paths, as can be particularly seen in the San Francisco area.

For the individual traveler who uses a GPS enabled cell phone or handheld, inclusion of pedestrian related network segments will have the advantage of being provided with more realistic routes that avoid unnecessary detours compared to systems that use street bound data only. While this study measured relative completeness of different network datasets by shortest path distances, actual routing applications will usually allow a search for alternative routes that are optimized for criteria other than travel distance, including safe routes, scenic routes, or routes with fewest slopes. In this case, additional edge attributes or turn cost need to be considered in the routing algorithm (16). Nevertheless, even if an alternative route, such as scenic route, is requested by the user of a routing system, a more complete network will provide a better search result than a sparse one, simply by avoiding unnecessary detours. The integration of heterogeneous (free) datasets is still a challenge (23), and the assessment of VGI data quality, especially OSM data, is an ongoing issue of high importance for successful geo-applications (11, 12).

References

(*1*)   Bharat, R. and Minakakis, L. Evolution of Mobile Location-based Services. *Communication of the ACM*, Vol. *46*, No. 12, 2003, pp. 61-65.

(*2*)   Bovy, P. H. L. and Stern, E. *Route Choice: Wayfinding in Transport Networks*, Kluwer Academic, Dordrecht, 1990.

(*3*)   Ching-Chien, C., Knoblock, C. A., and Shahabi, C. Automatically Conflating Road Vector Data with Orthoimagery. *GeoInformatica*, Vol. 10, No. 4, 2006, pp. 495–530.

(*4*)   Chiu, D. K. W., Lee, O. K. F., Leung, H.-F., Au, E. W. K., and Wong, M. C. W. *A Multi-Modal Agent Based Mobile Route Advisory System for Public Transport Network.* 38th Hawaii International Conference on System Sciences, 2005.

(*5*)   Cobb, M., Chung, M. J., Foley, H., Petry, F., Shaw, K., and Miller, H. V. A Rule-based Approach for the Conflation of Attributed Vector Data. *GeoInformatica*, Vol. 2, No. 1, 1998, pp. 7-35.

(*6*)   Corona, B. and Winter, S. Datasets for Pedestrian Navigation Services. In J. Strobl, T. Blaschke and G. Griesebner (eds.), *Proceedings of the AGIT Symposium*, Wichmann, Heidelberg, 2001, pp. 84-89.

(*7*)   Daamen, W., Bovy, P. H. L., Hoogendoorn, S. P., and V. d. Reijt, A.. *Passenger Route Choice concerning Level Changes in Railway Stations*. Transportation Research Board - 84th Annual Meeting, Washington, D.C. Transportation Research Board of the National Academies, 2005.

(*8*)   ETC Institute 2009-2010 Regional On-Board Transit Survey. http://www.atlantaregional.com/transportation/travel-demand-model/on-board-transit-survey. Accessed Mar 1, 2012

(*9*)   Golledge, R. G. Path Selection and Route Preference in Human Navigation: A Progress Report. In A. U. Frank and W. Kuhn (eds.), *Conference on Spatial Information Theory (COSIT'95)*, *LNCS 988*, Springer, Berlin, 1995, pp. 207-222.

(*10*)   Goodchild, M. F. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0 (Editorial). *International Journal of Spatial Data Infrastructures Research (IJSDIR)*, Vol. 2, 2007, pp. 24-32.

(*11*)   Goodchild, M. F. Spatial accuracy 2.0. In J.-X. Zhang and M. F. Goodchild (eds.), *Spatial Uncertainty, Proceedings of the Eighth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Science*s, World Academic Union, Liverpool, 2008, pp. 1-7.

(*12*)   Haklay, M. How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets, *Environment and Planning B, Planning and Design*, Vol. 37, No. 4, 2010, pp. 682 – 703.

(*13*)   Ham, S. A., Macera, C. A., and Lindley, C. Trends in Walking for Transportation in United States, 1995 and 2001. *Preventing Chronic Disease, Public Health Research, Practive, and Policy*, Vol. 2, No. 4, 2005.

(*14*)   Hochmair, H. H. Towards a classification of route selection criteria for route planning tools. In P. Fisher (ed.), *Developments in Spatial Data Handling*, Springer, Berlin, 2004, pp. 481-492.

(*15*)   Hochmair, H. H. Optimal Route Selection with Route Planners: Results of a Desktop Usability Study. In H. Samet, M. Schneider and C. Shahabi (eds.), *Proceedings of the 15th ACM International Symposium on Geographic Information Systems*, ACM, New York, NY, 2007, pp. 312-315.

(*16*)   Hochmair, H. H. Grouping of Optimized Pedestrian Routes for Multi-Modal Route Planning: A Comparison of Two Cities. In L. Bernard, H. Pundt and A. Friis-Christensen (eds.), *The European Information Society - Taking Geoinformation Science One Step Further, Lecture Notes in Geoinformation and Cartography*, Springer, Berlin, 2008, pp. 339-358.

(*17*)   Hoogendoorn, S. and Bovy, P. Pedestrian Route-Choice and Activity Scheduling Theory and Models. *Transportation Research Part B*, Vol. 38, 2004, pp. 169-190.

(*18*)   Iacono, M., Krizek, K., and El-Geneidy, A. (2008). *Access to Destinations: How Close is Close Enough? Estimating Accurate Distance Decay Functions for Multiple Modes and Different Purposes* Minneapolis, MN. *www.cts.umn.edu/Publications/ResearchReports/pdfdownload.pl?id=916*. Accessed Mar 1, 2012

(*19*)   Kim, N., Lee, H. S., Oh, K. J., and Choi, J. Y. Context-aware mobile service for routing the fastest subway path. *Expert Systems with Applications*, Vol. 36, No. 2, 2009, pp. 3319-3326.

(*20*)   Loutzenheiser, D. R. Pedestrian Access to Transit: Model of Walk Trips and Their Design and Urban Form Determinants Around Bay Area Rapid Transit Stations. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1604, 1997, pp. 40-49.

(*21*)   Muraleetharan, T. and Hagiwara, T. *Overall Level-of-Service of the Urban Walking Environment and Its Influence on Pedestrian Route Choice Behavior: Analysis of Pedestrian Travel in Sapporo, Japan*. Transportation Research Board - 86th Annual Meeting, Washington, D.C. Transportation Research Board of the National Academies, 2007.

(*22*)   Rietveld, P. The accessibility of railway stations: the role of the bicycle in The Netherlands. *Transportation Research Part D*, Vol. 5, No. 1, 2000, pp. 71-75.

(*23*)   Scheider, S. and Schulz, D. Specifying Essential Features of Street Networks. In S. Winter (ed.), *Conference on Spatial Information Theory (COSIT'07), LNCS 4736*, Springer, Berlin, 2007, pp. 169-185.

(*24*)   Senevirante, P. N. and Morrall, J. F. Analysis of factors affecting the choice of route of pedestrians. *Transportaton Planning and Technology*, Vol. 10, 1986, pp. 147-159.

(*25*)   Song, W., Keller, J., Haithcoat, T., and Davis, C. Automated Geospatial Conflation of Vector Road Maps to High Resolution Imagery. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, Vol. 18, No. 2, 2009, pp. 388-400.

(*26*)   Wu, Q. and Hartley, J. Accommodating User Preferences in the Optimization of Public Transport Travel. *International Journal of Simulation Systems, Science & Technology: Applied Modelling & Simulation*, Vol. 5, No. 3-4, 2004, pp. 12-25.

(*27*)   Zhao, F., Chow, L.-F., Li, M.-T., Ubaka, I., and Gan, A. Forecasting transit walk accessibility: Regression model alternative to buffer. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1835, 2003, pp. 34-41.

(*28*)   Zielstra, D. and Hochmair, H. H. A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2117, 2011, pp. 145-153.

(*29*)   Zielstra, D. and Zipf, A. Quantitative Studies on the Data Quality of OpenStreetMap in Germany. In R. Weibel and S. Fabrikant (eds.), *Proceedings of the Sixth International Conference on Geographic Information Science*, GIScience, Zurich, Switzerland, 2010, pp. 20-26.

List of Figure Captions and Table Titles

TABLE 1 Datasets used for different cities.

FIGURE 1 Random pedestrian routes tested in Berlin
FIGURE 2 Error of omission in Berlin (a) and error of commission in Miami (b)
FIGURE 3 Total lengths of approximately 1000 tested pedestrian routes for selected US and German cities using different data sources
FIGURE 4 Characteristics of observed OpenStreetMap trip distances for Berlin

TABLE 1 Datasets used for different cities.

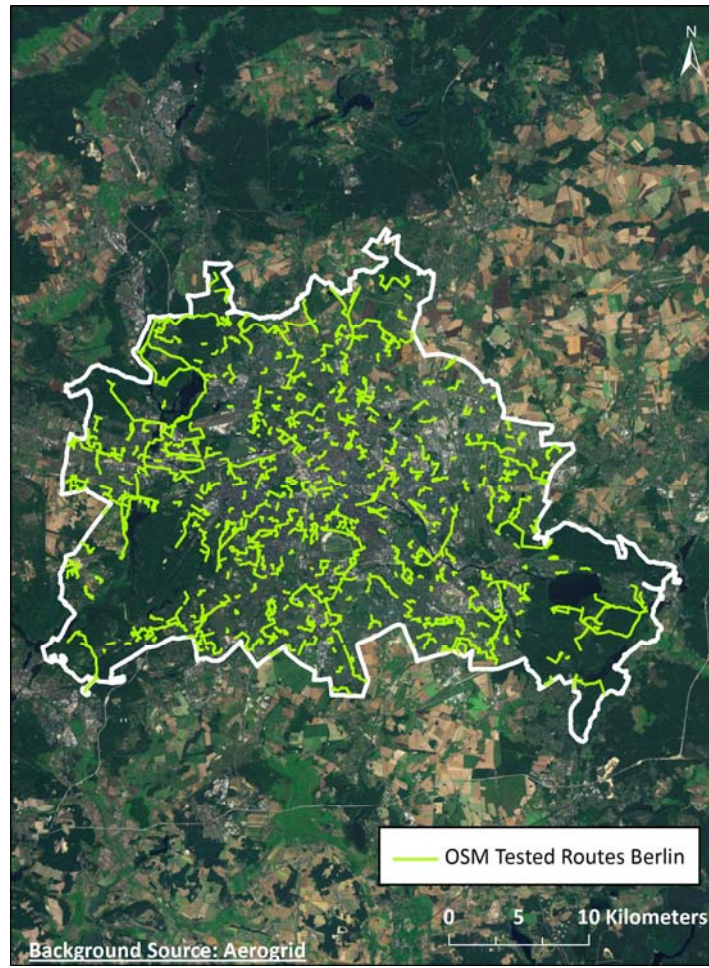|		Commercial		Free / Government	
		NAVTEQ	TomTom	OSM	TIGER/Line
USA	Miami	✓	✓	✓	✓
	San Francisco		✓	✓	✓
					ATKIS
Germany	Berlin		✓	✓	✓
	Munich		✓	✓	

FIGURE 1 Random pedestrian routes tested in Berlin

FIGURE 2 Error of omission in Berlin (a) and error of commission in Miami (b)

Total Length of Tested Pedestrian Routes in Miami

a)

Total Length of Tested Pedestrian Routes in San Francisco

b)

Total Length of Tested Pedestrian Routes in Berlin

c)

Total Length of Tested Pedestrian Routes in Munich

d)

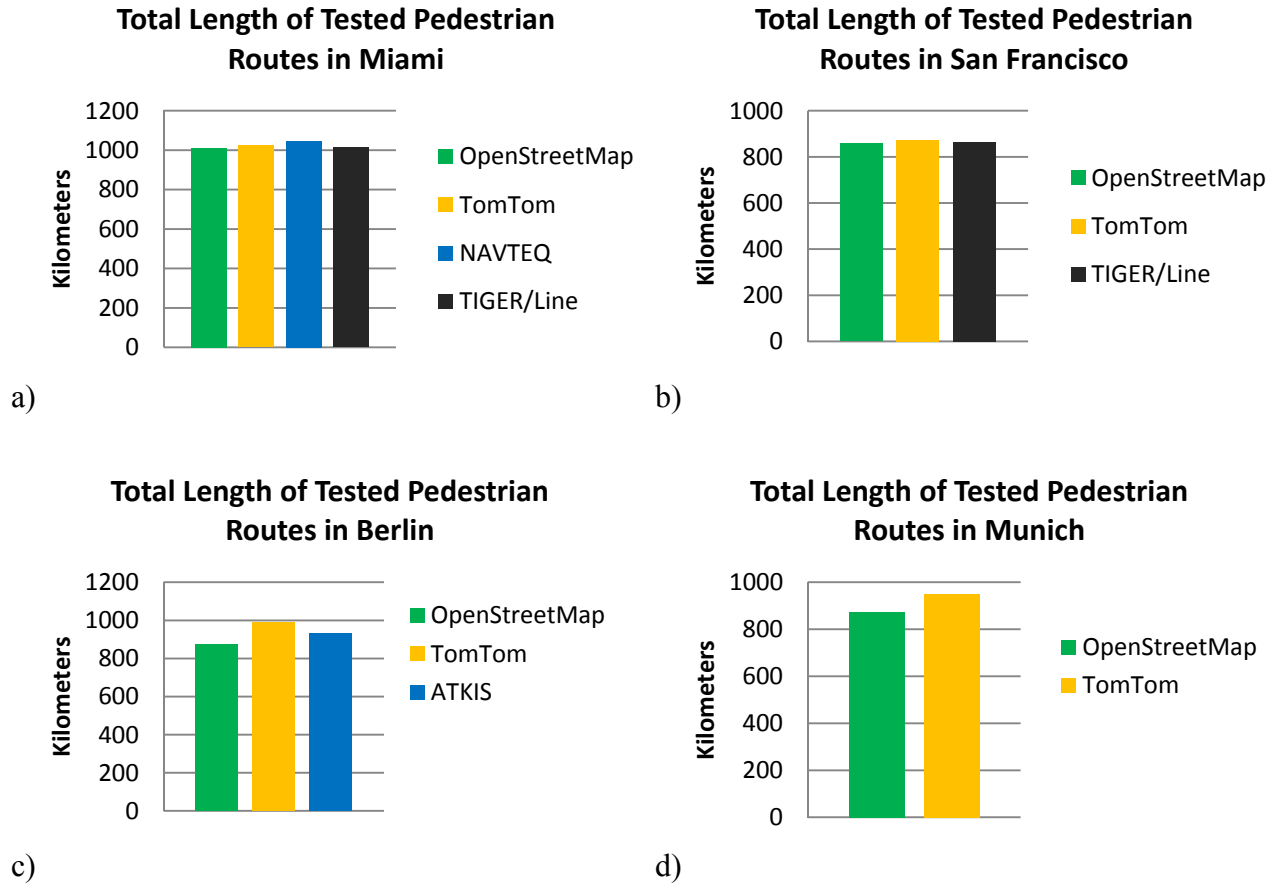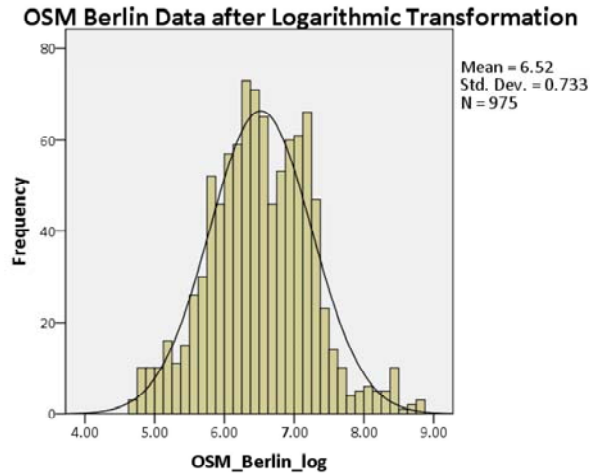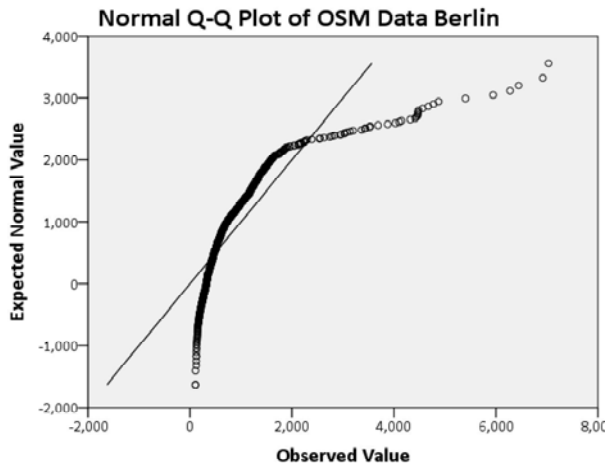**FIGURE 3 Total lengths of approximately 1000 tested pedestrian routes for selected US and German cities using different data sources**

a)   b)

FIGURE 4 Characteristics of observed OpenStreetMap trip distances for Berlin

Miami	OSM	TA	NAV	TIG
OSM		n.s.	n.s.	n.s.
σ(2-t.)		.024	.023	.725
TA	n.s.		n.s.	n.s.
σ(2-t.)	.024		.708	.064
NAV	n.s.	n.s.		n.s.
σ(2-t.)	.023	.708		.042
TIG	n.s.	n.s.	n.s.	
σ(2-t.)	.725	.064	.042	

San Francisco	OSM	TA	TIG
OSM		n.s.	—
σ(2-t.)		.067	.000
TA	n.s.		n.s.
σ(2-t.)	.067		.020
TIG	+	n.s.	
σ(2-t.)	.000	.020	

Berlin	OSM	TA	ATK
OSM		—	—
σ(2-t.)		.000	.000
TA	+		n.s.
σ(2-t.)	.000		.030
ATK	+	n.s.	
σ(2-t.)	.000	.030	

Munich	OSM	TA
OSM		—
σ(2-t.)		.000
TA	+	
σ(2-t.)	.000	

FIGURE 5 Results of Wilcoxon matched-pair signed rank tests after Bonferroni correction. OSM = OpenStreetMap, NAV = NAVTEQ, TA = TomTom, TIG = TIGER/Line, ATK = ATKIS