Geocoding Best Practices: Review of Eight Commonly Used Geocoding Systems

Jennifer N. Swift
Daniel W. Goldberg
John P. Wilson

Prepared for: Division of Cancer Prevention and Control
National Center For Chronic Disease Prevention and Health Promotion
Center for Disease Control and Prevention

Cover Photos:
Google Maps Street View, geocoded data in ArcGIS and USC Geocode Correction Service examples, by D. Goldberg and J. Swift (2008).

Preferred Citation:
Swift JN, Goldberg DW, and Wilson JP 2008 Geocoding Best Practices: Review of Eight Commonly Used Geocoding Systems. Los Angeles, CA, University of Southern California GIS Research Laboratory Technical Report No 10.

Table of Contents

List of Figures

List of Tables

Executive Summary

The purpose of this report is to provide a detailed account of current geocoding software to the Division of Cancer and Prevention and Control (DCPC), Centers for Disease Control and Prevention (CDC), based on a review of eight different commonly used proprietary and open source geocoding software packages available today. The packages tested consist of Centrus, Geolytics, ERSI Address Locator, Geocoder.us, Google Earth, Google Maps API, Yahoo API, and open source USC Geocoding Platforms. The reference dataset was first geocoded using the USC Geocoding Platform which incorporated parcel centroid matching accuracy, and then each match was verified and/or corrected using a combination of tools including Internet information searches, Google Earth and/or the Google Maps API. This is the third in a series of three reports which documents geocoding best practices for DCPC and CDC.

The different geocoding software packages were evaluated based on geocoding functionality, standardization, methodology and reference datasets utilized, and match rate. Many of these programs offer built-in functionalities as well as opportunities for customization that can be considered for inclusion in the design and coding of a standardized and centralized geocoder designed specifically to serve the needs of the cancer research community. The match rate was determined by comparing the results obtained with the eight software systems that were tested to the reference data set. The evaluation presented in this report also documented the requirements related to development and implementation, and the licensing and costs that are associated with the eight geocoding systems. The findings in this report should assist the CDC in their aim of developing of a standardized and centralized geocoder, freely available to the cancer research community.

1 Introduction

Numerous geocoding services and software exist today. Cancer registries utilize many different geocoding programs in order to carry out epidemiological research that incorporates geospatial mapping, as described in Goldberg et al. (2008a). This research report investigates the currently obtainable levels of geocoding accuracy by focusing on the methodologies and processes of geocoding that are associated with eight geocoding platforms that are frequently utilized by the general public as well as researchers.

At present, many proprietary and open source (free) geocoding systems are available to aid in geospatial mapping and support epidemiological research. The soundness of this research is highly dependent up the positional accuracy of geocoded addresses and the match rate (percentage of addresses geocoded) of the geocoding process. It is difficult for individual cancer registries to evaluate more than one or two of these options, assuming they wish to establish which software will best satisfy their priorities and produce the most accurate geocode output. Therefore, these researchers commonly rely on published comparisons found in the scientific literature. Accordingly, the present work is intended to update the results from related research conducted by Krieger et al. (2001, 2002a, 2002b), Whitsel et al. (2004), Yang et al. (2004), and Zhan et al. (2006), among others.

In order to determine what the currently available geocoding software has to offer and how the various functionalities of these programs might be utilized by the cancer research community, a comprehensive list of geocoding software has been assembled based on a review of all online and/or PC-based geocoders available today. The main goal of this research report is to present the results of an evaluation of eight different geocoding software packages available today: the Centrus US Street Point Database (Group 1 Software Inc. 2008a, 2008b), Geolytics GeocodeDVD (Geolytics 2008), ERSI Address Locator (ESRI 2008a), Geocoder.us (Locator Technologies 2008), Google Earth (Google 2008a, 2008b), Google Maps API (Google 2008c, 2008d), Yahoo API (Yahoo 2008), and the open source USC Geocoding Platform (Goldberg 2008b) (Table 1).

The objective in evaluating these particular geocoding applications is to provide a review of a variety of programs, requirements associated with development and implementation, and licensing and associated costs. Many of these programs offer built-in functionalities as well as opportunities for customization that can be considered for inclusion in the design and coding of a standardized and centralized geocoder designed specifically to serve the needs of the cancer research community. By identifying what software and respective functionality already exists versus what functionality remains to be developed, this report is intended to assist the Centers for Disease Control and Prevention (CDC) in their aim of developing of a standardized and centralized geocoder, freely available to the cancer research community.

2 Geocoders Evaluated

Presently there are many different commercial and open source geocoding options to choose from, many of which are available for immediate use via download or directly through the Internet (Appendix 1). For the purposes of this research, the most current and complete versions of eight different geocoding software packages available at the time of this writing were tested.

Table 1 Geocoding systems evaluated as part of this study, in alphabetical order

Name	Application	Commercial/Open Source	Coverage
Centrus US Street Point Data	PC-based	Commercial	United States (Worldwide cities)
ESRI Address Locator	PC-based	Commercial	Worldwide (User Defined)
Geocoder.us	Web-based	Commercial	United States
Geolytics GeocodeDVD	PC-based	Commercial	United States
Google Earth	Web-based	Commercial	Worldwide
Google Maps API	Web-based	Commercial	Worldwide
Yahoo Maps API	Web-based	Commercial	Worldwide
USC Geocoding Platform	PC/Web-based	Open Source	United States

A list of the geocoding software evaluated as part of this study is provided in Table 1. In each case, the same input datasets were geocoded in order to facilitate a comparison between the various software functionalities (Table 2). The examples included mixes of commercial/residential addresses picked at random from Internet searches for "Malibu CA 90265", "Long Beach, CA", "Altadena, CA, 91001", "Lancaster, CA", and "downtown "Los Angeles, CA". These locations were specifically chosen because of the distinct characteristics they possess. The Lancaster area was chosen because it epitomizes a classic grid style street pattern and because it contains especially difficult classes of addresses: single character street names, e.g. 123 K Street, as well as streets with special keywords transposed or as street names, e.g. 123 Avenue P. The Long Beach area was chosen because it represents a prototypical urban region, characterized by high density population resulting in large numbers of apartment buildings and complexes, and also because it contains many non-regular street segments. The Downtown Los Angeles area was chosen because it epitomizes a classic downtown with a high prevalence of large buildings taking up entire city blocks, truncated street segments, and a proliferation of one-way streets. The Malibu area was chosen because it is on the fringe of both urban and rural regions, possessing many of the characteristics of each. It has the many windy rural roads with just a few houses per road that one would expect to encounter in rural areas, as well as the more dense commercial regions found in an urban setting. The Altadena area was chosen because it is a prototypical suburban region consisting of primarily single family homes arranged in a regular pattern of streets. Additionally, it represents a class of population centers that are unincorporated regions of Los Angeles County, receiving services from both the county and other neighboring cities, e.g. policed by the LA County Sherriff's Department but contracting with the city of Pasadena for other services. This can potentially lead to incorrect city designated in addresses.

The software test evaluations and respective discussions are presented in alphabetical order within the relevant sections of this report. The compiled results in terms of geographic coordinates of the geocoded addresses (e.g. the geocoding results) are summarized in Appendix 2 It is recommended that additional similar datasets representative of other parts of the US be collected and tested accordingly in future phases of this research, in order to assess the extent to which the results may be extrapolated to other parts of the country. Prior to testing, a reference dataset was prepared by first geocoding the example addresses using the open source USC Geocoding Platform, since it implements parcel centroid matching accuracy (Goldberg 2008c). Then each match was verified and/or updated as appropriate using a combination of tools including Internet information searches, Google Earth and/or the Google Maps API, and the open source USC Geocoding Platform (Google 2008b). The reference dataset presented in Table 2 and Figure 1 is utilized as "ground truth" to facilitate comparison of the results generated through testing the eight geocoding packages evaluated as part of this study.

Table 2 Geocoded addresses utilized as the reference dataset in this study

ID	Address	Accuracy
1	23815 Stuart Ranch Road, Malibu, CA, 90265	Address range interpolation
2	3874 Cross Creek Rd, Malibu, CA, 90265	Address range interpolation
3	10936 Pacific View Dr, Malibu, CA 90265	Building centroid
4	21821 Castlewood Drive, Malibu, CA, 90265	Building centroid
5	5901 Filaree Heights Ave, Malibu, Ca 90265	Building centroid
6	22211 Pacific Coast Hwy, Malibu, CA, 90265	Building centroid
7	6902 Wildlife Rd, Malibu, CA, 90265	Building centroid
8	23732 Malibu Road Malibu, CA, 90265	Address range interpolation
9	6506 Westward Beach Rd Malibu, CA, 90265	Building centroid
10	3410 Civic Center Way, Malibu, CA, 90265	Street centroid
11	305 East Pacific Coast Highway, Long Beach, CA, 90806	Address range interpolation
12	1515 Hughes Way, Long Beach, CA, 90810	Building centroid
13	400 West Ocean Boulevard #1001, Long Beach, Ca	Address range interpolation
14	2400 E. Spring Street Long Beach, CA, 90806	Address range interpolation
15	1240 San Antonio Dr., Long Beach, CA, 90807	Address range interpolation
16	5021 East Anaheim Street, Long Beach, CA, 90804	Building centroid
17	3606 Arabella St., Long Beach, CA, 90805	Building centroid
18	4726 Boyar, Long Beach, CA, 90807	Address range interpolation
19	5915 E Seaside Walk, Long Beach,CA, 90803	Building centroid
20	5050 Linden Ave 107, Long Beach,CA, 90805	Exact parcel centroid
21	3055 Zane Grey Terrace, Altadena, CA, 91001	Building centroid
22	922 E Mendocino Street, Altadena, CA, 91001	Address range interpolation
23	617 Devirian Pl, Altadena, CA, 91001	Building centroid
24	2485 Lake Ave, Altadena, CA, 91001-2441	Nearest parcel centroid
25	2595 Fair Oaks Ave Altadena, CA, 91001	Building front door
26	2290 Country Club Drive, Altadena, CA, 91001	Building centroid
27	816 La Vina Ln, Altadena, CA, 91001-3754	Nearest parcel centroid
28	600 E. Mariposa St. , Altadena , CA, 91001	Address range interpolation
29	2851 Lake Avenue, Altadena Ca 91001	Building centroid
30	2333 N. Lake Ave. Unit H, Altadena, CA, 91001	Building centroid
31	44916 North 10th Street West, Lancaster, CA 93534	Building centroid
32	554 W. Lancaster Blvd., Lancaster, Ca 93534	Building front door
33	44750 60th Street West, Lancaster, CA, 93536-7620	Address range interpolation
34	43209 Crestwood Ct, Lancaster, CA, 93536	Address range interpolation
35	219 West Milling Street, Lancaster , Ca 93534	Address range interpolation
36	43625 N. Sierra Hwy. Suite A, Lancaster, CA, 93534	Address range interpolation
37	43436 16th Street West Lancaster, CA, 93534	Address range interpolation
38	1811 West Avenue J-12, Lancaster, CA, 93534	Building centroid
39	1309 W Ivesbrook St, Lancaster, Ca 93534	Address range interpolation
40	2025 W Ave K10, Lancaster, Ca 93536	Address range interpolation
41	801 East 4th Place, Los Angeles, CA, 90013	Building centroid
42	725 South Bixel Street Los Angeles, CA, 90017	Exact parcel centroid
43	333 South Figueroa Street Los Angeles, CA, 90071	Address range interpolation
44	740 S. Broadway Ave., Los Angeles, CA, 90014	Nearest parcel centroid
45	1850 Industrial St., Los Angeles, CA, 90021	Address range interpolation
46	11 South Hope Street, Los Angeles, CA, 90017	Street centroid
47	1130 S Flower St #420, Los Angeles, CA, 90015	Building centroid
48	930 Wilshire Blvd., Los Angeles, CA, 90017	Address range interpolation
49	333 S. Alameda Street, Los Angeles, CA, 90013-1735	Address range interpolation
50	318 E 2nd St, Los Angeles, CA, 90012	Address range interpolation

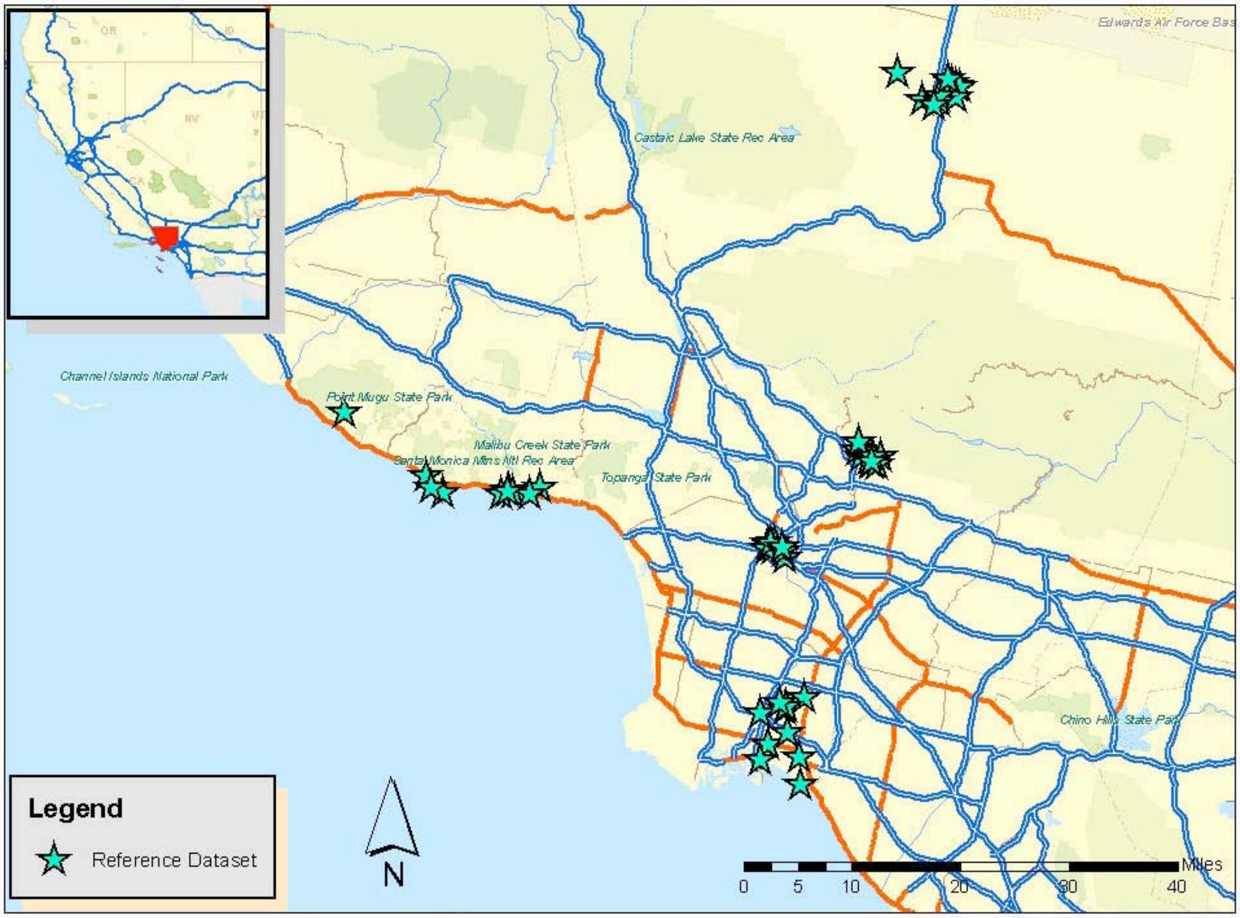


Figure 1 View of "true" locations for reference dataset used as listed in Table 2

Table 2 shows that the reference data were geocoded to four levels of accuracy. Building centroid accuracy means that the building associated with the address was unambiguously determined, and the location at the center of its roofline was recorded. Exact parcel centroid accuracy means that the parcel to which the address belonged could be unambiguously determined, but the specific building could not be determined. This is commonly the case for apartment complexes or other large parcels that contain more than one building. Nearest parcel accuracy means that although the exact parcel for an address could not be located, there was only a single option for a parcel in the same location leading to the conclusion that the address given was an alias for the parcel chosen. This is commonly the case when a parcel has been subdivided without the knowledge or consent of the property assessor's office, or in the case where multiple addresses share the same parcel. In either case, the United States Postal Service (USPS) effectively legitimizes the address because they deliver mail to the location using that address, be it officially recognized by the property assessor or not. In the case of address range interpolation accuracy, the geocode produced is the result of traditional linear interpolation where the output location is a proportional distance along the street segment equal to the proportion of address range represented by the address street number. These cases occur when the parcel for an address cannot be unambiguously determined. Most commonly this happens when the address does not exist in a property assessor parcel database (either rightfully or wrongfully) and it could potentially be associated with multiple nearest parcels, therefore being uncertain as to which parcel centroid to use. Street centroid accuracy occurs when the address street number does not correspond to the address range associated with a street segment, but the street can still be unambiguously determined. In these cases, the centroid of the street segment itself is used. These types of geocodes commonly occur when there is only a single street segment matching the other attributes of the address, but the address number associated with the input data is incorrect.

In this study, "positional accuracy" is defined as the differences in terms of linear (Euclidean) distance between the geocoded locations of the addresses provided in the reference dataset and the geocoded locations of the same addresses produced using the eight different software tools listed in Table 1. These distances were calculated using the traditional Great Circle Distance calculation (Weisstein 2008). The geocoders were compared by reviewing their respective methodologies, the proportion of addresses exactly matched versus partially matched or not at all relative to the reference dataset, and the variation in United States Census Bureau (USCB) Census Block groups in which the geocoded addresses fall, again with respect to the reference dataset.

Of the many web-based as well as PC-based geocoding programs that can be utilized in cancer research, each system has its own structure and workflow, functions it performs, and licensing and costs. First presented in Goldberg et al. (2008a), Figure 2 clearly illustrates the geocoding process, showing the critical components that should be common to any geocoder implementation (Goldberg 2008a; Goldberg et al. 2008b). Depending on specific users and/or developers, the implementation of these various geocoding programs can differ with respect to this conceptual representation of the geocoding process, as well as in terms of the various software components utilized. Each cancer registry and/or researcher will have unique geocoding user requirements, which are essentially the list of constraints that directly affect their choice of geocoding software functionalities (Goldberg et al. 2008a). Technical, budgetary, legal, and policy constraints will all influence the choice of geocoding software. Figure 2 shows the high-level decisions that may need to be made by individual cancer researchers, depending on their own requirements. This diagram also illustrates user requirements that geocoder vendors must take into account in order to develop useful geocoding services most appropriate for the cancer research community. Unfortunately many of the geocoding software listed in Appendix 1 and tested in this study do not provide information or metadata on their geocoding methodology at the level of detail shown in Figure 2.

2.1 PC-Based Applications

Centrus 2008 US Street Point Data

Centrus offers a suite of commercial geocoding software geared toward needs of organizations that require address-level geographic analysis (Group 1 Software Inc. 2008a). Centrus supports several levels of detail in geocoding databases which combine reference street networks from Tele Atlas (Tele Atlas 2008a, 2008b), NAVTEQ (NAVTEQ 2008), USCB Topographically Integrated Geographic Encoding and Referencing (TIGER)/Line files (United States Census Bureau 2008a, 2008b), and monthly updates based on USPS postal data (United States Postal Service 2008a). The most comprehensive standardization and geocoding tool offered by Centrus is the "Points Data Set", which also incorporates parcel centroid and elevation geocoding in the United States (US) based on parcel polygons acquired directly from counties. For the purposes of this study, the example address data in Table 2 were tested using the Centrus Points Data Set. The geocoder works by comparing the street address to be geocoded with records in the USPS ZIP+4 database (United States Postal Service 2008a), their enhanced street network files, and county property assessor parcel

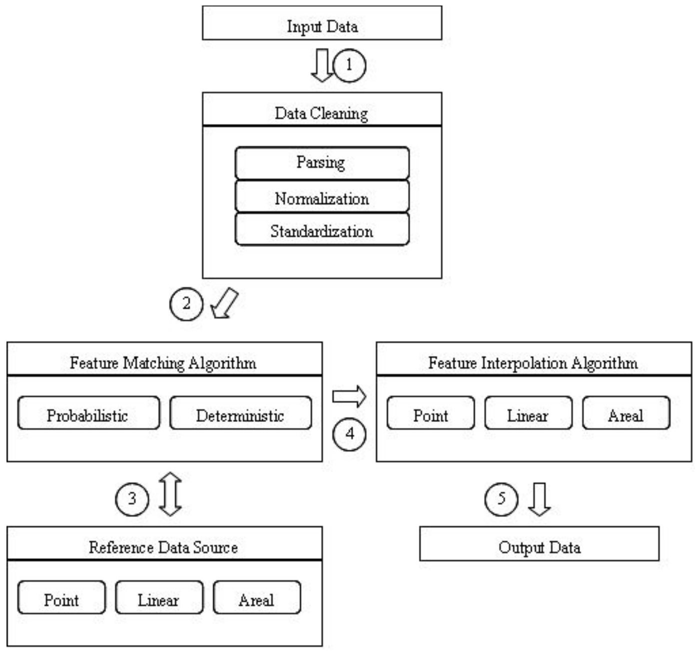
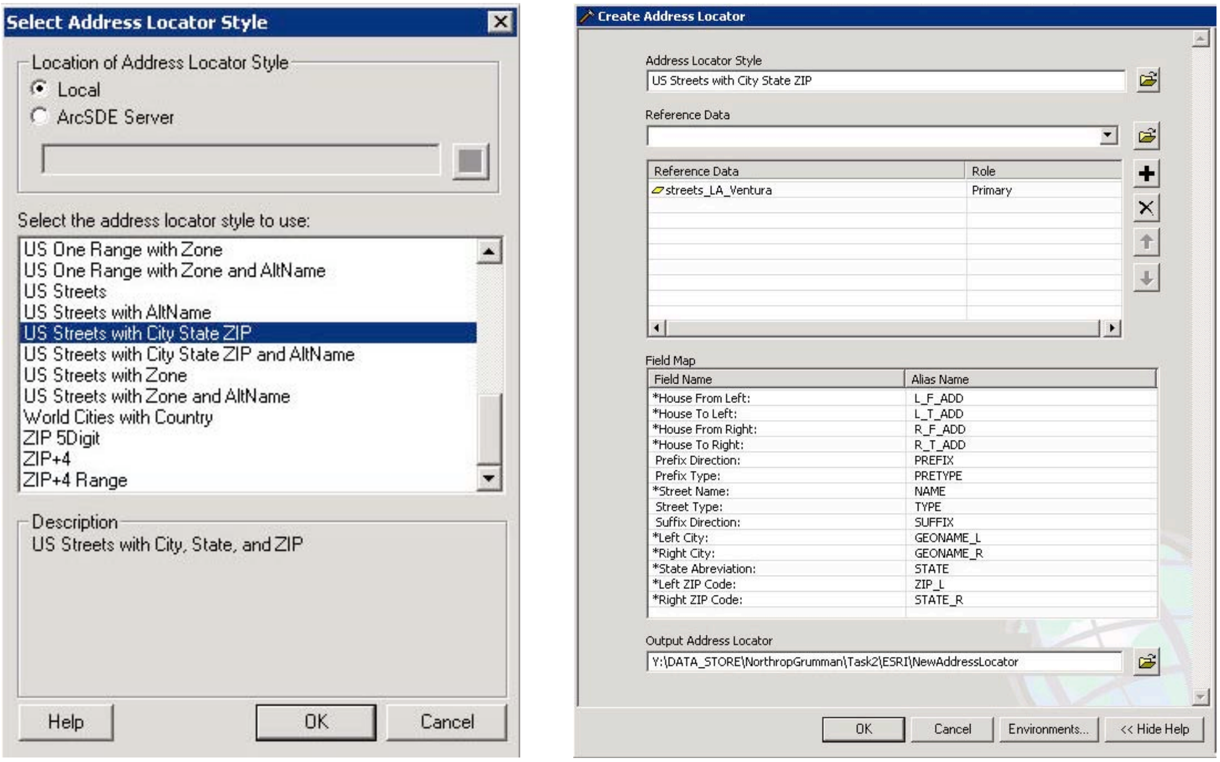


Figure 2 Generalized workflow of the geocoding process (Goldberg 2008a)

datasets (Group 1 Software Inc. 2008). In the output file, the geocoder provides a "match code" and a "location code" for each geocoded address. The alpha numeric match codes provide information about the standardization process, including whether or not a match was found, the resulting match type, and/or why a match was not found. The location codes contain details on the positional accuracy of the geocoded addresses, including the matched attribute of the input data (e.g. address, USPS ZIP code, city), the matched reference data feature type (i.e. street range, point level), the interpolation type (i.e. midpoint of street, center of parcel), and any errors (e.g. Census ID is uncertain).

ESRI Geocoding and Address Management

The ESRI Geocoding Addresses functionality was evaluated in ArcGIS 9.2, using ArcCatalog, ArcMap and ArcToolbox (ESRI 2008a). To utilize these tools, one first creates an ESRI "address locator" in ArcCatalog using one or more reference datasets chosen by the user and subsequently used to geocode a given set of addresses. The address locator consists of the user-generated and/or defined reference datasets including address attributes, indexes, and queries for geocoding, as well as the specification for how an address is standardized, the searching methods for possible matches, and an indication of which metadata is to be generated. Each address locator style is composed of several files that constitute the geocoding "rule base" (ESRI 2003). These files contain the match rule files that define how an address is matched against the reference candidates and how to standardize the address data into the desired format. Custom locator styles or locator styles provided by third parties can be used to define different roles for reference data feature classes and tables (Figure 3a and b and Figure 4). In addition, a composite address locator can be built that consists of a combination of more than one address locator, an implementation of the feature matching hierarchies defined in Goldberg (2008a) (Figure 3a). Search criteria can be set to filter addresses to specific address locators such that not all locators in the composite address locator will be used for finding the address.



a) Address Locator Style dialog box b) Create Address Locator dialog box

Figure 3 Specification of style (a) and reference dataset and attributes (b) in building an ESRI address locator (geocoder)

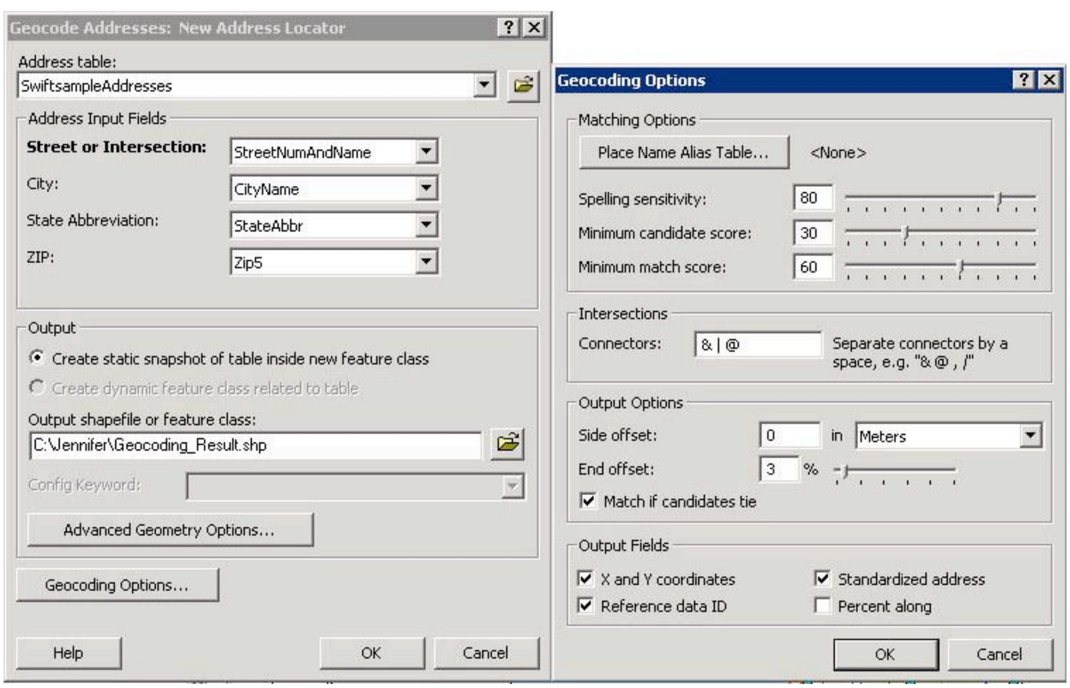


Figure 4. Geocode addresses dialog boxes for setting geocoding options

In effect, address locator "styles" can be seen as different combinations of reference datasets, matching rules, and other options (Figure 3a). For instance, the US Streets with Zone address locator style for street centerline data containing left and right address ranges, directionals, street names, street types, and left and right USPS ZIP codes can be implemented (Figure 3b). With the search criteria, a Los Angeles address indicated by the City field in the address will be passed to the Los Angeles address locator while a New York address will be filtered to the New York address locator. Thus the main strengths of this system include having control over the reference datasets, allowing interpolation offsets, support for probabilistic matching, and a user-friendly interface.

The ESRI StreetMap USA data was utilized as the reference dataset in this test, since ESRI users commonly have ready access to this dataset (Figure 4). StreetMap USA is provided for free to ESRI site license owners with any piece of ESRI software (i.e. ArcGIS, ArcIMS, ArcSDE, etc), in Smart Data Compression (SDC) format specifically prepared for nationwide streets display, routing, and geocoding (ESRI 2008b). It contains a USCB TIGER/Line files (2000 version) dataset enhanced by ESRI and Tele Atlas.

Geolytics

The Geolytics GeocodeDVD (or GeocodeCD) is a commercial geocoding software package that converts input addresses to latitude and longitude, then associates those with USCB Block group codes (2000 version) (Geolytics 2008). This software has inbuilt functionality for address parsing and standardization operations. The reference data for this geocoder are currently based on the USCB TIGER/Line files (2004 version), and also demographic data from the CensusCD 2000 Long Form (SF3). A user can choose to add up to 66 demographic variables pertaining to a USCB Block group or tract to their output file. The user is provided with an interface or dialog box which permits either manual (single) or batch mode data processing, as shown in Figure 5. This interface is divided into seven sections: (a) manual/batch page (top left), (b) options (middle left), (c) data page (top right), (d) help page, (e) configuration, (f) window size control and (g) scroll bar for viewing data. The various options allow address corrections for phonetic errors (i.e. Maine Street vs. Main Street, Greenwood Drive vs. Greenwood Road). However, the details concerning the exact phonetic algorithms performed are not provided. User specified options include allowing "phonetic match of state name in an address" (but not an abbreviation), "place-based USPS ZIP code match" (if a street is not found in a USPS ZIP code, scanning other USPS ZIP codes associated with the place is allowed), "fuzzy match (phonetic) of street name", "disregard parity for address match (matches even/odd addresses with even/odd address ranges)", "fuzzy address match" (closest address range to match the house number, rater than the accurate one), "fuzzy street type match", "geocode no matter what" (means that the geocoder will always use the closest rather than the most accurate match), "presort batch input", and "clean address list". Exact definitions of these terms are not provided in the software documentation, but the term "fuzzy" seems to refer to an attribute relaxation process that allows matches without requiring that all matches be exact (Goldberg 2008a). The set of results contains address match type which includes any corrections that were made to match the address (i.e. changing Road to Drive), geographic identifiers such as USCB Block group, longitude, latitude, place, and minor civil division (MCD), any errors encountered, and accuracy, for instance an exact match vs. fuzzy matching was used on the street type to correct a Road into Drive, and lastly the selected demographic data including racial, income, age, and housing distributions for the USCB Block group or tract that each address is located in.

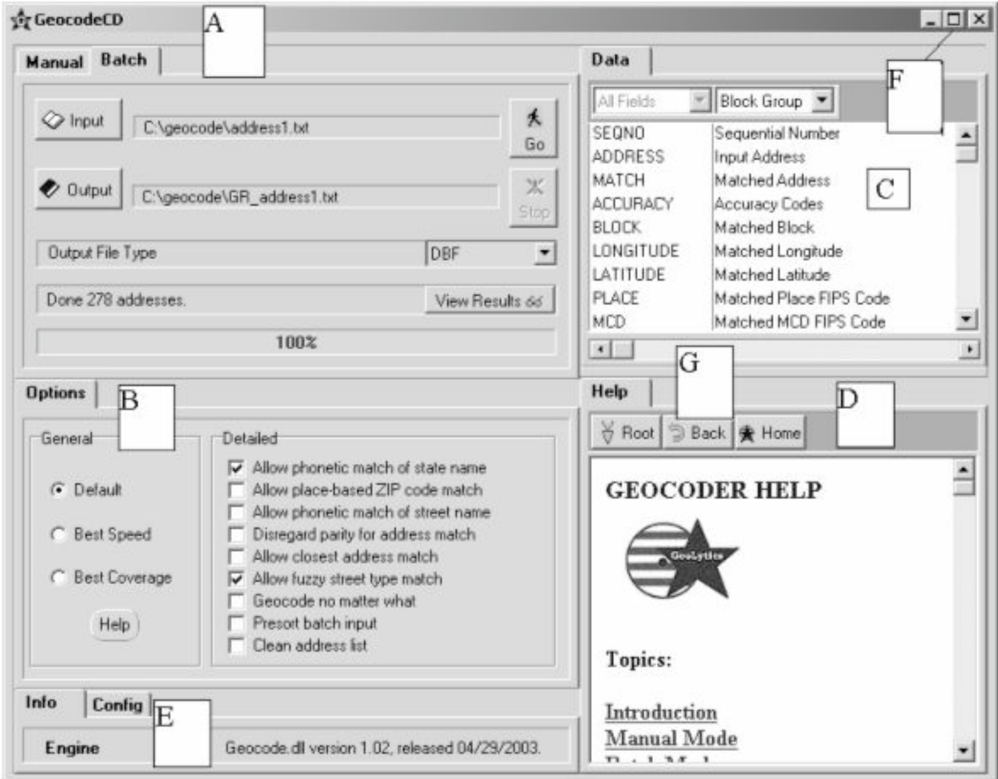


Figure 5 Geolytics GeocodeCD user interface (from Geolytics 2008)

2.2 Web-Based Applications

Geocoder.us

Geocoder.us is a widely used web-based geocoding tool offered by Locator Technologies (2008). This geocoder is based on the free and open source Perl module Geo::Coder::US (Erle and Walsh 2008). Geocoding single addresses via the main website is free, and the service utilizes USCB TIGER/Line files as the reference dataset. This tool takes a US address or intersection and returns a list of parsed address results. Geocoder.us can be used for simple address parsing, but not officially recognized address standardization because it is not a USPS CASS certified product (United States Postal Service 2008b). If the address can be parsed but not found, a single result is returned with latitude and longitude left undefined. If the address cannot be parsed, an empty list is returned. If the address cannot be found exactly, the closest result is returned. If a city and state, or a USPS ZIP code is not provided, the lookup will fail. No metadata or details on match type, standardization or other processing methodology are provided with the free online version. However, the original source code for the Perl module Geo::Coder::US (Erle and Walsh 2008) is available, from which these characteristics could conceivably be determined (i.e. reverse engineered).

According to Locator Technologies (2008) Geocoder.us geocoding functions can be accessed using several web-service oriented frameworks (i.e. XML-RPC, SOAP, REST). Geocoder.us also offers commercial batch processing services per 20,000 successful geocodes, and a small fee per record for each address beyond that. The commercial version supports distance calculations between two coordinates or two USPS ZIP codes, and there is also a lightweight geocoder (webpage) for mobile devices. In addition, a BETA version of Geocoder.us is being developed that will support geocoding to the centroid of a USPS ZIP code, or a city/state pair.

Google Earth

Google Earth is a downloadable, standalone program released by Google in 2005, which runs on Microsoft Windows, Mac OS X, and Linux (Google 2008a). The Virtual Globe program upon which it is based was originally created by Keyhole, Inc. and was acquired by Google in 2004. The free version supports single address geocoding. It offers advanced 3-dimensional (3D) globe-viewing features based on the superimposition of satellite images, aerial photography, and digital elevation model (DEM) data collected by NASA's Shuttle Radar Topography Mission (SRTM) (Google 2008b). Users can zoom in to view buildings and structures (i.e. bridges) in 3D, which have been submitted to Google by users who create the models using the 3D modeling program called SketchUp (Google 2008c). Though the basic Google Earth program is free to install on most computers (but not on servers), advanced functionality like overlaying spatially referenced data on top of Google Earth's 3D imagery requires the purchase of advanced commercial versions of Google Earth, or individual extensions that support specific functionalities (Google 2008d). No metadata or details on match type, address standardization or other geocoding processing methodology, or specifics on which reference data are utilized in geocoding are provided with the free online version.

Google Maps API

Google Maps is one of the most popular web-based GIS tools available, providing web mapping services and technologies for free, worldwide (Google 2008c, 2008d). It is used by countless map-based services as embedded maps on third-party websites through the utilization of the Google Maps API. The Google Maps site (http://maps.google.com) can be used to geocode single addresses, while developers can utilize the API to perform batch processing. Google Maps applications can provide street maps, a route planner, and an urban address/business locator for numerous countries around the world. Users can search an online map by entering an address, intersection, business name or attraction, or even just a general area, to speedily find information and driving instructions (US and Canada). Although Google uses the word "satellite", some of the high-resolution imagery is aerial photography taken from airplanes rather than from satellites. Nevertheless, high-resolution satellite images are available for most urban areas in Canada and the US, as well as parts of many other countries. The reference data used in Google Maps are provided by Tele Atlas and NAVTEQ, while the small patches of high-resolution satellite imagery are largely provided by DigitalGlobe and its QuickBird satellite, with some imagery also from government sources (Google 2008e). The main global imagery base called NaturalVue was derived from LANDSAT 7 imagery by MDA Federal (formerly Earth Satellite Corporation) (NASA 2006). Though this global image base provides the essential foundation for the entire Google Maps application, no specific details are available on which datasets are utilized in geocoding processes. No metadata or details on match type, address standardization or other geocoding processing methodology are provided with the free online version via Google Maps, but limited metadata regarding the match type and accuracy are available to developers when using the API.

USC Geocoding Platform

The open source USC Geocoding Platform (Goldberg 2008c) represents a proof of concept of the geocoding research underway at the USC GIS Research Lab. This system is based on the preliminary work of Bakshi et al. (2004) which identified the shortcomings of traditional linear-based interpolation geocoding methodologies, i.e., the parcel existence and homogeneity assumptions (Goldberg 2008a). Since its first creation in 2004, the system has evolved to include areal unit-based interpolation, attribute relaxation, and supports multiple user defined feature matching hierarchies including both the traditional postal address hierarchy structure, e.g. the NAACCR GIS Coordinate Code hierarchy (Hofferkamp and Havener 2008) and a reference feature uncertainty-based version (Goldberg 2008a). The reference data sets used for the evaluation performed in this report are USCB TIGER/Line files (2006b version), the Los Angeles County Assessor parcel files, and the administrative units from the 2000 USCB. The current implementation of the system performs strictly deterministic feature matching, meaning that probabilistic matching is not supported, although plans have been made to incorporate this functionality. In addition, the system includes a deterministic address parsing and normalization/standardization engine, which, while non-USPS CASS certified (United States Postal Service 2008b), has proven quite capable of parsing most normal USPS style address data into USPS *Publication 28* standards (United States Postal Service 2008c).

This system is available free of charge as a web service through the USC GIS Research Lab website (https://webgis.usc.edu) (Goldberg 2008c) and supports both batch- and single-record processing. PC-based versions are also available, but require that the reference data be stored locally on the user's machines, with plans also in place to alleviate this need by providing remote hosting for the reference data as well. The USC GIS Research Lab plans to make the entire system available as open source to the greater research community once it has reached an acceptable level of functionality and stability. The benefits of this approach are three-fold. The first is that it will be useful to other organizations in terms of reducing the up-front costs of geocoding data. Secondly, the researchers hope that it will stimulate a community of developers to contribute to building a more robust and complete version of a one-stop geocoding solution through an extensible framework such that experts can improve specific components which will then be available to the greater community at large. Finally, it is hoped that an open source geocoding platform will remove the uncertainty currently associated with the majority of results from most commercial geocoders in terms of the aspects typically held as "trade secrets", e.g. how the system processed a record, the decisions that were made, the rationale for why a choice was made, and any alternative options that could have been chosen.

Yahoo Maps API

Like Google Maps, Yahoo Maps is another free and very popular suite of web-based GIS tools that include geocoding functionality. The Yahoo Maps site (http://maps.yahoo.com) supports geocoding of single addresses, while the API can be utilized by developers to perform batch processing. The Yahoo Geocoding Web Service allows a user to find the specific latitude and longitude for addresses (Yahoo, 2008). This service can be used simply to geocode addresses, or it can be implemented by creating a Yahoo Maps application. Users can search an online map by entering an address, intersection, business name or attraction, just like Google Maps. More than one result may be returned if the given address is ambiguous. The precision of the address used for geocoding ranges from specific street address all the way up to country. A successful match is typically an address and/or location (such as a business), including street, USPS ZIP+4, ZIP+2, ZIP, city, state and country. Warnings are provided if the exact address was not found, and the closest available match will be returned. The Yahoo Maps API and free online single address geocoding tool on the Yahoo

Maps website (http://maps.yahoo.com) do not support highly detailed metadata on match type and accuracy, though some information is provided when an exact match is not found.

2.3 Licensing and Cost Comparison

A summary of the licensing and costs associated with each of the software tested as part of this study is provided in Table 3, with regard to initial start-up. The advantages of commercial products include reference data compilation, maintenance and updating, user-friendly interfaces, and technical support. The main disadvantages, however, are the high prices associated with startup and long-term maintenance and licensing. Also, in some cases variation of reference datasets is not supported and costs associated with customization (functionality, reference data, etc.) can be prohibitive.

The primary advantages of utilizing open source solutions are that the software is available for free and that the complete code base is available. The former means that portions of the up-front and recurring costs are alleviated because the software is available for free and no license fees are required. Because the source code is available, the geocoding software is fully extensible. One can build in their own custom components, ensure cross-platform compatibility, and take advantage of the freedom to customize the applications as required, e.g. individual choice of reference datasets, matching rules, and other options. The chief disadvantage of open source geocoders, particularly the USC Geocoding Platform, is that they lack the fundamental "user friendliness" inherent in the commercial product alternatives. Therefore, these geocoders are typically more complicated to use if one wishes to install and run them locally in their own organization because at least one individual at the organization would need to become somewhat of an expert in the particular open source solution. However, as existing more established open source products have shown, many of these issues become moot as the larger community begins to utilize and contribute to the development of the tools, e.g. Ubuntu Linux, and the MySQL and PostgreSQL database systems (Canonical Ltd. 2008, MySQL AB 2008, and PostgreSQL Global Development Group 2008).

3 Evaluation Results

The results of the evaluations of the geocoding software listed in Table 1 include comparisons of the positional accuracies, USCB Block groups and match rates of the geocoded addresses with respect to the reference dataset. A summary of the match results is presented in Table 4 and Table 5. Table 4 lists the original output accuracy and correction descriptions for each address geocoded, with respect to each piece of software tested. For instance, the Centrus accuracy descriptions are based on a hexadecimal output code attributed to each address geocoded. Table 4 also provides a comparison described as "Mismatch" between the eight geocoding software tested and the reference dataset with regards to USCB Block groups, again for each address geocoded. Table 5 also includes the total number of mismatches with respect to each program. The USCB Block group comparison is based on an overlay of the geocoded addresses on the 2000 USCB Block group layer, since most of the software evaluated did not provide this information in the metadata. The overall match rate in Table 5 is defined as the ratio between the total number of matches reported and the total number of geocodes attempted. The parcel or building match rate is defined as the total number of geocodes reported as corrected parcel or building centroid matches. Lastly the total number of all other matches, such as those requiring manual intervention, are grouped together and divided by the total number of geocodes attempted.

Table 3 Licensing and costs of geocoding systems evaluated as part of this study

Name	Licensing	Cost
Centrus US Street Point Data	Per user, enterprise	Pricing of versions is highly variable - dependent upon the user. Datasets: USPS ZIP+4 Centroid Data Set:: "Not available with Desktop" Points Data Set:: Not available for single state license. List price for National is $27,500 plus $25,000 additional software fee for Pointes-Level option. Enhanced Data Set: $1,000 GDT Data Set:: $2,750 Tele Atlas Points Data Set:: same as Centrus Points pricing NAVTEQ (US) Data Set:: $2,750
ESRI Address Locator	Per user, enterprise and site licenses	60-day trial (free) version of ArcGIS. Pricing of full version is highly variable - dependent upon the user
Geocoder.us	Per address/record	Single address processing is free. batch processing services at $50.00 per 20,000 addresses. After 20,000 records the cost is 0.025 cents per record sent (successfully matched), not for each record which is geocoded.
Geolytics GeocodeDVD	GeocodeDVD Geocode API	Per State: $600 For API: Internal product development pricing is $6,000/year. To embed the geocoder into a product that will be sold: either $1200 per seat or $30,000 per year for a Professional Use License.
Google Earth	Per user/computer (Seat) Enterprise	Per Seat: Google Earth Pro: $400 Per Seat: Google Earth Plus: $20 Enterprise: Pricing variable - dependent upon the user
Google Maps API	Personal, worldwide, royalty-free, for use only	Startup: N/A
USC Geocoding Platform	No limit for academic, government, and/or other non-profit/philanthropic usage	Startup: N/A
Yahoo Maps API	The Geocoding service is limited to 5,000 queries per IP address per day	Startup: N/A

According to Table 5, Centrus, the USC Geocoding Platform, ESRI, and Google Maps API have the highest match rates in terms of exact matches versus corrected matches, and the lowest overall mismatches with respect to the reference dataset USCB Block groups (red text). It is important to note that 58% of the initial geocodes were automatically or manually adjusted to generate the reference locations (Figure 1, Table 2). And in terms of comparability, it is also noteworthy that reference datasets used by different geocoders differ, as well as the methodologies used in geocoding.

Centrus 2008 US Street Point Data

Results of the evaluation of the Centrus Street Point Data geocoding are presented in Table 4 - Table 5, and Figure 6 - Figure 7. The methodology and reference datasets employed included address matching, USCB Block group, parcel centroid, street-level centroid and USPS ZIP code centroid matching. In terms of accuracy, address geocodes indicate a geocode made directly to an interpolated street address, parcel centroid, building centroid, matched street segment, or in the case of an intersection, two segments.

Table 4 Summary of accuracy and corrections performed in geocoding processing using software listed in Table 1.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
1	no	address, street range location, street side is unknown and the Census FIPS Block ID is assigned from the left side and there is no offset and the point is placed directly in the street, address interpolated onto a TIGER/Line segments that did not initial contain address ranges OR the original segment name was changed to match the USPA spelling (street type, predirectional and postdirectional)	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
2	no	address, street range location, address interpolated onto a TIGER/Line segments that did not initial contain address ranges OR the original segment name was changed to match the USPA spelling (street type, predirectional and postdirectional)	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
3	no	address, street range location, best location	no	86	no	address	no	fuzzy street type match	no	address	no	address range interpolation	no	exact address match	no	city centroid
4	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid

Table 4. Cont.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
5	no	address, point-level data location, center of a parcel	no	86	no	address	no	fuzzy street type match	no	address	no	address range interpolation	no	exact address match	no	city centroid
6	no	address, point-level data location, center of a parcel	no	100	yes	address - 36500 Pacific Coast Hwy CA 90265 – not exact match	no	address; if adjust zip exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
7	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
8	no	address, street range location, best location	no	100	no	address	yes	Address; adjust: add comma	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
9	no	address, point-level data location, center of a parcel	no	100	no	address	no	phonetic match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
10	yes	derived from a 5-digit ZIP code centroid, unclassified centroid accuracy accurate to at least County level, less than 80% of addresses in the ZIP code are in a single Census tract, Census ID is uncertain	no	67	no	address - 23498 Civic Center Way – not exact match	no	closest address match	no	address - could not match street number ; chose Toy Crazy, 3410 Civic Center Way, Malibu, CA 90265	no	street centroid	no	exact address match	yes	city centroid

Table 4. Cont.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
11	no	address, street range location, best location	no	100	no	address	yes	exact address match	yes	address	yes	address range interpolation	no	address; street number not found. Near location chosen	no	address range interpolation
12	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no		no	address range interpolation
13	no	address, street range location, best location	no	100	no	address	yes	does not work without a zip	no	address	no	address range interpolation	yes	city; Street could not be found. Center of the city. used	no	address range interpolation
14	no	address, street range location, best location	yes	100	yes	address; adjust: - add "," after "Street"	yes	phonetic match; adjust: add comma	no	address	no	address range interpolation	yes	ZIP; exact location could not be found. Center of the ZIP code. used	no	address range interpolation
15	no	address, street range location, best location	no	82	no	address	no	direction corrected	no	address; adjust: add street prefix "E"	no	address range interpolation	no	address; exact location could not be found; closest match: 1240 E San Antonio Dr, Long Beach, CA 90807 used	no	address range interpolation
16	no	address, point-level data location, center of a parcel	no	100	no	address	yes	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
17	no	address, point-level data location, center of a parcel	no	82	no	address	no	direction corrected	no	address; adjust: add street prefix "E"	no	address range interpolation	no	address; exact location could not be found; closest match: 3606 E Arabella St, Long Beach, CA 90805 used	no	address range interpolation

Table 4. Cont.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
18	no	address, point-level data location, center of parcel	no	86	no	address	yes	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
19	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	yes	ZCTA centroid
20	no	address, point-level data location, center of a parcel	no	100	no	address	no	fuzzy street type match; phonetic match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
21	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
22	no	address, street range location, best location	yes	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
23	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
24	no	address, point-level data location, center of a parcel	yes	100	yes	address; can chose 1 of 2	yes	exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
25	no	address, street range location, best location	no	82	no	address	yes	Address; match if adjusted: add comma	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
26	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
27	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
28	yes	address, point-level data location, center of a parcel	no	100	yes	address	yes	exact address match	yes	address	no	address range interpolation	no	exact address match	no	address range interpolation
29	no	address, point-level data location, center of a parcel	yes	82	yes	address - 2421 N Lake Ave - not exact match	yes	address; tried to adjust: add comma	yes	address; - chose "North Lake Villas"	yes	Street centroid	no	address; exact location could not be found, closest match: 2851 Lake Ave, Altadena, CA 91001 used	no	address range interpolation

Table 4. Cont.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
30	no	address, street range location, best location	yes	100	no	address	yes	exact address match	no	address	no	address range interpolation	no	address; exact location could not be found; closest match: 2333 Lake Ave, Unit H, Altadena, CA 91001	no	address range interpolation
31	no	address, point-level data location, center of a parcel	no	82	no	address	yes	exact address match	no	address; - chose 1 of 2: Essex House Hotel / Park Plaza	yes	Could not match	no	address; exact location could not be found; closest match: 44916 10th St W, Lancaster, CA 93534	no	address range interpolation
32	no	address, street range location, best location	no	100	no	address	no	address; adjust: ZIP exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
33	no	address, street range location, best location	no	100	no	address	yes	direction corrected	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
34	no	address, point-level data location, center of a parcel	yes	100	yes	address	no	exact address match	no	address	no	address range interpolation	yes	exact address match	yes	address range interpolation
35	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
36	yes	address, street range location, best location	no	82	no	address	yes	direction corrected	yes	address; if chose 1 of 2: Jeff Moffatt Law Office	yes	Could not match	yes	ZIP; exact location could not be found. Center of the ZIP code. used	no	address range interpolation
37	no	address, street range location, best location	no	100	no	address	yes	direction corrected; - adjust: add comma	no	address	no	address range interpolation	no	exact address match	no	address range interpolation

Table 4. Cont.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
38	no	address, street range location, best location	no	100	no	address	yes	closest address match ; fuzzy street type match; phonetic match; place based ZIP match	yes	address	yes	address range interpolation	no	exact address match	no	address range interpolation
39	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
40	no	address, point-level data location, center of a parcel	no	100	yes	address; if chose 1 of 2: 2298 W Ave E - not exact match	yes	closest address match; direction corrected	no	address	no	address range interpolation	no	exact address match	no	city centroid
41	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
42	no	address, point-level data location, center of a parcel	no	100	no	address	yes	address; - adjust: add comma	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
43	no	address, street range location, best location	no	100	no	address	yes	address; adjust: add comma	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
44	no	address, point-level data location, center of a parcel	no	86	no	address	no	closest address match; fuzzy street type match; place based ZIP match	yes	address; - could not match street number	no	address range interpolation	no	address; exact location could not be found, closest match: 740 S Broadway, Los Angeles, CA 90014	no	nearest parcel centroid
45	no	address, street range location, best location	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation

Table 4. Cont.

ID	Mismatch	Centrus	Mismatch	ESRI Mach Score	Mismatch	Geocoder.us	Mismatch	Geolytics	Mismatch	Google Earth	Mismatch	Google Maps	Mismatch	Yahoo Maps	Mismatch	USC Geocoding Platform
46	no	derived from a 5-digit ZIP code centroid, unclassified centroid accuracy accurate to at least County level, less than 80% of addresses in the ZIP code are in a single Census tract, Census ID is uncertain	no	67	yes	address; 600 S Hope St - not exact match	yes	closest address match	no	address; - could not match street number	no	Street centroid	yes	address; Street number could not be found. nearby location.	no	ZCTA centroid
47	no	address, point-level data location, center of a parcel	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	nearest parcel centroid
48	no	address, street range location, best location	no	100	no	address	no	exact address match	no	address	no	address range interpolation	no	exact address match	no	address range interpolation
49	no	address, street range location, best location	no	100	yes	address	yes	exact address match	yes	address	no	address range interpolation	no	exact address match	no	address range interpolation
50	no	address, street range location, best location	no	100	yes	address	no	exact address match	yes	address	no	address range interpolation	no		no	address range interpolation

Table 5 Summary of match rates of geocoding systems evaluated as part of this study. Best overall match rates are highlighted in red.

Name	Overall Match Rate	Parcel/Building Match Rate	All Other Match Rates	Number of USCB Block group Mismatches Compared to Reference Dataset
Centrus US Street Point Data	48/50	27/50	23/50	3
ESRI Address Locator	48/50	Not reported	12/50	6
Geocoder.us	42/50	Not reported	12/50	10
Geolytics GeocodeDVD	27/50	Not reported	23/50	20
Google Earth	42/50	Not reported	8/50	8
Google Maps API	48/50	Not reported	2/50	5
USC Geocoding Platform	48/50	9/50	31/50	3
Yahoo Maps API	39/50	Not reported	11/50	5

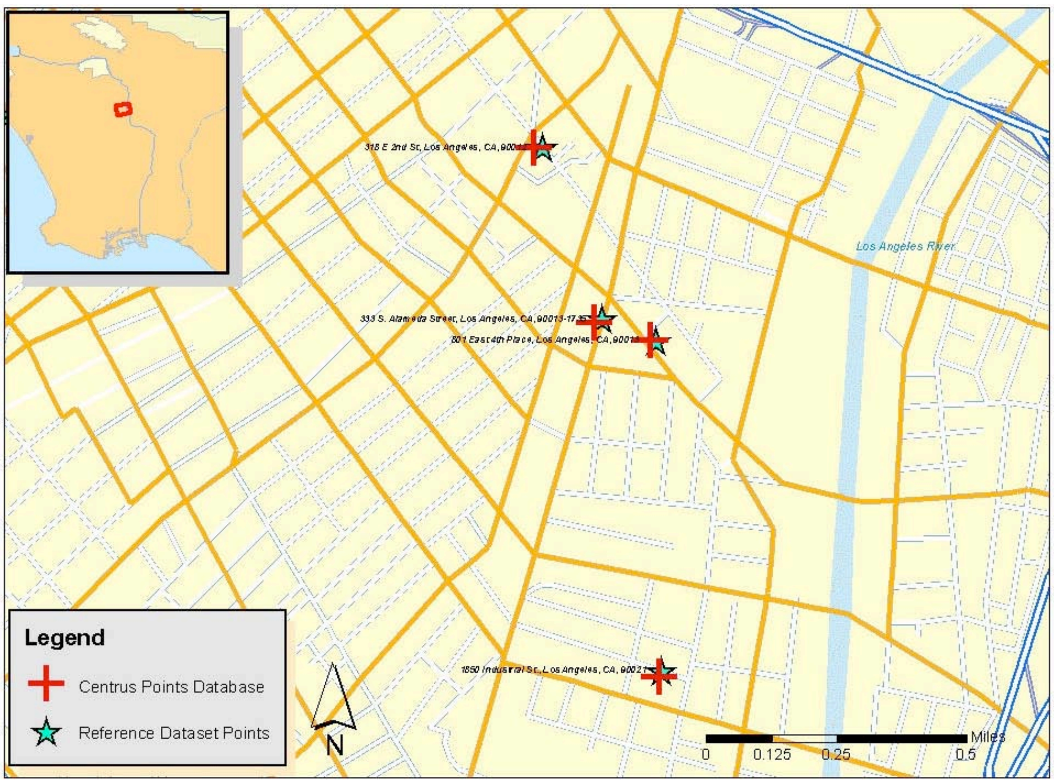


Figure 6 Partial view of Centrus geocoding test results and the corresponding reference dataset locations

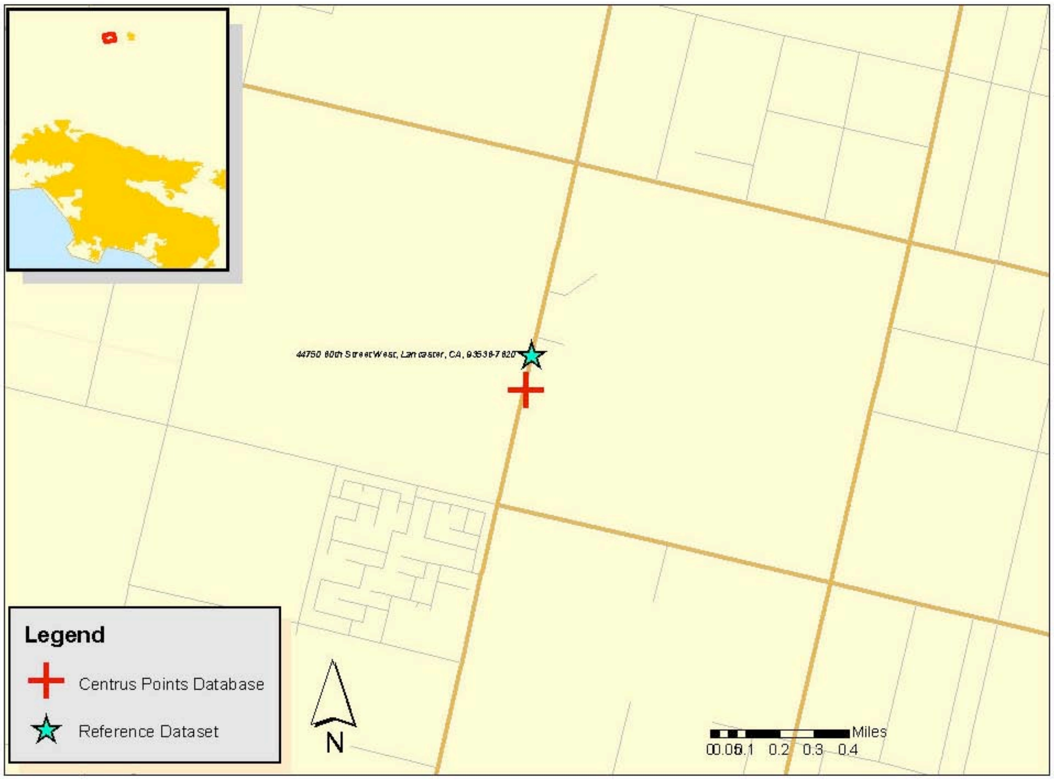


Figure 7 Partial view of Centrus geocoding test result and corresponding reference dataset location

Street centroid geocodes indicate that no match was made to the input address number, or that the input did not include a street number, and that a match was made to a street segment within the search area based on the input USPS ZIP code. If there is no USPS ZIP code, then the search area is based on the input city, depending on the user's address and match settings. Street centroids have a range of confidence depending on the search area used to obtain the matching street segment. USPS ZIP+4 centroid locations have a range of confidence depending on how the USPS ZIP+4 centroid was determined.

The Centrus evaluation yielded a total of 48 out of 50 address matches, with two unmatched (Table 4). Of the 48 matches, 39 were standardized with no change, six had the prefix direction corrected, the street name was corrected on three and the street suffix on another three, and one had the USPS ZIP code adjusted. To summarize, 21 geocodes (42%) were address matched, 27 geocodes (54%) were matched to parcel centroids, and two were matched to USPS ZIP code centroids (4%). Figure 6 - Figure 7 provide illustrations of portions of the geocoded locations with respect to the positions of the reference data points. In terms of positional accuracy, 10 of the geocoded addresses fall within a distance of 5 m from the reference locations, five are within 10 m, and the remaining 34 were placed within 1000 m of the reference data points.

ESRI Geocoding and Address Management

The addresses in Table 2 were geocoded by running the ArcToolbox Geocode Addresses tool using an address locator created with the abovementioned reference data sources and techniques. The methodology employed included choosing the address style "US Streets with City State ZIP" (which exactly matches the input dataset), setting the spelling sensitivity and acceptable "match score" levels, and standardization of input addresses (Figure 4). Spelling sensitivity controls how much variation (how many candidates) the geocoding service will allow when it searches for likely candidates in the reference dataset for the address to be geocoded. Spelling sensitivity runs from 0 to 100. The spelling sensitivity does not affect the geocode "match score". This is different from the match rates provided in Table 5. To illustrate, a schematic of the ESRI geocoding process is provided in Figure 8 (ESRI 2003). When an address is entered, the address locator points the input address to a rule base file, and the address is broken into address elements (e.g. street name, street type, and number) to be standardized, such as making all the directionals or street types the same (for example, Street, St., and Str. would all come back the same). There are specific files in the ESRI default rule base for each address locator style, which can be modified by the user (ESRI 2003). The address locator then assigns each standardized address element to a specific category used for matching and ESRI geocoder "scoring". The geocoder uses preprogrammed weights for each element in an address, and then the scores are accumulated to generate an overall score not to exceed 100 (e.g. Figure 9). Thus a "match score" indicates the extent of the overall address match. Some of these settings can be controlled through properties in the address locator (Figure 4), while others, such as modifying a rule base file (e.g. match weighting), require additional programming effort "outside of the box" (ESRI 2003). Lastly, the address locator presents the best matches based on the score and the location of address being matched (Figure 9).

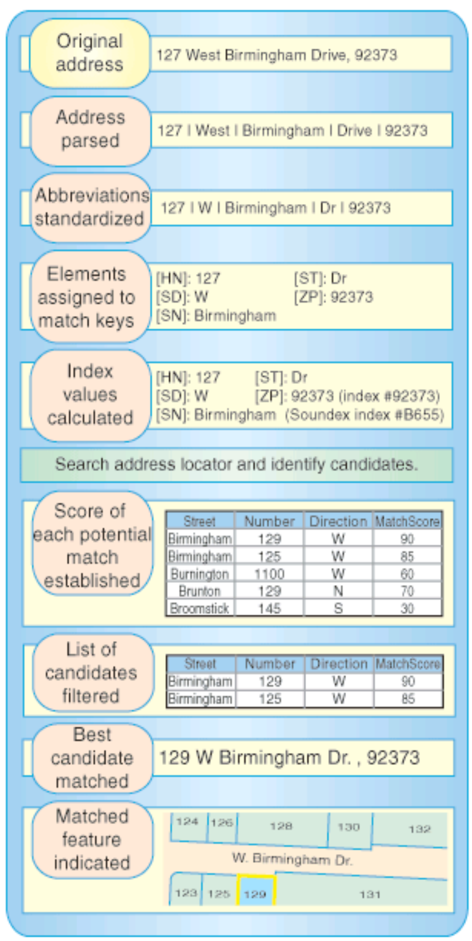


Figure 8 Summary of ESRI geocoding process (ESRI 2003)

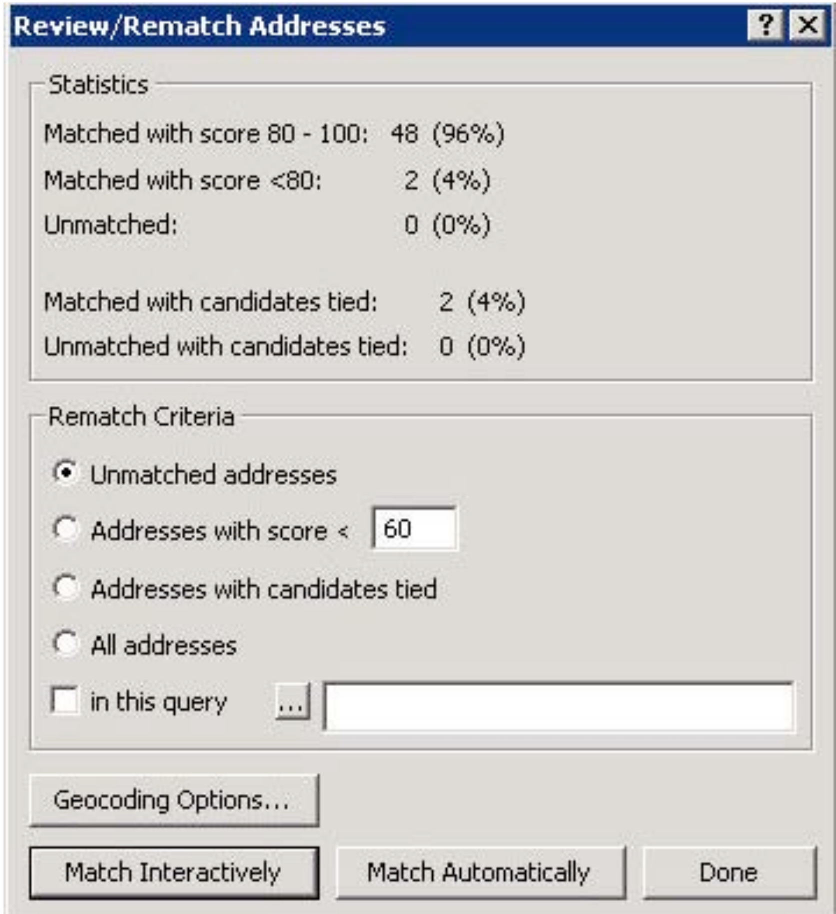


Figure 9 ESRI geocode test results summary

When the geocoder searches for likely candidates in the reference data, it uses a user-specified threshold (minimum score) to determine whether a potential candidate may be considered (Figure 4, dialog box "Geocoding Options"). If the score exceeds the minimum, it is considered a match; if it does not, it is not considered a match. For this study, a spelling sensitivity of 80 was utilized, the minimum candidate score was set to 30, and the minimum match score to 60. The offset value is a constant applied to all addresses, a straight distance used to adjust the location of the geocoded address away from the centerline of a street toward the address being geocoded. The offset value was set to 3 m when this test was performed. The output file consists of an ESRI shapefile containing the geocoded points representing the input address locations. Table 4 - Table 5 and Figure 9 - Figure 11 provide summaries of the settings chosen and results produced. A total of 48 out of 50 addresses (96%) were reported as matched, while two were matched with a score of < 80%. 38 out of 50 (76%) of the geocoded addresses were reported as having a score of 100, while 12 addresses (24%) listed in Table 4 had scores of less than 100. Figure 10 - Figure 11 provide examples of locations of the geocoded addresses with respect to the positions of the reference data points. In terms of positional accuracy, eight of the geocoded addresses fall within a distance of 5m from the reference locations, six are within 10 m, and the remaining 36 were placed within 1000 m of the reference data points.

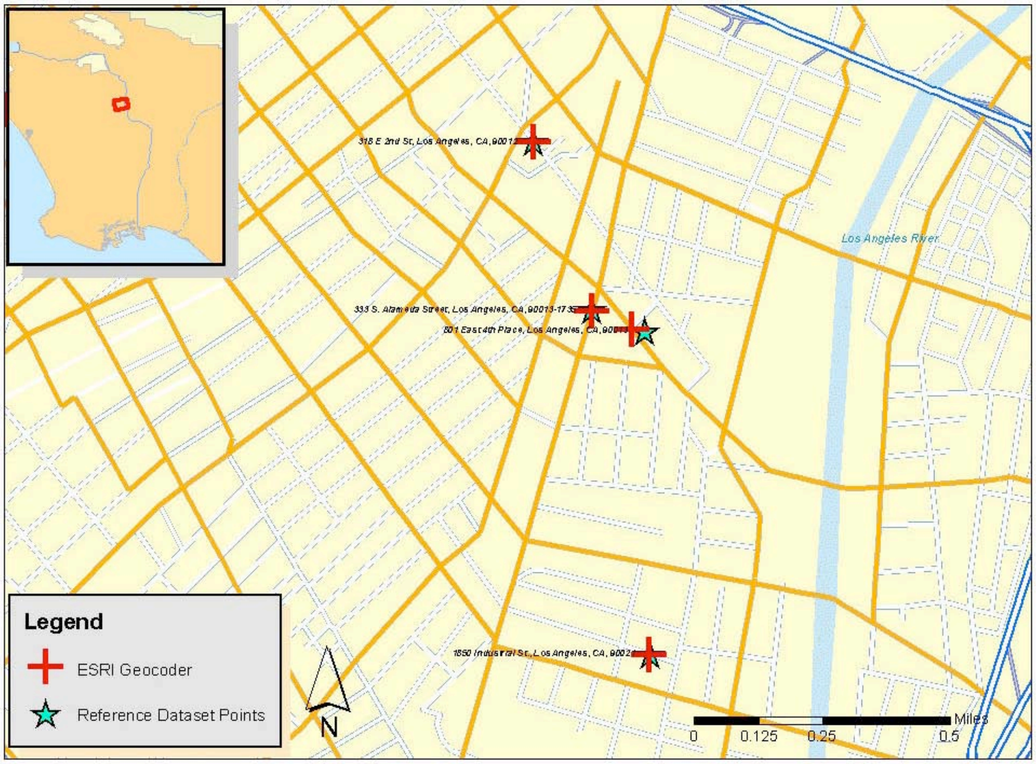


Figure 10 Partial view of ESRI address locator geocoding test results and the corresponding reference dataset locations

Figure 11 Partial view of ESRI geocoding test result and the corresponding reference dataset location

Geolytics

The methodology employed in testing the Geolytics software allowed phonetic match of state and street names, place-based USPS ZIP code matching, fuzzy street types and closest address match. The results of geocoding the addresses in Table 2 are presented in Table 4 and Figure 12 - Figure 13. A total of 27 out of 50 addresses (54%) were exactly matched, six more (12%) were exactly matched after manually adjusting the addresses, two (4%) were not matched at all (no coordinates provided), and 25 (50%) were matched using one or a combination of two or more matching options. For instance five of the 26 (10%) were matched to the closest address, five (10%) were based on fuzzy street type, four (8%) were phonetic matches, and two (4%) were place-based USPS ZIP code matches. Figure 12 - Figure 13 provide different views of several of the geocoded locations with respect to the geographical positions of the reference data points.

Although a number of the geocoded addresses illustrated in Figure 12 - Figure 13 are aligned with the reference locations, it is apparent that several do not fall close to the reference data locations and are therefore considered to be of low positional accuracy. Thirteen of the geocoded addresses fall within a distance of 5 m from the reference locations, two are within 1 0m, and the remaining 34 were placed within 1 km of the reference data points.

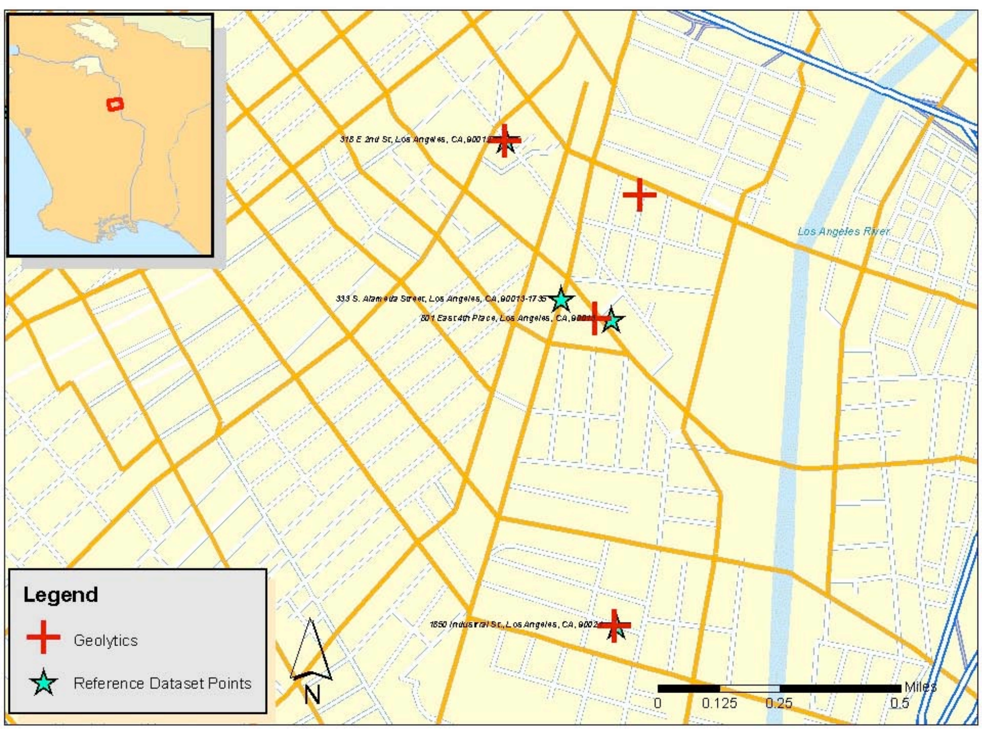


Figure 12 Partial view of Geolytics geocoding test results and corresponding reference dataset locations

Figure 13 Partial view of Geolytics geocoding test result and corresponding reference dataset location

Geocoder.us

The results of geocoding the addresses in Table 2 using Geocoder.us are presented in Table 4 and Figure 14 - Figure 15. Forty-tow out of 50 of the addresses (84%) were reported as successful matches; there were eight (16%) as listed in Table 4. No detailed metadata was provided. Figure 14 - Figure 15 provide illustrations of portions of the geocoded locations with respect to the positions of the reference data points. Most of the geocoded addresses fall close to the reference data locations, while several are obviously quite different and thus are of low positional accuracy. Two of the geocoded addresses fell within a distance of 5 m from the reference locations, one is within 10 m, and the remaining 45 geocoded were placed within 1 km of the reference data points.

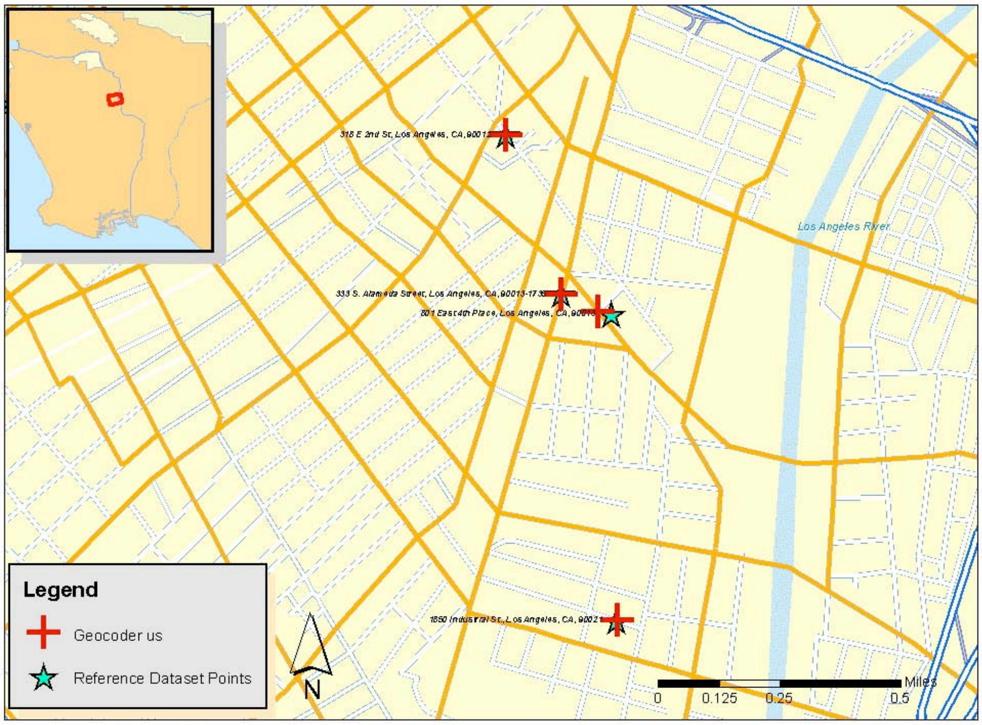


Figure 14 Partial view of Geocoder.us geocoding test results and corresponding reference dataset locations

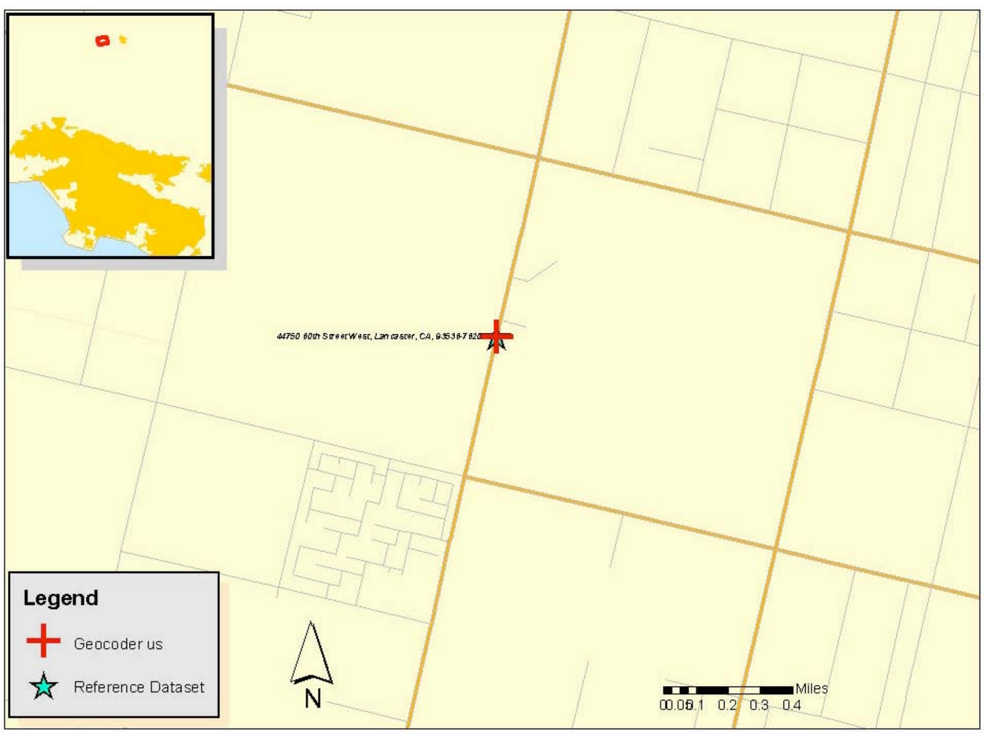


Figure 15 Partial view of Geocoder.us geocoding test result and corresponding reference dataset location

Google Earth

The results of geocoding the addresses in Table 2 using the free version of Google Earth are presented in Table 4 - Table 5 and Figure 16 - Figure 17. A total of 42 of 50 addresses (84%) were reported as exact matches, while eight out of 50 (16%) required manual matching. In three of the eight cases (6% of the total), street number could not be matched, thus approximate coordinates (nearest street address) were accepted. In another three cases (6%), the closest match out of two alternatives was chosen. Figure 16 - Figure 17 provide illustrations of portions of the geocoded locations with respect to the positions of the reference data points. In most cases the addresses geocoded using Google Earth aligned well with the reference dataset. None of the geocoded addresses fell within a distance of 5 m from the reference locations, six are within 10 m, and the remaining 44 geocoded were placed within 1 km of the reference data points.

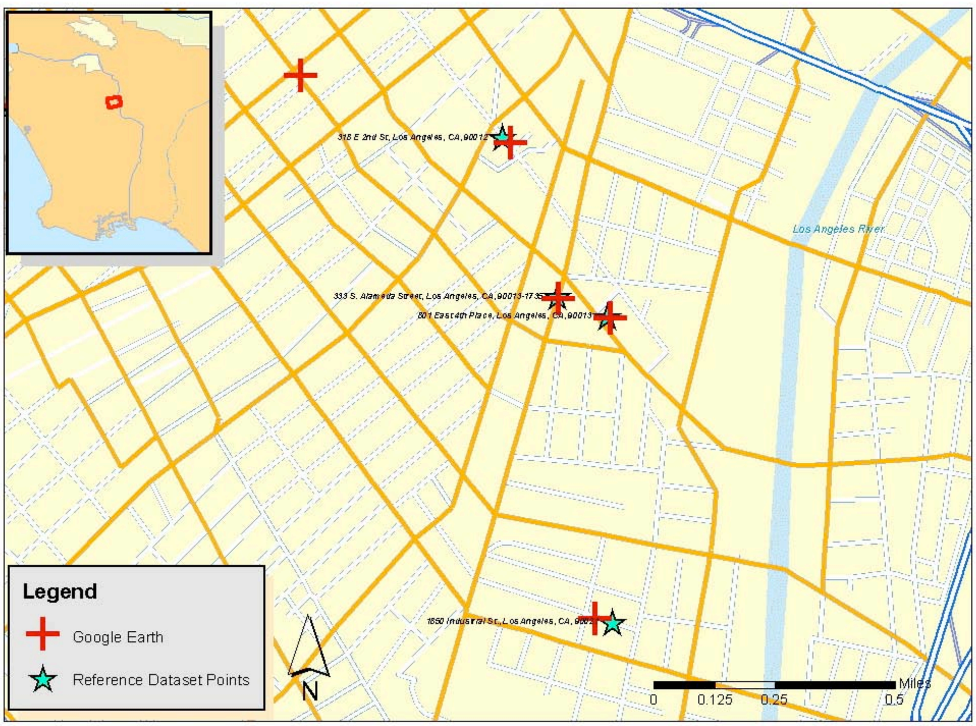


Figure 16 Partial view of Google Earth geocoding test results and corresponding reference dataset locations

Figure 17 Partial view of Google Earth geocoding test result and corresponding reference dataset location

Google Maps API

For the purposes of this study, the Google Maps API was utilized to evaluate the addresses in Table 2 through an implementation created for use in the USC GIS Research Lab Geocoding Correction Services (Goldberg 2008b). The results are presented in Table 4 - Table 5 and Figure 18 - Figure 19. A total of 48 out of 50 addresses (96%) were reported as matches, while two out of 50 (4%) could not be matched. Of the 48 matches, 45 matches (90%) were matched based on address range interpolation, and three (6%) were assigned coordinates based on street centroids. Figure 18 - Figure 19 provide illustrations of portions of the geocoded locations with respect to the positions of the reference data points. Eleven of the geocoded addresses fell within a distance of 5 m from the reference locations, six are within 10m, and the remaining 31 geocoded were placed within 1 km of the reference data points.

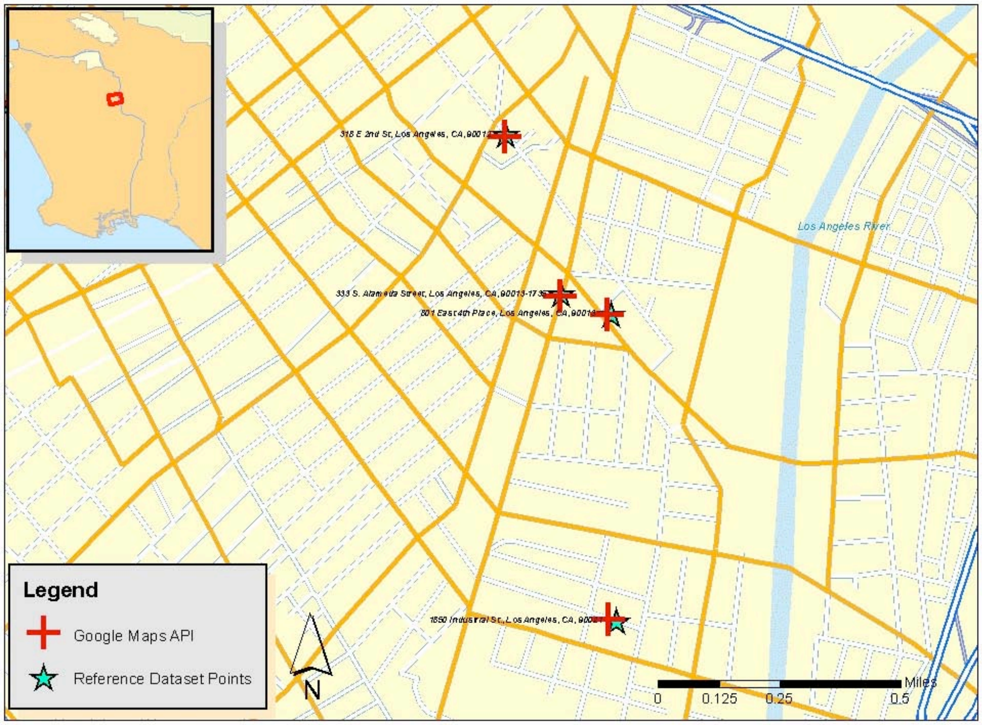


Figure 18 Partial view of Google Maps geocoding test results and corresponding reference dataset locations

Figure 19 Partial view of Google Maps geocoding test result and corresponding reference dataset location

USC Geocoding Platform

For the purposes of this study, the USC Geocoding Platform was utilized to evaluate the addresses in Table 2 (Goldberg 2008c). No additional corrections were performed on this dataset, other than running this utility as-is. The results of geocoding the addresses in Table 2 using the USC Geocoding Platform are presented in Table 4 - Table 5, Figure 20 - Figure 21. A total of 34 of 50 addresses (68%) were reported as address range interpolation matches. Nine geocodes (18%) were reported as locating the nearest parcel centroid, four (8%) to city centroids, and another two (4%) were assigned to USCB ZCTA centroids. Twenty-one of the geocoded addresses fall within a distance of 5 m from the reference locations, two within 10 m, and the remaining 27 geocoded were placed within 1 km of the reference data points.

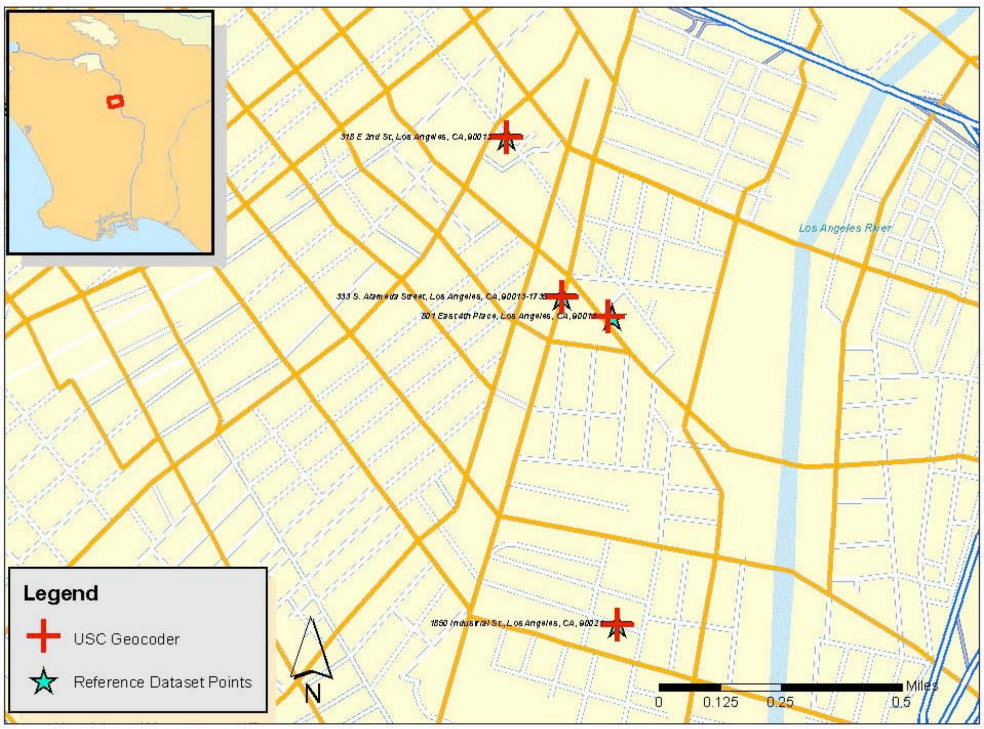


Figure 20 Partial view of USC Geocoding Platform geocoding test results and corresponding reference dataset locations

Figure 21 Partial view of USC Geocoding Platform geocoding test result and corresponding reference dataset location

Yahoo Maps API

The results of geocoding the addresses in Table 2 using the Yahoo Maps API are presented in Table 4 - Table 5 and Figure 22 - Figure 23. A total of 39 out of 50 addresses (78%) were reported as exact matches. Of the 11 remaining matches (22%), eight were based on the closest matching address or location, two (4%) were matched to USPS ZIP code centroids, and one (2%) was assigned to the center of a city. Figure 22 - Figure 23 provide illustrations of portions of the geocoded locations with respect to the positions of the reference data points. One of the geocoded addresses fell within a distance of 5 m from the reference locations, two are within 10 m, and the remaining 46 geocoded were placed within 1 km of the reference data points.

A summary of the distances in meters of each geocoded position with respect to the reference dataset locations is provided in Table 6 and Figure 24. The variations in the positional accuracies in the results are easy to discern in Figure 24. Overall, the Centrus software, ESRI, Geocoder.us, Google Maps and the USC Geocoding Platform performed better in terms of positional accuracy compared to the other four software evaluated in this study.

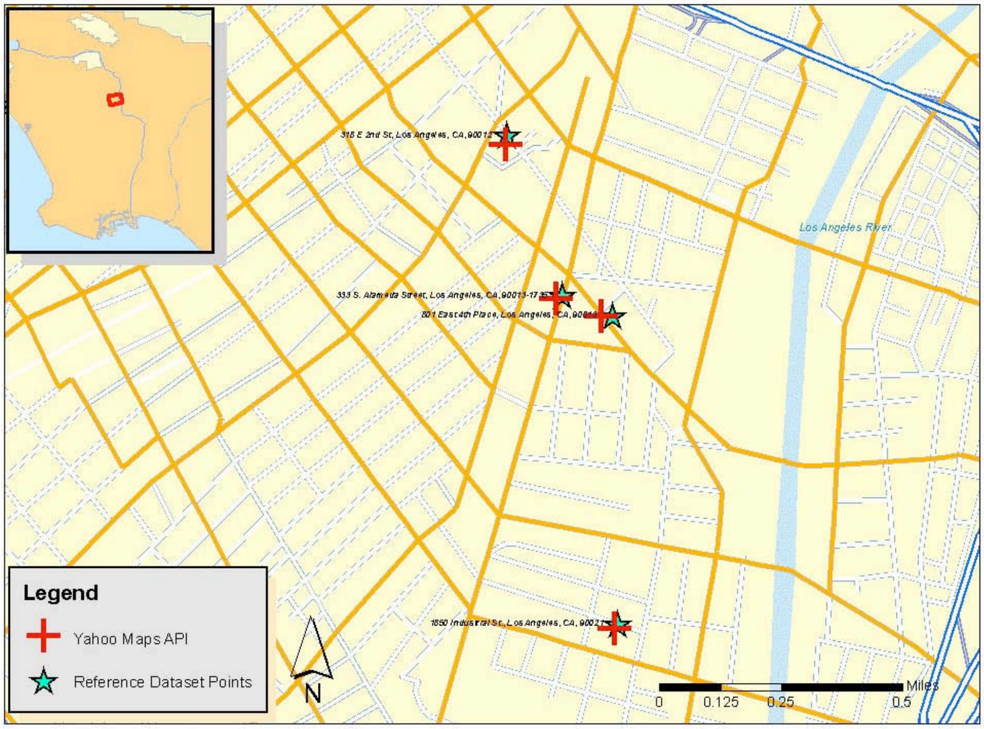


Figure 22 Partial view of Yahoo Maps geocoding test results and corresponding reference dataset locations

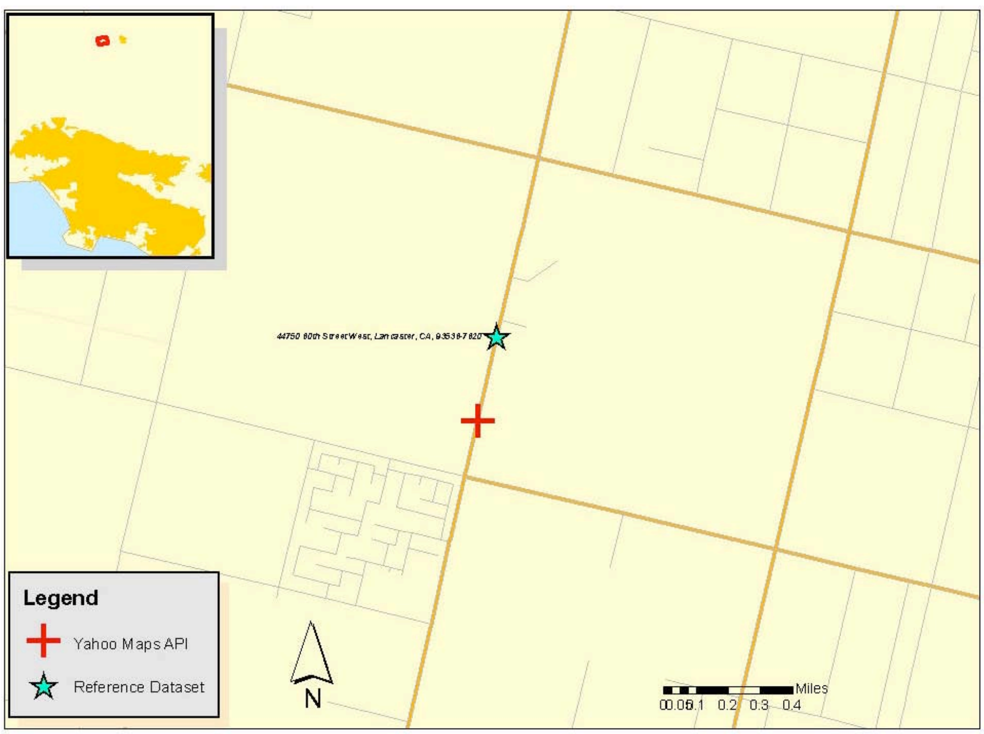


Figure 23 Partial view of Yahoo Maps geocoding test results and corresponding reference dataset location

Table 6 Accuracies of geocoded data in terms of linear distance (meters) from reference dataset positions

ID	Centrus	ESRI	Geocoder.us	Geolytics	Google E	Google M	Yahoo M	USC
1	101.9	133.3	308.2	60.0	156.5	317.8	73.7	0.5
2	97.8	74.1	85.4	76.3	39.6	73.2	91.1	0.04
3	457.2	321.5	85.1	365.7	651.5	77.9	319.4	19496.2
4	6.7	122.9	7.2	237.9	296.1	8.2	6.4	6.6
5	59.4	48.0	5.1	45.2	110.1	33.0	84.4	4549.6
6	11.2	310.4	48.7	26391.9	525.1	11.5	18.6	11.5
7	6.3	96.3	5.7	113.8	94.2	7.4	69.2	6.3
8	1163.3	1029.1	1126.6	1060.3	901.2	1093.4	1100.3	0.4
9	29.6	193.2	35.4	187.8	316.6	22.1	93.7	29.3
10	10503.5	81.3	7.8	623.3	623.1	0.5	796.2	7810.9
11	34.7	0.8	41.7	1.3	11.6	29.7	42.5	0.2
12	3.6	124.4	42.9	153.5	253.7	3.2	57.4	217.5
13	40.8	2.6	53.4	1.9		26.0	216516.3	0.6
14	50.5	2.5	51.2	1.8	25.3	29.3	2558.8	0.6
15	162.8	7.5	121.6	1.6	20.7	155.9	48.1	0.4
16	54.7	38.4	10.4	39.0	62.7	3.7	37.3	37.6
17	0.8	51.6	31.0	56.8	48.4	3.1	18.0	56.1
18	68.8	5.7	51.4	1.8	163.2	53.3	45.5	0.3
19	4.6	9.6	41.4	9.5	10.6	3.0	22.2	1450.8
20	8.0	83.4	127.4	82.5	86.1	89.6	77.6	81.8
21	19.9	125.5	52.5	58.5	83.9	21.2	142.6	59.0
22	23.0	4.7	10.7	1.1	3.9	14.0	20.3	0.3
23	8.1	15.0	31.3	12.3	27.6	7.6	27.8	11.4
24	1.5	24.6	34.3	29.5	25.9	0.2	28.4	0.4
25	5.5	17.4	49.1	22.2	415.4	16.9	6.3	20.3
26	48.5	135.1	19.4	149.4	506.1	83.8	4.7	115.8
27	1.3	238.6	71.3	280.4	1772.6	35.3	20.2	0.2
28	2230.4	2.7	22.9	0.9	630.7	51.2	28.3	0.3
29	6.0	45.0	1241.6	806.2		1238.3	18.7	41.0
30	48.7	58.1	25.6	56.7	60.2	57.5	39.3	55.3
31	29.8	142.7	68.9	146.2	274.5		64.1	144.3
32	24.1	8.7	19.5	7.2	37.5	15.9	24.0	61.2
33	164.1	17.3	161.6	2.1	739.4	152.1	427.4	0.3
34	2.5	74.3	21.2	71.2	25.7	0.3	15.6	189.3
35	60.4	64.0	20.3	66.4	189.2	61.0	109.5	0.2
36	690.6	678.4	662.8	700.0	717.4		2134.9	0.4
37	59.6	44.1	17.2	1.2	658.7	41.2	68.7	0.4
38	40.2	68.2	1175.9	66.2	1038.2	1176.1	180.4	67.2
39	1.5	40.3	38.3	42.6	277.8	0.3	17.2	283.9
40	1.1	91.2	31.2	10785.5	5539.7	0.6	45.0	2636.6
41	14.8	40.7	12.4	43.9	53.4	11.7	36.0	13.7
42	4.2	90.9	16.8	89.9	265.0	8.2	67.0	88.8
43	39.3	106.5	22.4	111.7	108.6	0.1	44.3	109.4
44	1.3	91.4	770.2	75.9	9593.7	13.4	66.4	0.4
45	20.0	5.9	60.8	1.2	8.0	29.0	19.3	0.4
46	760.7	6.6	6.2	414.6	405.6	0.6	10427221.5	828.7
47	10.4	55.6	46.4	59.3	62.1	38.2	52.2	10.7
48	26.7	3.4	45.4	1.3	16.2	62.1	59.9	0.4
49	24.4	1.5	7.2	1.4	434.5	6.6	22.3	0.3
50	28.1	0.1	32.9	1.0	4.9	6.1	30.5	0.4
Mean	27.4	53.6	39.9	58.9	159.8	21.6	46.8	11.4

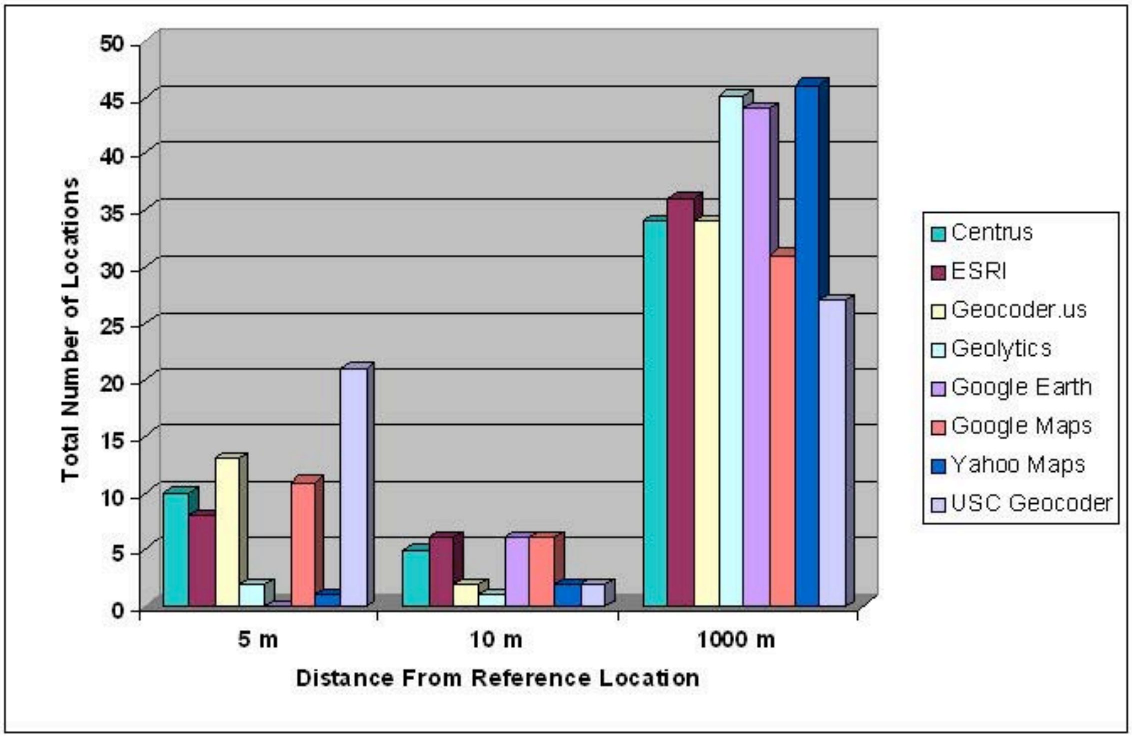


Figure 24 Summary of number of points within 5, 10 and 1 km distance from the reference data locations, as listed in Table 6

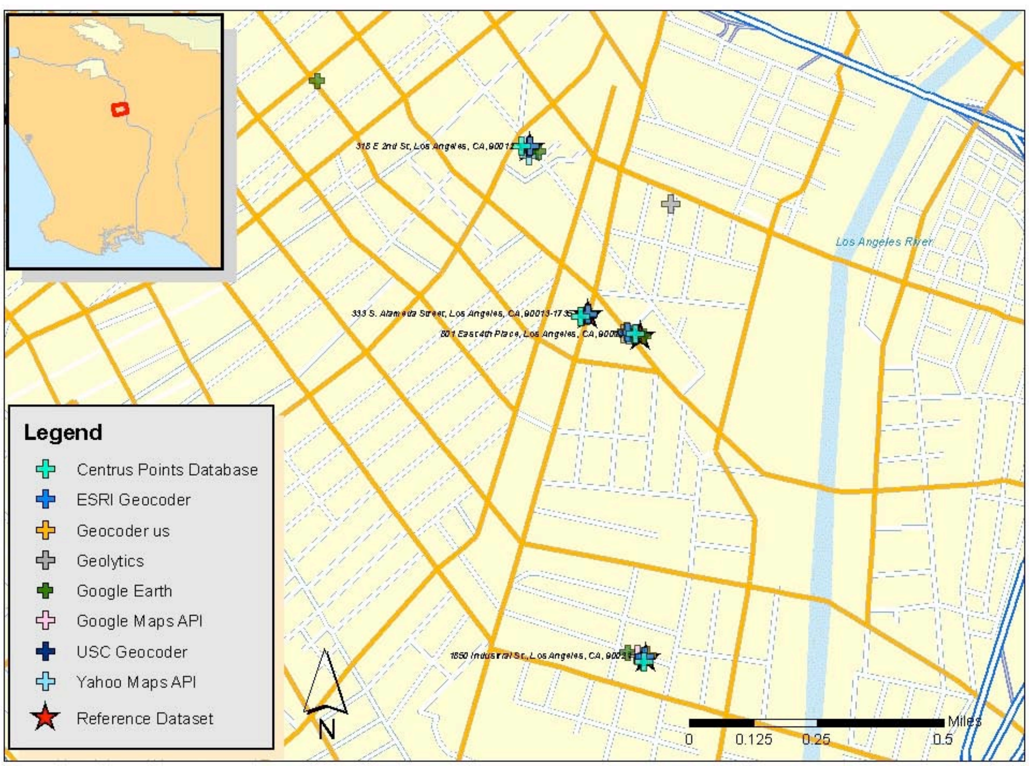


Figure 25 Partial view of all software tested utilizing geocoding software listed in Table 1

4 Summary and Conclusions

This research report is the third in a series of three reports on best geocoding practices and presents the results of an in-depth investigation of eight currently available geocoding platforms. The geocoders evaluated span the gambit in terms of cost, functionality, customization, and reporting capability. Depending on the needs, requirements, technical ability and budget, each of these geocoders represent suitable options for a geocoding solution. No one geocoder stands out as above and beyond the others, with each having their strong and weak points, as is clearly evident from Table 4. All of the geocoders performed the best in certain circumstances, and likewise all of the geocoders performed the worst in certain circumstances. Many of these cases were due to the nature of the geocoder implementations in terms of the reference data sources utilized, the specific matching strategies included, and/or the interpolation procedures utilized. While many of the cases that resulted in less than perfect results occurred simultaneously with each of the geocoders, there was usually was a single geocoder that succeeded when the others failed. Likewise, on addresses where most of the geocoders succeeded, there are several cases where a single geocoder did not. Most striking is that this disparity even occurs between two supposedly equivalent products, Google Earth and Google Maps.

The results indicate that there are indeed some patterns as to when the geocoders fail or succeed. Some of these were expected, some were not. For instance, in the cases where the input address was perfectly correct and the address existed in the reference data sources, the geocoders that utilized data sources that were more accurate, e.g. Tele Atlas (if price is used as a proxy for quality), produced more accurate results. In contrast, the results suggest that when the input data are incorrect or the reference data are incomplete or inaccurate, the simplest methods may be the most effective, e.g. deterministic feature matching and address range interpolation with USCB TIGER/Line files out performs more complex strategies and higher quality Tele Atlas reference data.

Perhaps most importantly, the findings indicate that price alone is not a good guide to geocode accuracy. The most expensive geocoder tested, Centrus, resulted in USCB misclassifications on several input addresses, all of which were handled properly by the less expensive commercial versions (ESRI) as well as the free online web services (e.g. Google, Yahoo, and/or Geocoder.us). However, these free versions were more prone to produce erroneous results in other cases where the Centrus geocoder got the USCB classification correct. Likewise, the open source USC Geocoding Platform clearly has some shortcomings with regard to functionality (i.e. the lack of probabilistic matching), which caused it to fall short in cases where both the free and paid commercial geocoders succeeded.

As anticipated, it is clear that each of these geocoding systems offer unique and/or innovative functionalities that could be utilized in the development of a comprehensive and user-friendly geocoder geared toward serving the cancer research community (Goldberg et al. 2008a, 2008b). Taken as a whole, the geocoders tested were capable of geocoding every address correctly such that there was at least one geocoder for each record that did not misclassify the resulting USCB associated values. This evidence may support the conclusion that perhaps the best approach to improving geocoding results would be a composite geocoding solution borrowing from the strengths of each of the geocoders surveyed in this report. Future related studies should continue evaluating additional geocoding software such as Microsoft Virtual Earth, MapInfo MapMarker, MapQuest and TerraServer, and others that may be commonly used by the cancer research community, as well as increasing the coverage of input data to look at other parts of the US. The latter would provide a basis for comparison and extrapolation of geocoder evaluation results across the country.

5 Acknowledgements

The development of this research report was funded by the CDC. It draws upon the work of Goldberg (2008a, 2008b, 2008c), which has been funded in part with Federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, and from the Centers for Disease Control and Prevention.

6 References

Bakshi R, Knoblock CA, and Thakkar S 2004 Exploiting Online Sources to Accurately Geocode Addresses. In Pfoser D and Cruz IF (eds) *Proceedings of the 12th ACM International Symposium on Advances in Geographic Information Systems* pp 194-203

Canonical Ltd., 2008 *Ubuntu Home Page*. WWW document, http://www.ubuntu.com (July 15th 2008)

Erle S, Walsh J 2008 *Geo::Coder::US*. WWW document, http://search.cpan.org/~sderle/Geo-Coder-US/US.pm (July 15th 2008)

ESRI 2003 *ArcGIS 9 Geocoding Rule Base Developer Guide*. ESRI. WWW document, webhelp.esri.com/arcgisdesktop/9.2/pdf/Geocoding_Rule_Base_Developer_Guide.pdf (July 15th 2008)

ESRI 2008a *Defining the Address Locator Components*. ArcGIS 9.2 Desktop Help. WWW document, http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Defining_the_address_locator_components (July 15th 2008)

ESRI 2008b *StreetMap USA*. ESRI Data & Maps Media Kit, ESRI Data & Maps/StreetMap USA DVD

Geolytics 2008 *Geolytics GeocodeCD User Guide*. East Brunswick, NJ, 10 pages

Goldberg DW 2008a *A Geocoding Best Practices Guide*. In preparation. Springfield, IL North American Association of Central Cancer Registries. 245 pages

Goldberg DW 2008b *Free Geocode Correction Services*. WWW document, https://webgis.usc.edu/Services/Correction/ (July 15th 2008)

Goldberg DW 2008b *Geocoding Services*. WWW document, https://webgis.usc.edu/Services/Geocode/ (23 June 2008)

Goldberg DW, Swift JN, and Wilson JP 2008a *Geocoding Best Practices: Reference Data, Input Data and Feature Matching*. Los Angeles, CA, University of Southern California GIS Research Laboratory Technical Report No 8

Goldberg DW, Swift JN, and Wilson JP 2008b *Geocoding Best Practices: Geocoding User Requirements Analysis*. Los Angeles, CA, University of Southern California GIS Research Laboratory Technical Report No 9

Google 2008a *Google Earth User Guide*. WWW document, http://earth.google.com/userguide/v4/ (July 15th 2008)

Google 2008b *Welcome to Google Earth*. WWW document, http://earth.google.com/ (July 15th 2008)

Google 2008c *Google Code*. Developer Resources and Featured News. WWW document, http://code.google.com/ (July 15th 2008)

Google 2008d *Google Maps API Reference*. WWW document, http://code.google.com/apis/maps/documentation/reference.html#GGeoAddressAccuracy (July 15th 2008)

Google 2008e *Google Earth Help Center*. Data and Imagery. WWW document, http://earth.google.com/support/bin/topic.py?topic=13327 (July 15th 2008)

Group 1 Software Inc. 2008a *Centrus - Business Geographics, Data Quality, Real-Time Customer Matching*. WWW document, http://www.centrus.com (July 15th 2008)

Group 1 Software Inc. 2008b *Centrus Desktop User's Guide*. Software Release 4.07.00.L. April 2008. Pitney Bows Corp, 118 pages

Hofferkamp J and Havener L (eds) 2008 *Standards for Cancer Registries: Data Standards and Data Dictionary, Volume II* (12th Edition). Springfield, IL North American Association of Central Cancer Registries

Krieger N Waterman P Chen JT 2002a Geocoding and monitoring of US socioeconomic inequities in mortality and cancer incidence: does the choice of area measure and geographic level matter? The Public Health Disparities Geocoding Project. *American Journal of Epidemiology* 156(5) pp 471-482

Krieger N Waterman P Chen JT Soobader MJ Subramanian SV Carson R 2002b Zip code caveat: bias due to spatiotemporal mismatches between ZIP codes and US Census-defined geographic areas. The Public Health Disparities Geocoding Project. *American Journal of Epidemiology* 92(7) pp 1100-1102

Krieger N Waterman P Lemieux KS Zierler S Hogan JW 2001 On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *American Journal of Public Health* 91(7) pp 1114-1116

Locator Technologies 2008 *Geocoder.us/.NET - Find the latitude & longitude of any US address - for free.* WWW document, http://geocoder.us/ (23 June 2008)

MySQL AB 2008 *MySQL: The world's most popular open source database*. WWW document, http://www.mysql.com (July 15th 2008)

NASA 2006 *LANDSAT Data Base Map for Google Earth*™. WWW document, http://landsat.gsfc.nasa.gov/news/news-archive/dyk_0002.html (July 15th 2008)

NAVTEQ 2008 *NAVSTREETS*. WWW document, http://developer.navteq.com /site/global/dev_resources/170_navteqproducts/navdataformats/navstreets/p_navstreets.jsp (July 15th 2008)

PostgreSQL Global Development Group 2008 *PostgreSQL: The world's most advanced open source database*. WWW document, http://www.postgresql.org (July 15th 2008)

Tele Atlas Inc. 2008a *Dynamap Map Database*. WWW document, http://www.teleatlas.com/OurProducts/MapData/Dynamap/index.htm (July 15th 2008)

Tele Atlas Inc. 2008b *MultiNet Map Database*. WWW document, http://www.teleatlas.com/OurProducts/MapData/Multinet/index.htm (July 15th 2008)

United States Census Bureau 2008a *American Community Survey*. WWW document, http://www.census.gov/acs. (July 15th 2008)

United States Census Bureau 2008b *MAF/TIGER Accuracy Improvement Project*. WWW document, http://www.census.gov/geo/mod/maftiger.html (July 15th 2008)

United States Postal Service 2008a *Address Information System Products Technical Guide*. WWW document available online at: http://ribbs.usps.gov/files/Addressing/PUBS/AIS.pdf (July 15th 2008)

United States Postal Service 2008b *CASS Mailer's Guide*. WWW document available online at: http://ribbs.usps.gov/doc/cmg.html (July 15th 2008)

United States Postal Service 2008c *Publication 28 – Postal Addressing Standards*. WWW document available online at: http://pe.usps.com/text/pub28/welcome.htm (July 15th 2008)

Weisstein EW 2008 *Great Circle. MathWorld--A Wolfram Web Resource*. WWW document, http://mathworld.wolfram.com/GreatCircle.html (July 8th 2008)

Whitsel EA Rose KM Wood JL Henley AC Liao D. and Heiss G. 2004 Accuracy and Repeatability of Commercial Geocoding. *American Journal of Epidemiology* 160(10) pp 1023-1029. WWW document, http://aje.oxfordjournals.org/cgi/reprint/160/10/1023 (July 15th 2008)

Yang D-H Bilaver LM Hayes O and Goerge R 2004 Improving Geocoding Practices: Evaluation of Geocoding Tools. *Journal of Medical Systems* 28(4) pp 361-370

Yahoo 2008 *Yahoo! Developer Network*. Yahoo! Maps Web Services - Geocoding API. Finding Latitudes and Longitudes. WWW document, http://developer.yahoo.com/maps/rest/V1/geocode.html (July 15th 2008)

Zhan FB Brender JD Lima IDE Suarez L and Langlois PH 2006 Match Rate and Positional Accuracy of Two Geocoding Methods for Epidemiologic Research. *Annals of Epidemiology* 16(11) pp 842-849

7 List of Terms

Abbreviation	Description
CASS	Coding Accuracy Support System
CDC	Centers For Disease Control and Prevention
DEM	Digital Elevation Model
MCD	Minor Civil Division
NAACCR	North American Association of Central Cancer Registries
NGC	Northrop Grumman Corporation
SRTM	Shuttle Radar Topography Mission
TIGER	Topographically Integrated Geographic Encoding and Referencing
USCB	United States Census Bureau
USPS	United States Postal Service
ZCTA	ZIP Code Tabulation Area

Appendix 1 List of Currently Available Geocoders

Name	Reference URL	Application	Commercial/ Open Source	Coverage
Ajmsoft	http://www.ajmsoft.com/ac/geocode.php	Web	Open Source	US
Ajmsoft	http://www.ajmsoft.com/ac/geo/GeoPE.php	PC	Open Source	US
AltaMap Geocoder - Geomicro - AltaMap Desktop Professional and	http://www.geomicro.com/capabilities/geocoding.asp	PC	Commercial	US and Canada
Alteryx - SRC	http://www.extendthereach.com/products/alteryx_overview.srct	PC	Commercial and Open Source	US or World
Andre Lewis and Bill Eisenhauer - GeoKit	http://geokit.rubyforge.org/ and http://geokit.rubyforge.org/api/index.html	PC	Open Source	US or World
android.location.Geocoder	http://code.google.com/android/reference/android/location/Geocoder.html	PC	Open Source	US or World
ArcMap - ESRI	http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=1740&pid=1738&topicname=Geocode_Addresses_(Geocoding) and http://www.lib.unc.edu/reference/gis/faq/geocode.html	PC	Commercial	World
ArcWeb Services - ESRI	http://www1.arcwebservices.com/v2006/develop.jsp	PC	Commercial	World
Batch Geocode	http://www.batchgeocode.com/	Web	Open Source	World
Brian Beck - geopy	http://exogen.case.edu/projects/geopy/	PC	Open Source	World
Centrus - for ESRI ArcGIS - Centrus Desktop and Geostan API	http://www.centrus.com/ and http://www.centrus.com/documents/arcgis.pdf	PC	Commercial	US and Canada
CLEAN_Address() - Runner Technologies	http://www.runnertechnologies.com/clean_addr_features.html	Web	Commercial	US, Canada, 240 countries
Dan Egnor	http://ofb.net/%7Eegnor/google.html	PC and Web	Open Source	World
DOTS Address Geocode - US	http://www.serviceobjects.com/products/geocode_web_service.asp?zut=ggl1030	Web	Commercial	US
Excel Geocoding Tool v3x	http://www.juiceanalytics.com/writing/excel-geocoding-tool-v2/	PC	Open Source	World
Explorer Geocoder - SRC	http://www.extendthereach.com/products/OpenSourceGeocoder.srct	PC	Open Source	US
EZ-Locate	http://www.geocode.com/	PC and Web	Commercial	US and Canada
Fannie Mae Property GeoCoder	https://www.efanniemae.com/sf/refmaterials/geocoder/	Web	Open Source	US
FFIEC Geocoding System	http://www.ffiec.gov/Geocode/default.aspx	Web	Open Source	US

Appendix 3 cont.

Name	Reference URL	Application	Commercial/ Open Source	Coverage
GeoBase	http://www.geobase.info/spatial-functions/address-lookup/address-	PC	Commercial	World
Geocode.Service	http://sourceforge.net/projects/geocode/	PC	Open Source	US and World
GeocodeDVD and Geocode API Toolkit - Geolytics, Inc.	http://www.geolytics.com/USCensus,Geocoding-Products,Categories.asp	PC and Web	Commercial	US
GeoCoder Object - Melissa Data.com	http://www.melissadata.com/geocoder/geocoderobject.htm	PC	Commercial	US
Geocoder.us	http://geocoder.us/	Web	Commercial and Open Source	US
GeoIPTC - Kalimages	http://www.geoiptc.com/EN/Index.html	PC	Commercial	World
GeoNames	http://www.geonames.org/export/free-geocoding.html	Web	Open Source	World
GeoPinPoint Suite ActiveX/Java/Linux/Unix/Windows	http://www.dmtispatial.com/geocode/index.html	PC	Commercial	Canada
GIS Dynamics	http://www.gisdynamics.com/services/mapping/?gclid=CI2s-dfu25MCFRpciAodS299YQ	PC	Commercial	US and World
Global Mapper	http://globalmapper.com/	PC	Commercial	US
Google Earth	http://earth.google.com/#utm_campaign=en&utm_source=en-ha-na-us-google&utm_medium=ha&utm_term=geocoding	Web	Open Source	World
Google Earth Pro	http://earth.google.com/enterprise/earth_pro.html	Web	Commercial	World
Google Maps	http://maps.google.com	Web	Open Source	World
GPS Insight Map Book Tool	http://www.gpsinsight.com/blog/?p=164?source=google&gclid=CPigweux4JMCFR0ZagodTAVoVw	Web	Open Source	World
GPSVisualizer	http://www.gpsvisualizer.com/geocoding.html	Web	Open Source	World
HoudahGeo	http://www.houdah.com/houdahGeo/	PC	Commercial	World
iBegin	http://geocoder.ibegin.com/	PC	Commercial and Open Source	US and Canada
iMapstools U.S. Geocoder	http://imaptools.com/geocode-us.html	PC	Commercial	US, Canada and world
Instant Geocoder - Questsoft	http://www.questsoft.com/instant_geocoder.php	PC	Commercial	US
iTouchMap.com	http://itouchmap.com/latlong.html	Web	Open Source	World
John Coryat - USNaviguide	http://maps.huge.info/geocoder/	Web	Open Source	World
LinkageWiz	http://www.linkagewiz.com/?gclid=CNLCyOf43ZMCFQwxiQodc04EZQ	PC	Commercial	SA, Canada, UK, France, Australia
Manifold	http://www.manifold.net/info/geocoder_comp.shtml	PC	Commercial	US, Canada, 11 countries

Appendix 3 cont.

Name	Reference URL	Application	Commercial/ Open Source	Coverage
Map Suite USA	http://gis.thinkgeo.com/Products/GISComponentsforNETDevelopers/Map	PC	Commercial	US
MapBlast - MSN Maps & Directions	http://www.mapblast.com/(lsfmcn45ct2qlzf24nyuvv45)/Home.aspx	Web	Open Source	World
MapBuilder	http://www.mapbuilder.net/	Web	Open Source	World
MapChannels Geocoder	http://www.mapchannels.com/geocoding.aspx	PC and Web	Open Source	World
Map-In-A-Box Solo 2006 - add-on to MapInfo Professional 6.0	http://www.mappingsolutions.com/product.asp?section=Products&page=Map-In-A-Box%20Solo%202006%20for%20MapInfo	PC	Commercial	US and Canada
MapInfo MapMarker for US	http://www.cmcus.com/Products/Vendors/MapInfo/mapmarker.asp?source=google&campaign=MapMarker&_kk=geocoder&_kt=379c5bf1-dc3a-45b2-b06d-c0f3be53f6e7	PC	Commercial	US
MapInfo MapMarker Plus for US / MapMarker Plus with Parcel Precision for the US / MapMarker Plus for Other Countries, i.e. MapMarker Plus Canada	http://www.cmcus.com/Products/Vendors/MapInfo/mapmarker.asp?source=google&campaign=MapMarker&_kk=geocoder&_kt=379c5bf1-dc3a-45b2-b06d-c0f3be53f6e7	PC	Commercial	US and Canada, and other countries"
MapMarker World	http://extranet.mapinfo.com/products/Overview.cfm?productid=1891	PC	Commercial	World
Mapquest	http://www.mapquest.com and http://help.mapquest.com/jive/entry.jspa?externalID=36&categoryID=4	Web	Open Source	US and World
Mapstraction	http://www.mapstraction.com/geocode.php?geocoder=mapquest&map=mapquest	PC	Open Source	World
Matthew Kanehttp	http://www.cs.indiana.edu/~markane/i590/geocoder.html	PC and Web	Open Source	US
MGeocoder: Google Maps API Extension for Geocoding	http://brainoff.com/gmaps/mgeocoder.html	PC	Open Source	World
Microsoft MapPoint 2006 Web Service	http://www.microsoft.com/mappoint/products/webservice/default.mspx	PC	Commercial	World
MS Virtual Earth	http://msdn.microsoft.com/en-us/library/cc161074.aspx and http://msdn.microsoft.com/en-us/library/bb545004.aspx	PC	Commercial	World
MyGeoPosition	http://www.mygeoposition.com/	Web	Open Source	World
NAC Geocoding/Address Matching Services - NAC Geographic	http://www.nacgeo.com/geocode.asp	PC and Web	Commercial	US + 25 Other Countries
NacGeo	http://www.nacgeo.com/geocode.asp	Web	Commercial	World
Ontok	http://www.ontok.com/api	Web	Open Source	US
PxPoint	http://www.proxix.com/Products/Software/Geocoding/	PC	Commercial	US
RubyForge Geocoder	http://rubyforge.org/projects/geocoder/	PC	Open Source	?

Appendix 3 cont.

Name	Reference URL	Application	Commercial/ Open Source	Coverage
SRC	http://www.extendthereach.com/products/OpenSourceGeocoder.srct	PC	Open Source	World
Stephen Morse - Converting Addresses to/from Latitude/Longitude in One Step	http://www.stevemorse.org/jcal/latlon.php	Web	Open Source	World
TerraServer - Web Service and Online Search	http://terraserver-usa.com/webservices.aspx	PC and Web	Open Source	US
Travel GIS	http://www.travelgis.com/geocode	Web	Open Source	World
worldKit Geocoder	http://worldkit.org/geocoder/ and http://worldkit.org/ and http://worldkit.org/doc/rss.php	PC	Open Source	US
Yahoo	http://developer.yahoo.com/maps/rest/V1/geocode.html	Web	Open Source	World
Zeemaps - Zee Source	http://www.zeemaps.com/geocoding.do	PC	Open Source	World

Appendix 2 Summary of Geocoded Addresses

Lat Centrus	Lon Centrus	Lat Google Earth	Lon Google Earth	Lat Google Maps	Lon Google Maps	Lat Yahoo	Lon Yahoo	Lat Geoco-der.us	Lon Geoco-der.us	Lat Geo-Lytics	Lon Geo-Lytics	Lat ESRI	Lon ESRI	Lat USC	Lon USC
34.038	-118.692	34.040	-118.693	34.040	-118.693	34.038	-118.692	34.038	-118.693	34.038	-118.694	34.038	-118.692	34.03706	-118.692
34.036	-118.685	34.036	-118.685	34.036	-118.685	34.036	-118.685	34.036	-118.685	34.035	-118.685	34.036	-118.685	34.0353	-118.685
34.087	-118.975	34.086	-118.980	34.086	-118.980	34.088	-118.977	34.089	-118.977	34.088	-118.973	34.088	-118.978	34.03045	-118.779
34.051	-118.645	34.051	-118.645	34.051	-118.645	34.051	-118.645	34.053	-118.644	34.053	-118.643	34.052	-118.644	34.0511	-118.645
34.029	-118.828	34.029	-118.828	34.029	-118.828	34.029	-118.827	34.029	-118.828	34.029	-118.827	34.029	-118.828	34.03045	-118.779
34.040	-118.658	34.040	-118.658	34.039	-118.658	34.039	-118.659	34.047	-118.945	34.039	-118.653	34.039	-118.655	34.0395	-118.658
34.014	-118.794	34.014	-118.794	34.014	-118.794	34.014	-118.794	34.014	-118.795	34.014	-118.795	34.014	-118.794	34.01443	-118.794
34.034	-118.692	34.034	-118.692	34.034	-118.692	34.034	-118.692	34.034	-118.693	34.034	-118.695	34.034	-118.693	34.03056	-118.704
34.017	-118.816	34.017	-118.816	34.017	-118.816	34.017	-118.817	34.016	-118.818	34.015	-118.819	34.016	-118.818	34.01666	-118.816
34.056	-118.805	34.035	-118.694	34.035	-118.694	34.037	-118.686	34.037	-118.688	34.037	-118.688	34.035	-118.693	34.03045	-118.779
33.790	-118.189	33.790	-118.189	33.790	-118.189	33.790	-118.189	33.790	-118.189	33.790	-118.189	33.790	-118.189	33.78993	-118.189
33.829	-118.214	33.829	-118.214	33.829	-118.214	33.829	-118.213	33.829	-118.212	33.827	-118.214	33.829	-118.212	33.827	-118.214
33.767	-118.197	33.767	-118.196	33.767	-118.196	32.664	-116.273	33.767	-118.197			33.767	-118.197	33.76741	-118.197
33.811	-118.163	33.812	-118.163	33.812	-118.163	33.803	-118.189	33.812	-118.164	33.812	-118.163	33.812	-118.163	33.8118	-118.163
33.842	-118.174	33.842	-118.175	33.842	-118.175	33.842	-118.176	33.842	-118.176	33.842	-118.176	33.842	-118.176	33.84174	-118.176
33.784	-118.135	33.783	-118.135	33.783	-118.135	33.783	-118.135	33.783	-118.135	33.783	-118.136	33.783	-118.135	33.78275	-118.135
33.863	-118.150	33.863	-118.150	33.863	-118.150	33.863	-118.150	33.863	-118.150	33.863	-118.150	33.863	-118.150	33.86268	-118.15
33.845	-118.175	33.845	-118.175	33.845	-118.175	33.844	-118.175	33.844	-118.175	33.843	-118.175	33.844	-118.175	33.84417	-118.175
33.749	-118.125	33.749	-118.125	33.749	-118.125	33.749	-118.125	33.749	-118.125	33.749	-118.125	33.749	-118.125	33.76112	-118.13
33.848	-118.186	33.848	-118.187	33.848	-118.187	33.848	-118.186	33.848	-118.186	33.848	-118.186	33.848	-118.186	33.84819	-118.186
34.197	-118.120	34.197	-118.120	34.197	-118.120	34.196	-118.120	34.197	-118.120	34.198	-118.120	34.196	-118.120	34.19714	-118.12
34.186	-118.131	34.186	-118.131	34.186	-118.131	34.186	-118.131	34.186	-118.131	34.186	-118.131	34.186	-118.131	34.18587	-118.131
34.198	-118.160	34.197	-118.160	34.198	-118.160	34.198	-118.160	34.197	-118.160	34.197	-118.160	34.197	-118.160	34.19748	-118.16
34.189	-118.132	34.189	-118.132	34.189	-118.132	34.188	-118.132	34.189	-118.131	34.188	-118.131	34.189	-118.131	34.18854	-118.132
34.190	-118.150	34.190	-118.149	34.190	-118.149	34.190	-118.150	34.190	-118.149	34.188	-118.146	34.190	-118.149	34.19032	-118.149
34.185	-118.122	34.185	-118.122	34.186	-118.121	34.185	-118.122	34.186	-118.123	34.189	-118.123	34.186	-118.123	34.18416	-118.122
34.208	-118.160	34.208	-118.159	34.208	-118.159	34.208	-118.160	34.210	-118.160	34.221	-118.149	34.210	-118.160	34.20751	-118.16

Appendix 2 **cont.**

Lat Centrus	**Lon Centrus**	**Lat Google Earth**	**Lon Google Earth**	**Lat Google Maps**	**Lon Google Maps**	**Lat Yahoo**	**Lon Yahoo**	**Lat Geoco-der.us**	**Lon Geoco-der.us**	**Lat Geo-Lytics**	**Lon Geo-Lytics**	**Lat ESRI**	**Lon ESRI**	**Lat USC**	**Lon USC**
34.196	-118.160	34.190	-118.137	34.190	-118.137	34.189	-118.137	34.190	-118.137	34.185	-118.133	34.190	-118.137	34.18955	-118.137
34.195	-118.132	34.184	-118.132	34.184	-118.132	34.195	-118.131	34.188	-118.131			34.195	-118.131	34.19469	-118.131
34.186	-118.132	34.186	-118.132	34.186	-118.132	34.186	-118.132	34.186	-118.132	34.186	-118.132	34.186	-118.131	34.18596	-118.132
34.700	-118.147	34.700	-118.148			34.699	-118.148	34.699	-118.148	34.697	-118.148	34.699	-118.148	34.69872	-118.148
34.698	-118.139	34.698	-118.139	34.698	-118.139	34.698	-118.139	34.698	-118.139	34.698	-118.139	34.698	-118.139	34.69799	-118.14
34.694	-118.236	34.694	-118.237	34.694	-118.237	34.692	-118.236	34.696	-118.237	34.689	-118.237	34.696	-118.237	34.69571	-118.237
34.668	-118.187	34.668	-118.187	34.668	-118.187	34.668	-118.187	34.669	-118.187	34.668	-118.187	34.669	-118.187	34.66961	-118.187
34.697	-118.133	34.697	-118.133	34.697	-118.133	34.697	-118.132	34.697	-118.132	34.698	-118.132	34.697	-118.133	34.69684	-118.133
34.676	-118.133	34.676	-118.133			34.699	-118.145	34.676	-118.133	34.675	-118.133	34.676	-118.133	34.68175	-118.134
34.672	-118.160	34.672	-118.160	34.672	-118.160	34.672	-118.159	34.671	-118.160	34.666	-118.160	34.672	-118.160	34.67146	-118.16
34.679	-118.162	34.690	-118.164	34.689	-118.164	34.679	-118.165	34.679	-118.163	34.674	-118.172	34.679	-118.163	34.6788	-118.163
34.703	-118.154	34.703	-118.154	34.703	-118.154	34.703	-118.154	34.703	-118.154	34.703	-118.157	34.703	-118.154	34.70271	-118.157
34.666	-118.167	34.666	-118.167	34.666	-118.167	34.666	-118.166	34.762	-118.176	34.631	-118.210	34.666	-118.166	34.68698	-118.154
34.045	-118.236	34.045	-118.236	34.045	-118.236	34.045	-118.237	34.045	-118.237	34.045	-118.237	34.045	-118.237	34.04462	-118.236
34.050	-118.265	34.050	-118.265	34.050	-118.265	34.051	-118.265	34.051	-118.264	34.051	-118.262	34.051	-118.264	34.05062	-118.264
34.054	-118.256	34.054	-118.255	34.054	-118.255	34.055	-118.255	34.055	-118.255	34.055	-118.255	34.055	-118.255	34.0551	-118.255
34.044	-118.254	34.050	-118.249	34.044	-118.254	34.045	-118.254	34.045	-118.254	33.960	-118.278	34.045	-118.254	34.04397	-118.254
34.036	-118.233	34.036	-118.234	34.036	-118.234	34.036	-118.233	34.036	-118.233	34.036	-118.233	34.036	-118.233	34.0358	-118.233
34.053	-118.263	34.047	-118.259	34.047	-118.259	34.049	118.256	34.050	-118.256	34.049	-118.256	34.047	-118.259	34.05256	-118.264
34.042	-118.264	34.042	-118.265	34.042	-118.265	34.042	-118.264	34.042	-118.265	34.043	-118.265	34.042	-118.264	34.04203	-118.264
34.051	-118.260	34.050	-118.260	34.050	-118.260	34.050	-118.260	34.051	-118.260	34.051	-118.260	34.051	-118.260	34.0507	-118.26
34.045	-118.238	34.045	-118.238	34.045	-118.238	34.045	-118.238	34.045	-118.238	34.048	-118.236	34.045	-118.238	34.04485	-118.238
34.049	-118.242	34.049	-118.241	34.049	-118.241	34.049	-118.241	34.049	-118.241	34.049	-118.241	34.049	-118.241	34.04909	-118.241

Lat = Latitude (in decimal degrees)
Lon = Longitude (in decimal degrees)