

UC Santa Barbara

Core Curriculum-Geographic Information Systems (1990)

Title

Unit 29 - Discrete Georeferencing

Permalink

<https://escholarship.org/uc/item/35j897pz>

Authors

Unit 29, CC in GIS
National Center for Geographic Information and Analysis

Publication Date

1990

Peer reviewed

UNIT 29 - DISCRETE GEOREFERENCING

UNIT 29 - DISCRETE GEOREFERENCING

- [A. INTRODUCTION](#)
- [B. STREET ADDRESS](#)
 - [Using addresses in GIS](#)
 - [Method](#)
 - [Example - Addmatch using TIGER](#)
- [C. POSTAL CODE SYSTEMS](#)
 - [US ZIP Codes](#)
 - [Canadian Postal Code](#)
 - [Problems](#)
- [D. US PUBLIC LAND SURVEY SYSTEM](#)
 - [PLSS References](#)
- [E. GEOLOC GRID](#)
 - [GEOLOC References](#)
 - [Precision](#)
- [F. CENSUS SYSTEMS](#)
 - [Converting to georeferences](#)
- [G. ISSUES CONCERNING DISCRETE GEOREFERENCING](#)
 - [Hooks](#)
 - [Purpose](#)
 - [Standardization](#)
- [REFERENCES](#)
- [DISCUSSION OR EXAM QUESTIONS](#)
- [NOTES](#)

This lecture concludes the module on geocoding. Several important practical issues are raised here that will be important particularly for those who will be working with economic and demographic databases.

UNIT 29 - DISCRETE GEOREFERENCING

A. INTRODUCTION

- the georeferencing methods covered so far (latitude- longitude, Cartesian, projections from latitude/longitude to the plane) are continuous
 - this means that there is no effective limit to precision, as coordinates are measured on continuous scales
- will now look at discrete methods - systems of georeferencing for discrete units on the earth's surface
- many of these methods are indirect
 - this means that the method provides a key or index, which can then be used with a table to determine latitude/longitude or coordinates
 - for example: a Zip code is an indirect georeference
 - rather than give latitude/longitude for a place directly, it provides a unique number which can be looked up on a map if coordinates are needed
- because these methods are indirect, it is important to consider the precision of these systems
 - precision is related directly to the size of the discrete unit which forms the basis of the georeferencing system
- many methods of indirect or discrete georeferencing are in common use
 - following are 5 of the most common

B. STREET ADDRESS

- the precision of street addresses as georeferences varies:
 - is highest for apartments or houses in cities
 - is lowest for rural addresses or post office box numbers, where the address may indicate only that the place is somewhere in the area served by the post office

Using addresses in GIS

- general approach is to match address to a list of streets (called address matching or "addmatch")
 - spelling and punctuation variations make this difficult
 - e.g., Ave. or Avenue, apartment number before or after street number
 - a failure rate of 10% is regarded as good, 40% is not uncommon. In such cases it is necessary to find the street by hand, which may take as much as 5 minutes per address in large cities

Method

1. identify the block containing address from table of address ranges in each block
 - i.e., 551 B St. lies in the block running from 501 to 599

2. estimate position of house using the coordinates of the end points
 - the exact position of the house can be estimated by linear interpolation
 - i.e., 551 is roughly half way down the block
 - such estimates are crude
 - in many countries (e.g. India) addresses are not sequential along the street, but reflect date of construction
 - if the street is curved the estimate can be improved by using intermediate points (called shape points)
 - shape points are associated with the same information that block endpoints have, including building numbers and other georeferences
- databases to support addmatching exist in most industrialized countries.
 - in the US, DIME files were developed for this purpose in the late 1960s by the Bureau of the Census, and are now being replaced by more comprehensive TIGER files
 - see Unit 8 for an introduction to TIGER

handout - TIGER system: An overview (9 pages)

Example - Addmatch using TIGER

handout - Map of west central Columbia, MO

- note: intersection of West Blvd and W. Broadway is W of center

handout - Portion of the TIGER file (Boone County, MO)

demonstration - the solution of this problem could be demonstrated using the TIGER file for Boone County, MO

- TIGER files can be readily accessed and displayed using the SAFARI package from Geographic Data Technologies, Inc

Problem: find the latitude and longitude of 950 West Broadway

Procedure:

1. search the TIGER file for Boone County for features with the name "West Broadway" or equivalent (W. Broadway, Broadway W. etc)
 - get about 30 matches for the length of W. Broadway
2. find the record that lists the address range which includes address 950:
 - record #6714 covers the block from Greenwood to West Blvd, and includes the following data:
 - longitude 92.3503 to 92.3527
 - latitude 38.9519 to 38.9522 (indicating that the street has been coded from east to west)
 - ZIP code 65203 on both sides
 - census tract 6 on the left side, 7 on the right

- address ranges 900 to 998 on the left, 901 to 999 on the right
 - no shape points, so we assume the block is straight
- 3. determine the coordinates of number 950:
 - assume that the houses are evenly spaced along the street, and that the full range of addresses is used (this is not necessarily a good assumption, but it's the best that can be done without more information).
 - longitude is: $92.3503 + \{(950-900) * (92.3527-92.3503) / (998- 900)\} = 92.3515$
 - latitude is: $38.9519 + \{(950-900) * (38.9522-38.9519) / (998- 900)\} = 38.9521$
 - note that the results are given to the same precision as the block endpoints
 - we could have calculated more digits, but they would have been meaningless given the accuracy of the inputs
- problems with determining georeferences by address matching:
 - cases where matching fails (10 - 40% common)
 - rural areas and box numbers where there are no street addresses
 - long blocks with uneven houses
 - street addresses do not always identify a parcel or lot, and some parcels have many street addresses (e.g., apartments, condominiums)
- address matching is very commonly used to determine georeferences for marketing and retailing, health and the collection of social statistics

C. POSTAL CODE SYSTEMS

- postal code systems have been set up in many countries
- these often provide a high level of spatial precision

US ZIP Codes

- in the US, zip codes are designed to assist with mail sorting and delivery
 - the codes are hierarchically nested, states are uniquely identified by one or more sets of the first 2 numbers
 - a 5 digit ZIP code identifies the area served by a single post office
 - this gives precision of many city blocks
 - the 9 digit ZIP potentially provides a much higher level of spatial resolution, but problems exist
 - buildings may have different codes for different floors
 - overlapping and fragmented boundaries

Problems:

- addresses associated with a single zip code were developed from lists of addresses representing postal walks, rather than from maps. Addresses were seen as points along the streets rather than parcels of land
 - as a result, the area associated with a single 5 digit zip code does not necessarily have a well- defined geographical boundary
 - therefore, areas are sometimes not well defined, and they may overlap

- it is possible for the faces of a city block to have different ZIPs
 - boundaries of the zip code areas have been interpolated and files giving the coordinates of 5 digit ZIP code boundaries are available from a number of vendors
- warning: some of these have used simple Thiessen polygons to delineate associated areas
 - i.e. the area of a ZIP code has been defined as the area closest to the corresponding post office, instead of the true area

overhead - Rennie's ZIP code map of Los Angeles

- note unusual shapes of zones and boundaries

Canadian Postal Code

- the first 3 digits of the Canadian postal code define a Forward Sortation Area which is a useful unit for mapping (average population around 20,000) and is hierarchically nested within provinces
- the full 6 digits provide resolution of a few block faces
- files exist which allow the 6 digit code to be converted to census reporting zones and latitude/longitude

Problems

- postal code systems have great potential as discrete georeferences
 - however, they have not been designed for this purpose, hence the problems noted above
 - since their purpose is, in principle, internal to the postal system, it is also difficult to ensure stability through time (codes frequently change)
- however, there is great demand for statistics based on postal georeferences because of their applications in retailing and marketing and the ease with which they can be merged with customer account data

D. US PUBLIC LAND SURVEY SYSTEM

- PLSS is the basis for land surveys and legal land description over much of the US
 - unlike the previous systems, it is designed to reference land parcels
- because it is a comprehensive, systematic approach it is possible to use it as a georeference
 - commonly used by agencies such as the Bureau of Land Management and the US Forest Service, and within the oil and gas industry.
- packages exist to convert PLSS descriptions to latitude/longitude

PLSS References

handout - US public land survey system (not included, see Strahler and Strahler 1987, pp. 485-487).

- begin with a surveyed principal Meridian, several of which were laid out as north-south baselines in the Western US
- the area on both sides of the meridian is then blocked off in 6 mile by 6 mile areas, identified by township and range numbers
 - since this is a square grid system the township and ranges must be offset as one moves NS along the meridians
- the 36 square mile sections within each township are numbered from the top in a standard order
- each section is divided into four quartersections, and these can be further divided if higher spatial resolution is needed, as for example in describing the location of an oil well
- PLSS is most effective where the simple rules were followed closely, however:
 - much of the Northeast was settled long before the advent of the PLSS
 - there are major variations in the Southwest where the PLSS runs up against areas of early Spanish land tenure
 - errors in the early surveys have become embedded in the system and must be replicated in packages which offer PLSS to latitude/longitude conversion

E. GEOLOC GRID

handout - GEOLOC description (3 pages)

- an elaborate and more systematic example is provided by the GEOLOC geographical referencing system (see Whitson and Sety, 1987), which can be used to index every 100 acre parcel in the continental US

GEOLOC References

- the first level of partition consists of 2 rows and 3 columns, each partition or tile being 25 degrees of longitude by 13 degrees of latitude
 - these tiles are ordered row by row from the top left (Pacific Northwest) and numbered 1 to 6
- at the next level, each tile is divided into 26 rows of one half degree latitude and 25 columns of one degree longitude, the area covered by one 1:100,000 USGS quadrangle.
 - each of these subtitles is given a two letter designation using a letter to represent the row (A through Z) and one to represent the column (A through Y)
- each subtile is divided into 4 rows and 8 columns of 7.5 minute quads, numbered row by row from 1 to 32

- at the next level, these are divided into 4 rows and 2 columns, designated by assigning the letters A through H row by row
- finally, each of these divisions is divided into 5 rows, lettered A through E, and 10 columns numbered 0 through 9 to produce 50 cells of approximately 100 acres each
- an example of a full designator for a 100 acre parcel (in the Los Angeles area) is 4FG19DC6

Precision

- hierarchically nested systems like GEOLOC, and to some extent PLSS, allow the user to vary spatial precision depending on the application
 - 4FG19 would identify a 7.5 minute quadrangle, or an area roughly 9 miles across
 - the full 4FG19DC6 gives an area roughly 2000 ft across

F. CENSUS SYSTEMS

- note: this topic was introduced in Unit 8; a handout and overhead from that unit are reproduced here
- the major source of social and economic data in many countries is the Census
- statistics are collected and reported using a complex system of several different types of reporting zones:
 - political or administrative units used for reporting (province, county, city, electoral district)
 - units defined for ease of data collection (block, block group, enumeration district) but often too small to use for data reporting due to privacy regulations
 - units designed to be homogeneous for ease of analysis (census tract)
- in the US the major units are:
 - block group (formerly enumeration district)
 - the smallest reporting unit, about 1000 population
 - census tract
 - primarily in large cities, about 5000 population, intended for analysis
 - Minor civil division (mostly on township boundaries)
 - County
 - State

overhead - Hierarchy of census areas, 1990 census

handout - US census units

Converting to georeferences

- for the larger units, the main method of converting from census zone to georeference is through boundary files, which are digitized boundaries established for most of the major units and readily available from vendors or the Bureau

- for a smaller unit such as the block group (formerly ED) it is often possible to obtain from the Census Bureau a representative point or centroid which can be used as a georeference
 - for units with uneven population distribution the centroid may be located in the area of highest population density

G. ISSUES CONCERNING DISCRETE GEOREFERENCING

Hooks

- is useful to consider how many different reference systems are related to specific datasets
 - i.e. TIGER has street addresses, census zones and lat/long associated with each record
 - allows linking of many different data sources

Purpose

- many of these systems were set up for special purposes, and have only later become the basis for general georeferencing
 - e.g. post office does not have a mandate to maintain these systems for georeferencing purposes, therefore will only add ZIPs when mail is delivered to the location
 - zones may change without notice or record
 - e.g. census is only updated every 10 years
- as a result, these systems do not necessarily have "quality control" in the georeferencing sense
 - no agency maintains a file of new addresses

Standardization

- general purpose systems such as GEOLOC use regular divisions of the earth's surface, while special purpose systems tend to use irregular divisions
- in the past, efforts have been made to impose greater regularity on discrete georeferences
 - e.g. "gridiron" system of rectangular street networks (Washington, DC)
 - in the last century some city names were changed so that no two places in a single state had the same name
 - introduction of the ZIP code
- however, such standardization efforts generally are not consistent or long-term
 - rectangular street networks are no longer in fashion
 - referencing systems such as PLSS are now fairly chaotic despite simple principles
 - ZIPs are not consistent
- given their usefulness, is it possible to set up a single, common system of discrete

georeferencing?

REFERENCES

Strahler, A.N. and A.H. Strahler, 1987. *Modern Physical Geography*, 3rd edition, Wiley, New York. Contains a thorough description of the US PLSS.

U.S. Department of Commerce, Bureau of the Census, 1988. *Tiger/Line File: Boone County, Missouri*, Technical Documentation, Washington, D.C.

Whitson, J. and M. Sety, 1987. "GEOLOC Geographic Location System", *Fire Management Notes*, 46:30-32.

DISCUSSION OR EXAM QUESTIONS

1. Determine the resources available to you in geocoding street addresses for your local area. What sources exist for obtaining (a) street index (DIME or TIGER) files, (b) address matching software, (c) maps with address ranges marked on streets? Estimate the time it would take to geocode 1000 addresses in this area using various combinations of these resources. What percentages of hits and misses would you anticipate? Estimate the cost per address which you would have to charge a sponsoring agency for such a project.

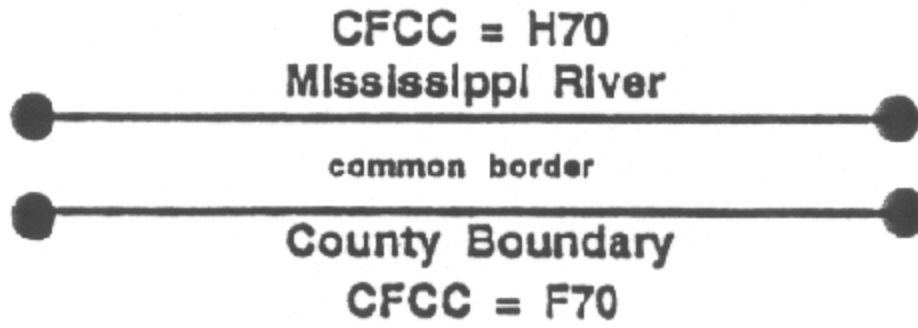
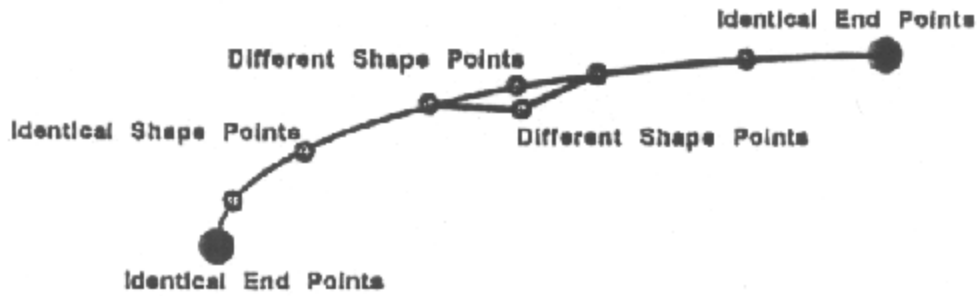
2. Discuss the usefulness of the PLSS as a georeferencing system in your local area. How complete is it? What local agencies or organizations make use of the PLSS? What is its relationship to the local system of land tenure?

3. Determine the 5 discrete georeferences described in this unit for your own residence. What problems do you have in doing this? What is the potential or actual precision of each method?

4. Discuss the ways in which the system of discrete georeferencing in the US (or your own country) might be improved. What is the appropriate level or agency of government to sponsor or undertake such an improvement? Which existing system of georeferencing should it be based on? Who are the potential users of such a system, and how might cost be shared?

Last Updated: August 30, 1997.

UNIT 29 IMAGES



West Central Columbia, Missouri





Renie's zip code map of Los Angeles

