

# **To Parcel Geocode or Not to Parcel Geocode: That is the question**

Colleen Reid, Rachel Broadwin  
and Edmund Seto

WREN 2007

**CalEPA, OEHHA, Air Toxicology and Epidemiology Branch**

# Background

- We are examining the association btw mortality and fine particles (including its species/components - e.g., OC, EC, NO<sub>3</sub>, K, Se, Fe).
- Not clear there is any threshold before experiencing adverse health effects such as mortality, particularly in vulnerable populations (elderly, comorbidity, children etc.)

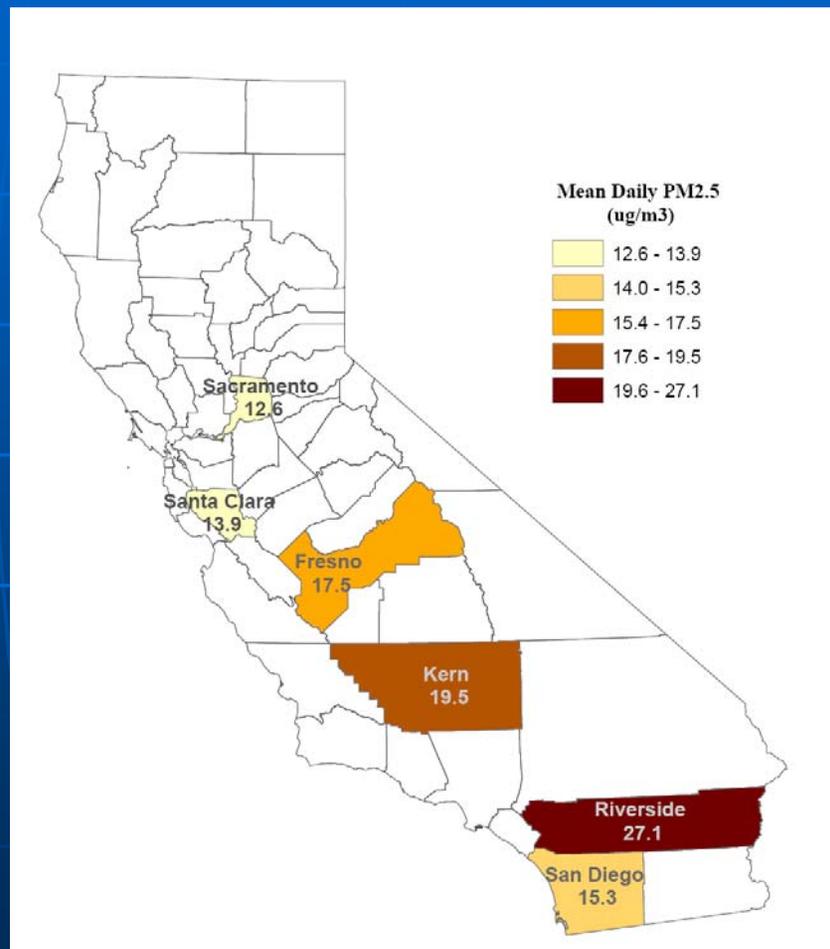
# Time-Series Analyses

- Relate counts of daily mortality to daily pollution levels, controlling for time-varying variables such as temperature, humidity, day of week, and season.
- Regression model run for each county and then results combined by a meta-analyses.

# Case-crossover Regression Analysis

Similar to matched case control - each person serves as his/her own control. The exposure level on day of death is compared to other days when the person did not die.

# CA Counties with Monitors Measuring Fine Particles



# Background (cont.)

- So far analyses have been based on the county of residence
- Now we have residential address data for a few of the years. Want to compare the association btw mort and fine particles using a finer scale.
- Concerns about accuracy of geocoding.

# Sources

- Mortality Addresses – previously standardized by CA DHS, Office of Health Info and Research (Vital Statistics) – Years 1999-2001
- Software for Streetfile Geocoding– ESRI ArcGIS v9.1
- Streetfile Database– GDT v.13 (2003) – (GDT now TeleAtlas)- more accurate than free TIGER database
- Software for Parcel Geocoding – CA Env Health Tracking Program’s Table Address Locator
- Parcel Database - TeleAtlas Address Points 2005 database
- Online aerial photos – Google Maps
- Census Density for 2000 – ESRI Data & Map ArcGIS 9 Media Kit

# What is geocoding?

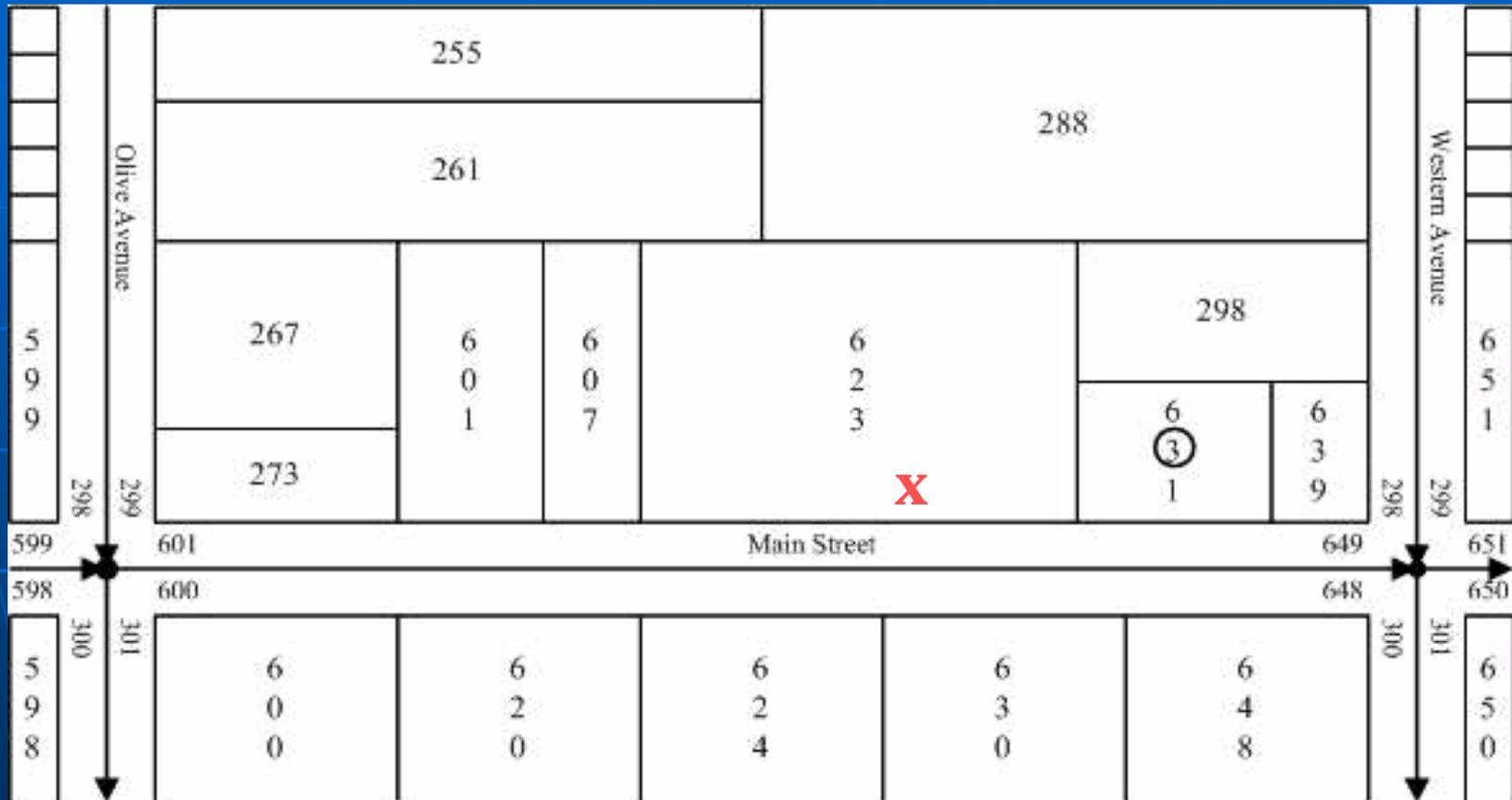
Used to assign geographical positions to residential addresses

# What are parcels?

- A plot of land
- In most cases, each residential parcel has one street address associated with it.

# Parcel versus Streetfile Geocoding

## Example – geocoding 631 Main Street



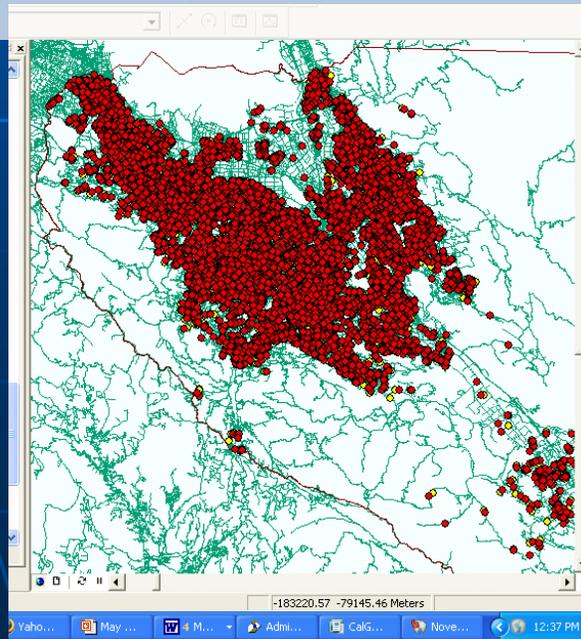
**X=location for streetfile geocode, ○ =location of parcel geocode**

# Streetfile Geocoding

Streetfile geocoding – address is parsed into house number, directional (N, S, E, W), street name, street suffix, city, & zip code and matched to street segments that correspond.

These potential matches are scored according to an algorithm in the geocoding service and then the best scored segment is assigned as a match to the address, as long as the score is at least as high as the set minimum match score.

If there's a match, the point is placed by interpolation between the beginning and ending address range assigned to the line segment.



# Parcel Geocoding

- Geocoding with parcels takes the whole address and matches it to the point of the parcel that corresponds to that address. No interpolating, so the point is more accurate; however some parcels missing from database.
- Possible for many addresses to correspond to one parcel (condominiums, apartment buildings...) – and also for one address for more than one parcel (rural areas with farms...)
- In order to use this service, you have to send an email to Craig Wolff at EHIB and request a password and some information on what you will use it for.

# Flow Chart of Geocoding

**Compared one county during one year**

7950 addresses

EHIB Address Points  
“Parcel Geocoding”

ArcGIS streetfile  
geocoding

**6579 matched (82.8%)**  
(min match score=60  
we could not change this  
score)

**7749 matched (97.5%)**  
(min match score = 77 because  
below this lead to matches when  
street number missing)

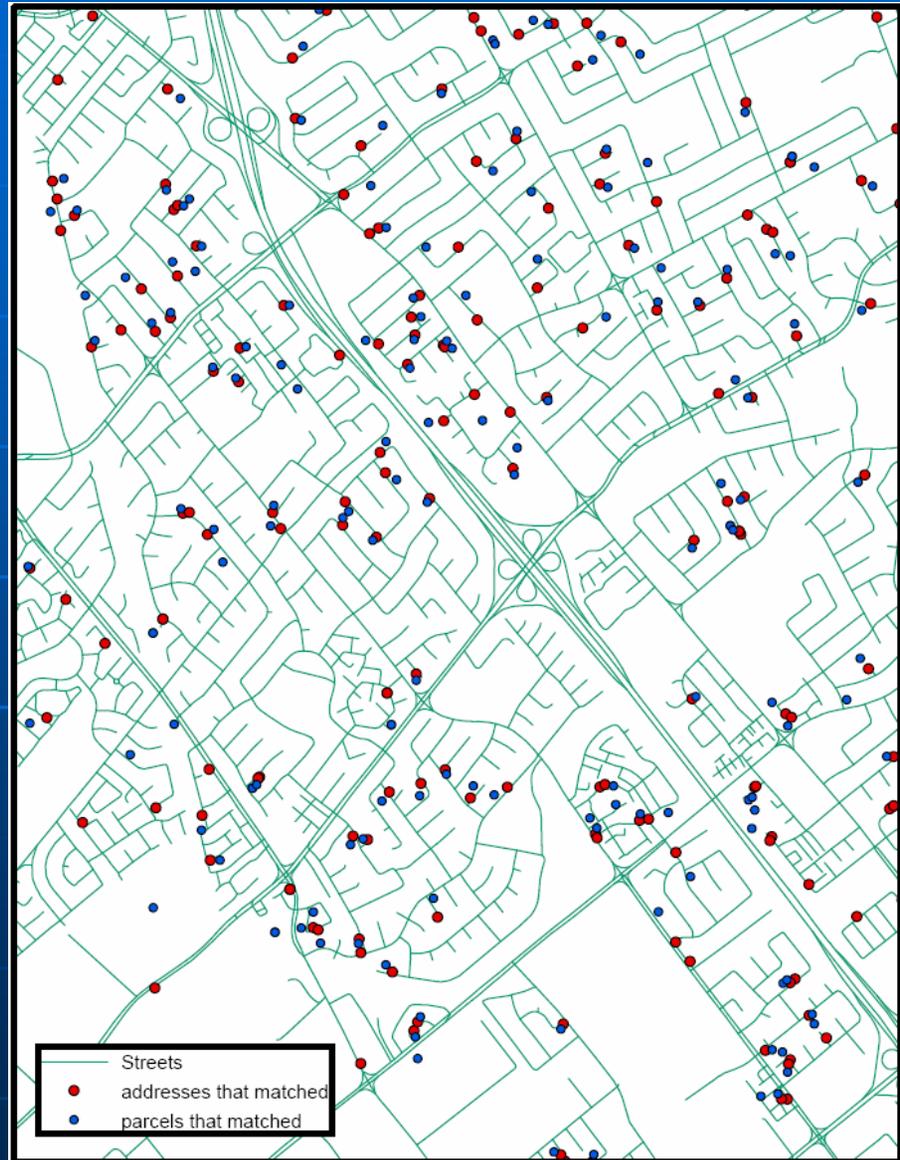
**All of the addresses that matched with a score between 60 and 77 in parcel geocoding matched with a perfect score of 100 in the streetfile geocoding.**

# Comparing Geocoding Methods

- Streetfile geocoding
  - Easier Access
  - Higher match rate
- Parcel Geocoding
  - More accurate
  - Lower match rate

# Comparing Geocoding Methods (cont.)

- But which geocoding method is better?



# COMPARISON OF MATCH RATES

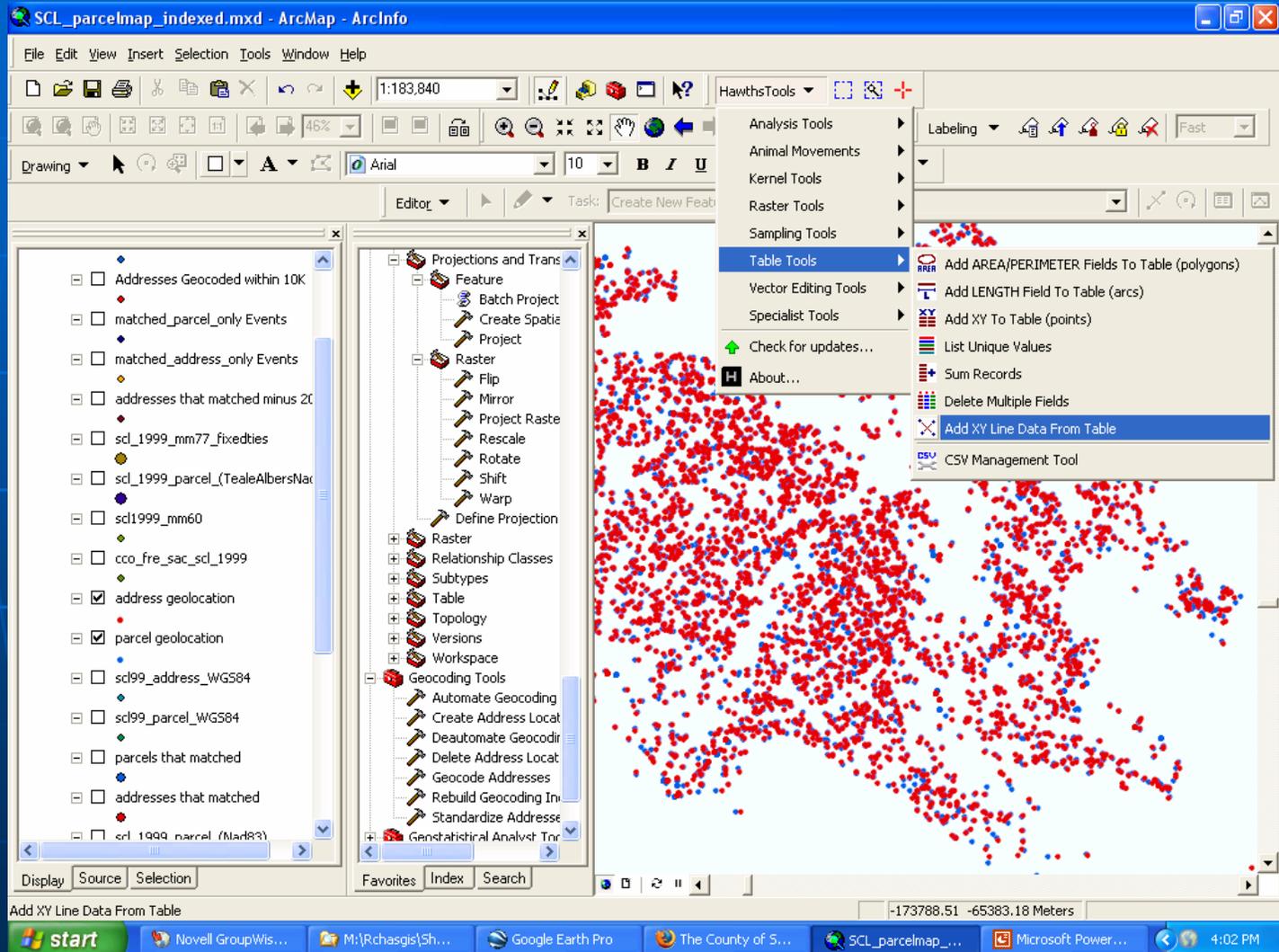
Out of total of 7950 addresses attempted to geocode in both methods:

Matched in both street and parcel geocoding:	6519 (82%)
Matched in streetfile but not in parcel geocoding:	1230 (16%)
Matched in parcel but not in street geocoding:	60 (1%)
No match with either geocoding method:	141 (2%)

	Streetfile	Parcel
Matched	7749 (97%)	6579 (83%)
No Match	201 (2%)	1371 (17%)

No addresses matched to the same location with both streetfile and parcel geocoding.

# Drawing lines to connect parcel and streetfile geocodes - Hawth's Tools



<http://www.spatial ecology.com/htools>

# DISTANCE BETWEEN STREET AND PARCEL GEOLOCATIONS

**N= 6519 addresses matched in both methods, distance given in meters**

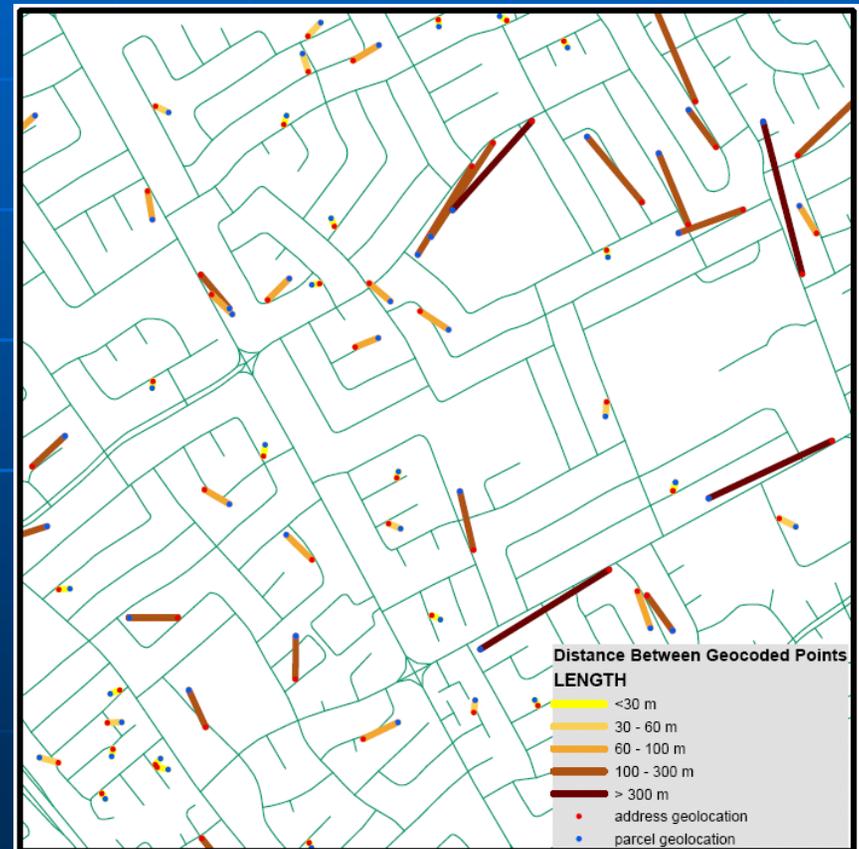
**Minimum: 2.8**

**Maximum: 3400**

**Mean: 86**

**Median: 44**

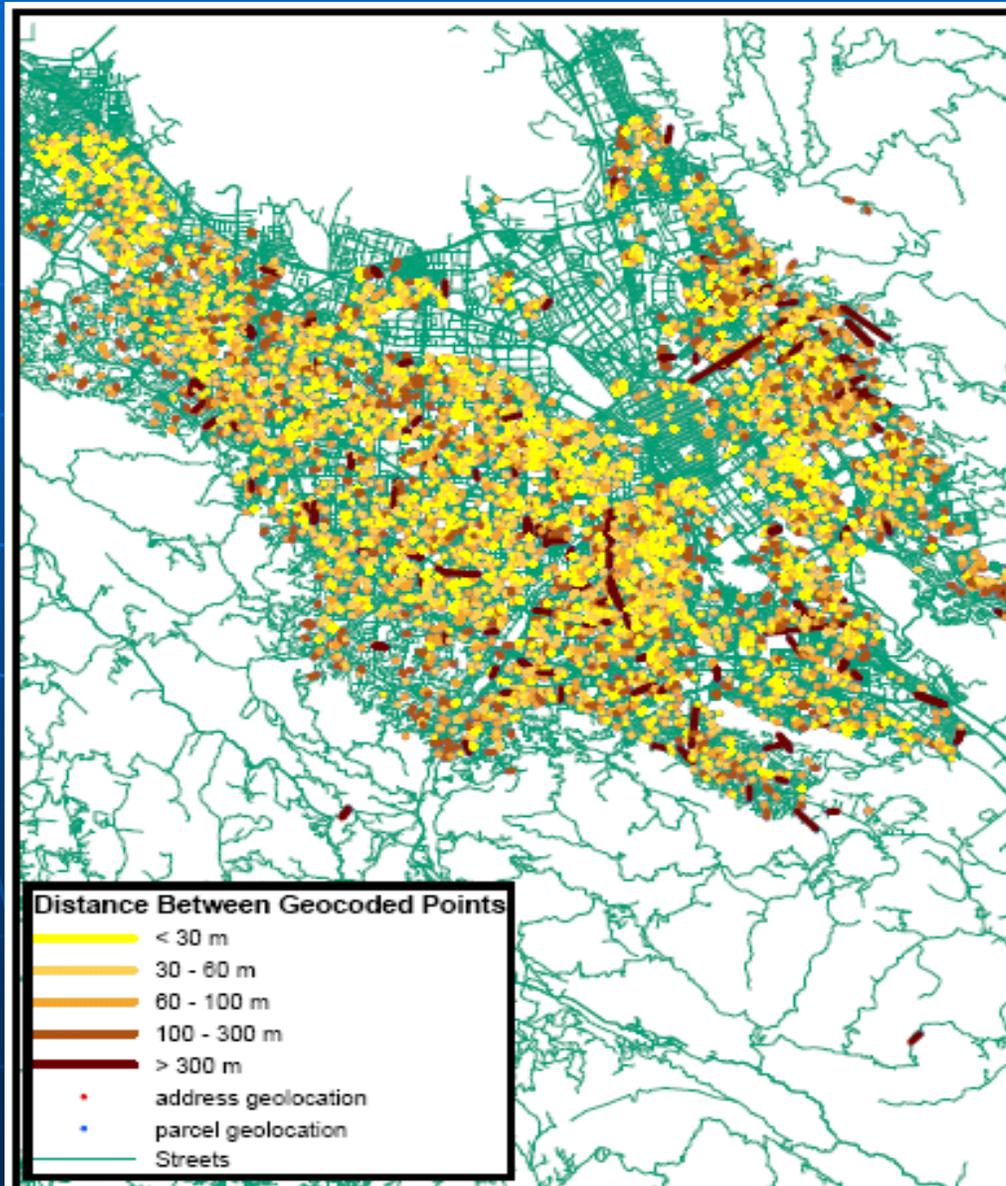
**Stand Dev: 160**



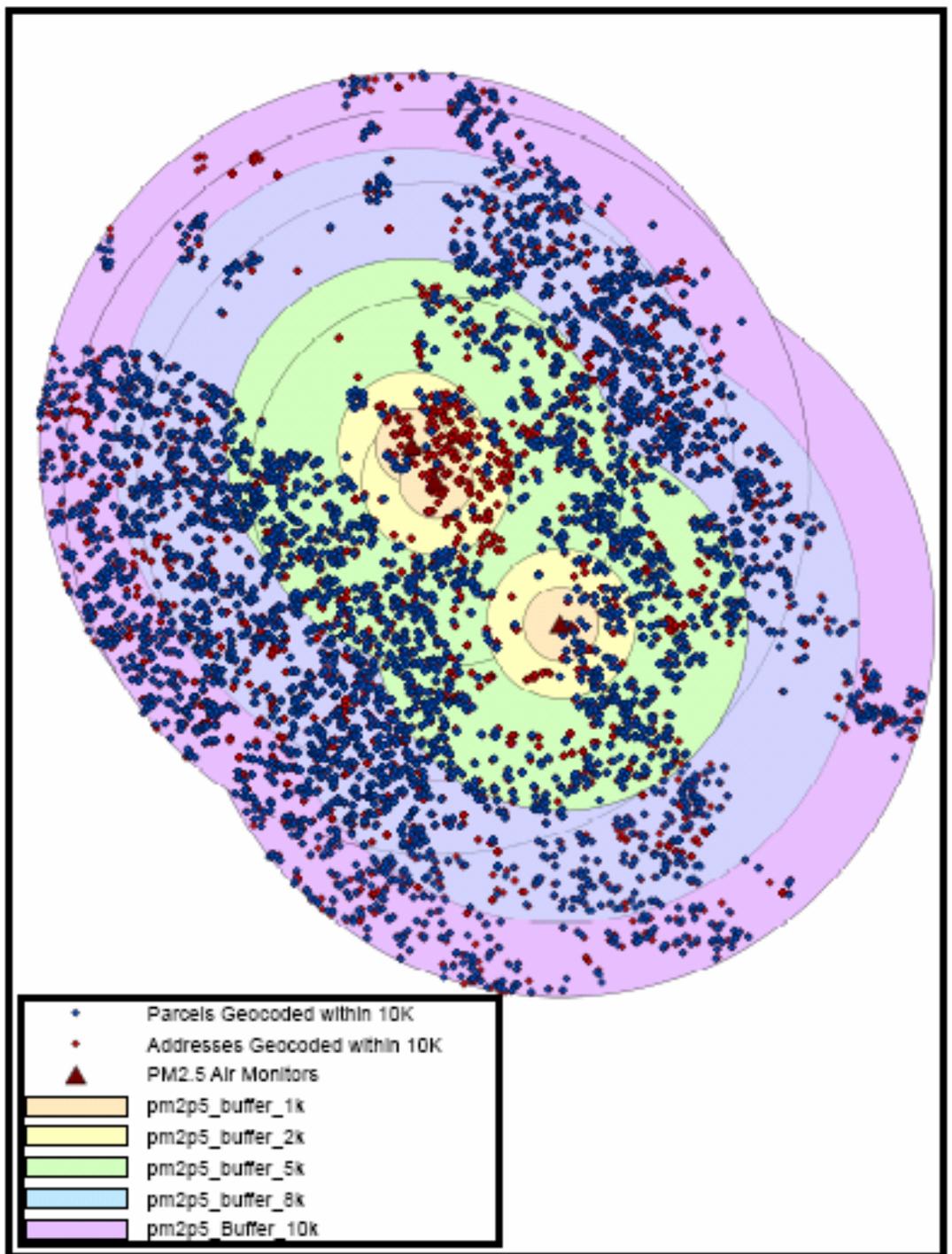
## Number of addresses by distance between their geolocated points in streetfile versus parcel geocoding methods

< 30 m (within the same parcel)	<b>2228 (34%)</b>
30 - 60 m (within neighboring parcels)	<b>1790 (27%)</b>
60 - 100 m (on the same block)	<b>1188 (18%)</b>
100 - 300 m (different blocks)	<b>997 (15%)</b>
300 - 3400 m (different neighborhoods)	<b>316 (5%)</b>

# Distances between Parcel and Streetfile Geocoding Methods



But how does  
this affect the  
exposure  
classification?

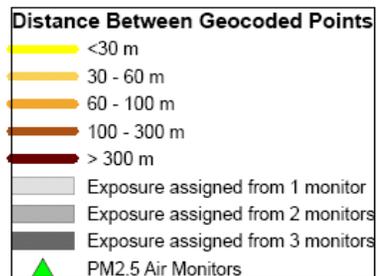
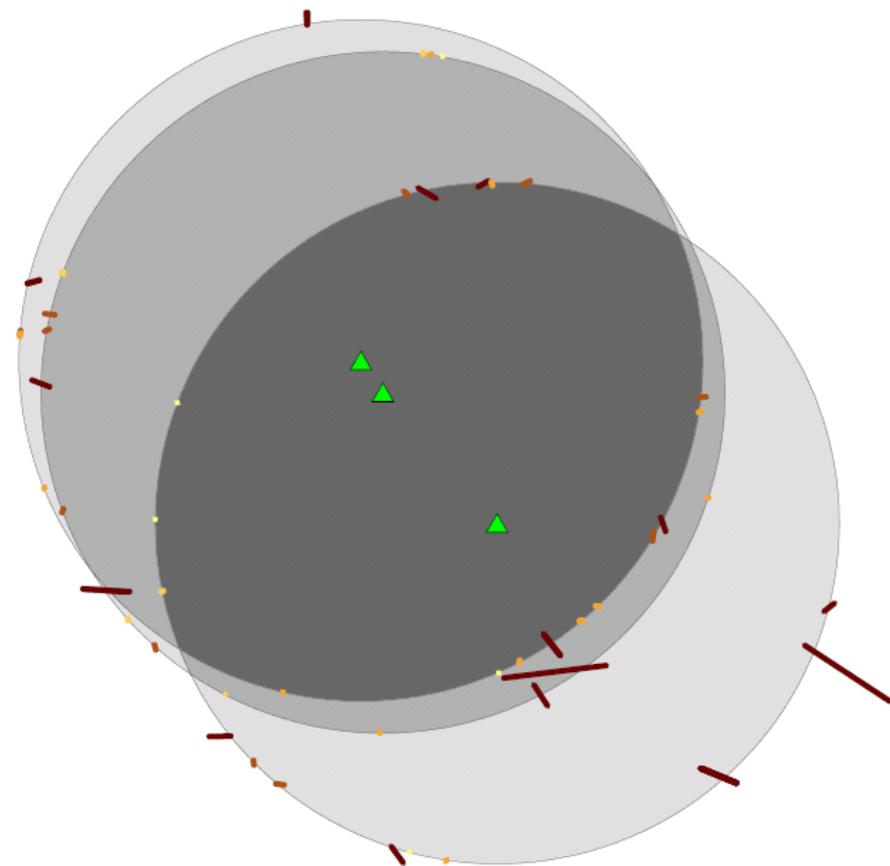


# How Streetfile vs Parcel Geocoding Methods Affect Exposure Assignment

Out of the 6519 addresses that geocoded in both methods, 3794 (58.2%) had either a parcel or streetfile geocode that fell within 10K of at least 1 of the 3 air pollution monitors for this county. Within that group, the number of addresses where the parcel and streetfile geocodes differed by whether they fell within the same buffer:

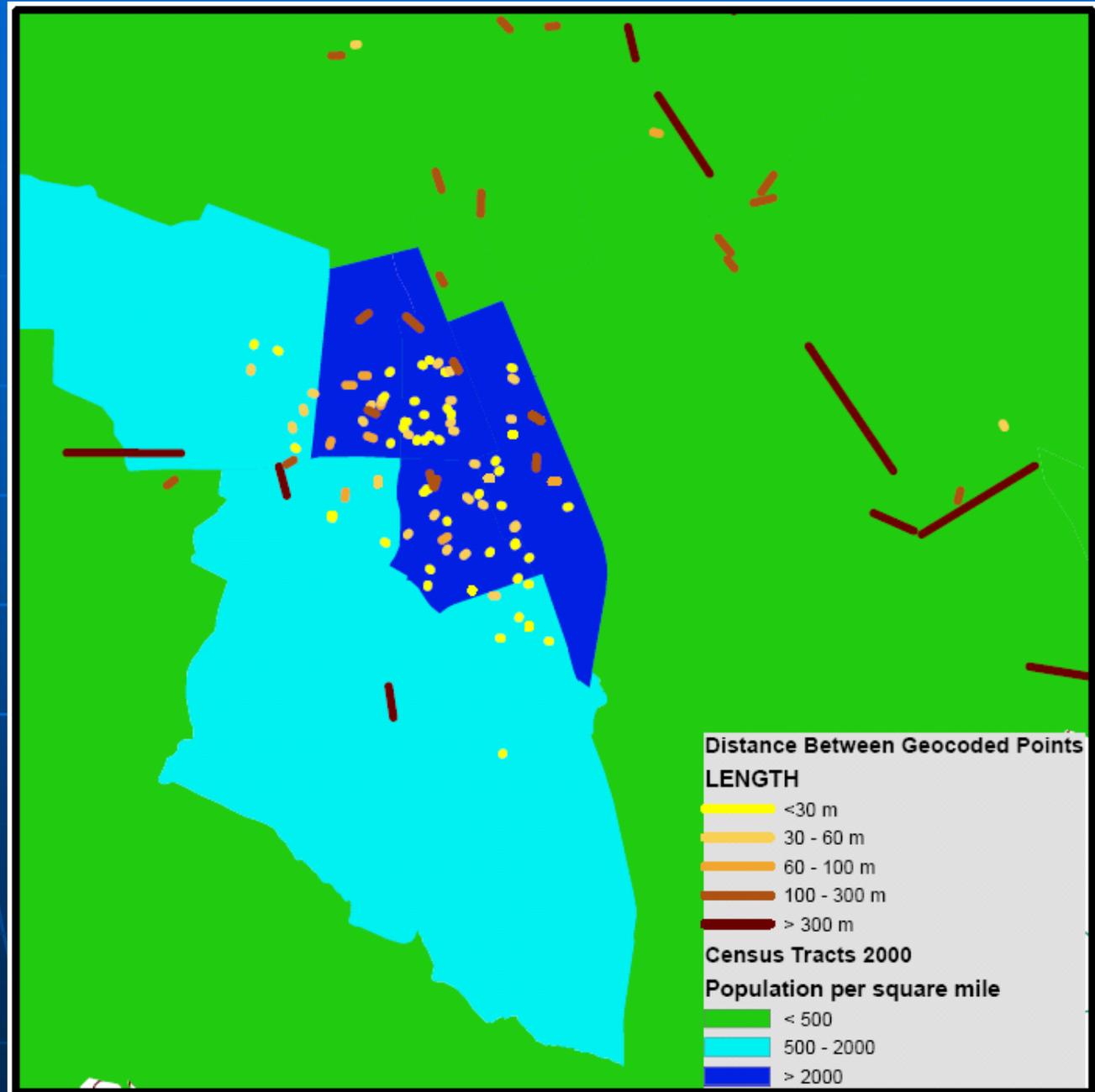
- In 1K buffers: 1 address out of 34 (2.9%)
- In 2K buffers: 3 addresses out of 151 (2.0%)
- In 5K buffers: 58 addresses out of 1360 (4.3%)
- In 8K buffers: 86 addresses out of 2969 (2.9%) (see next slide)
- In 10K buffers: 63 addresses out of 3794 (1.6%)

Example for  
the 8k buffer  
where  
**ADDRESSES  
CHANGE**  
**EXPOSURE  
CATEGORY**  
**N = 86 addresses**

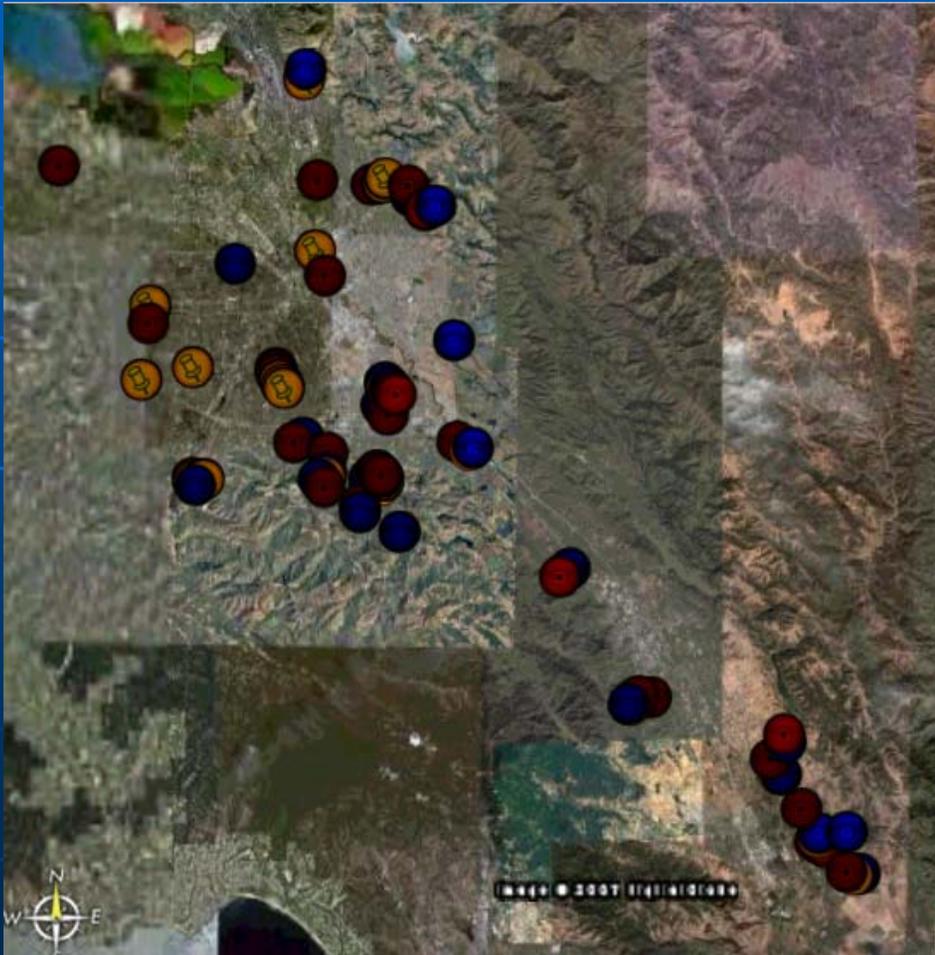


# DISTANCES BETWEEN GEOCODED POINTS

BY DENSITY  
OF CENSUS  
TRACT



# How accurate is each geocoding method?



**Compared with Google Earth Aerial Photography and Google Earth Address Matching (database updated continually, often a cross between streetfile interpolation & parcel geocodes)**

# Parcel Map from County Assessor – Picture Only, Not georeferenced

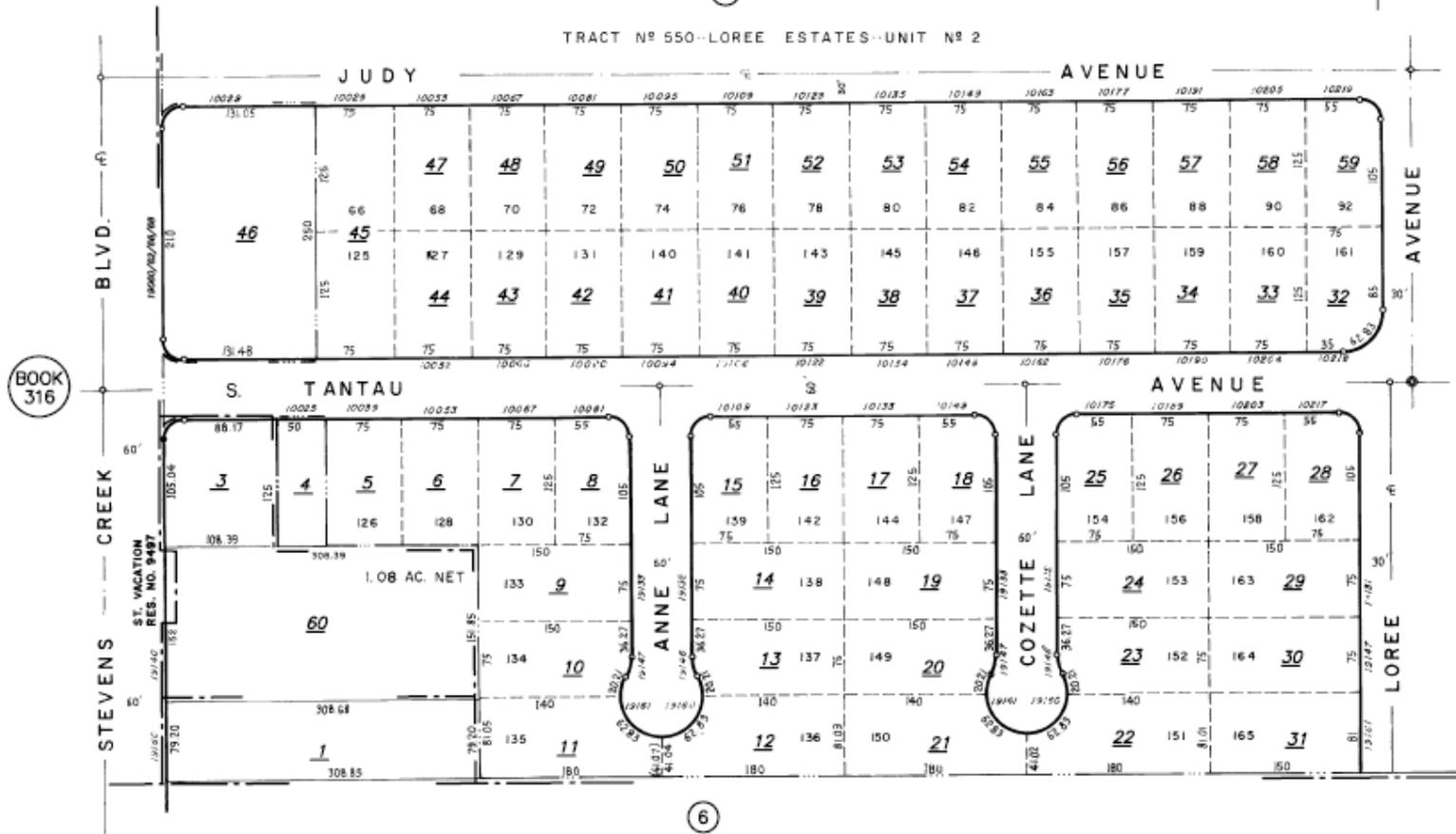
OFFICE OF COUNTY ASSESSOR — SANTA CLARA COUNTY, CALIFORNIA

BOOK <b>375</b>	PAGE <b>7</b>
--------------------	------------------



(11)

TRACT N° 550--LOREE ESTATES--UNIT N° 2

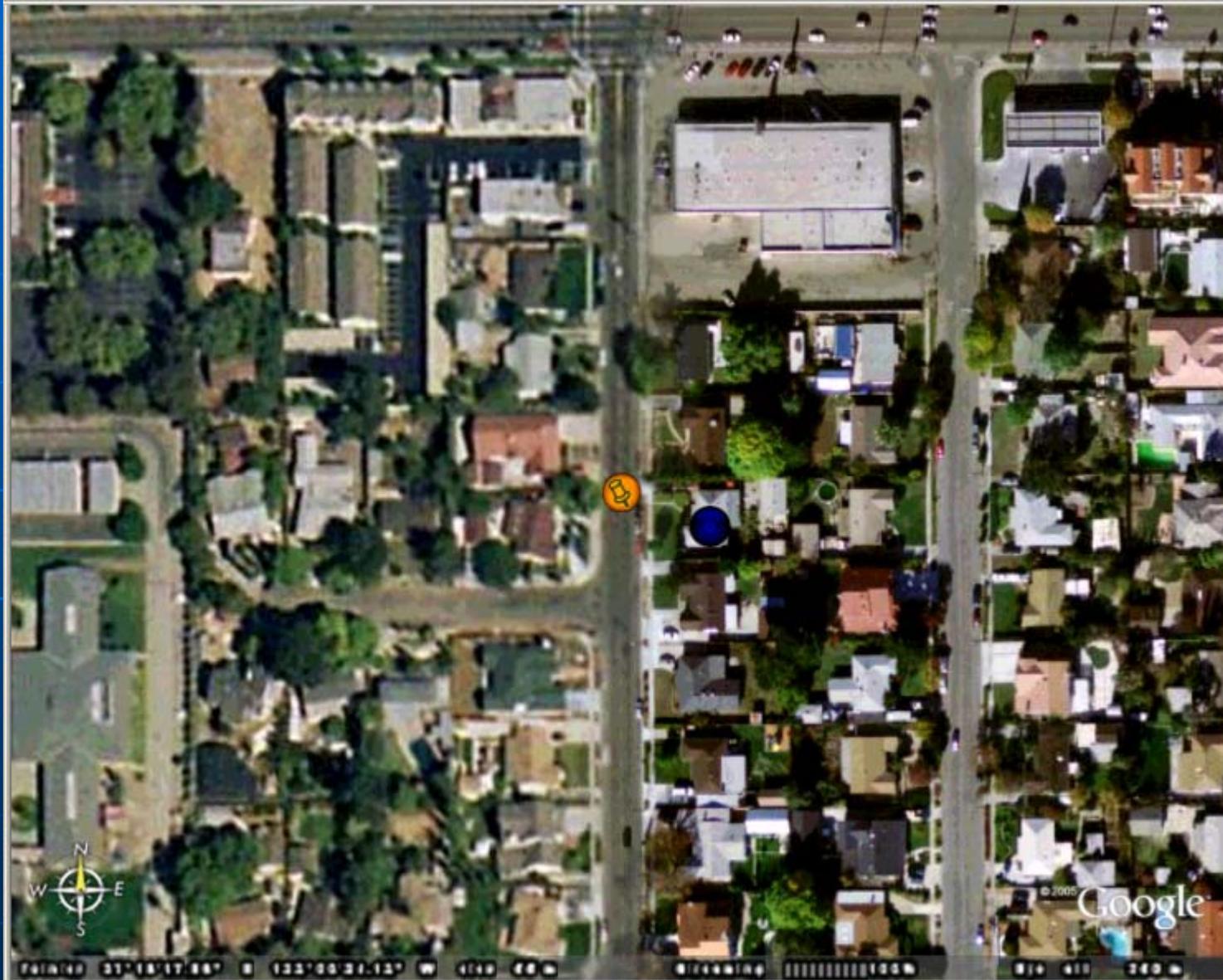


BOOK  
316

(8)

LAWRENCE E. STONE — ASSESSOR  
Cadastral map for assessment purposes only  
Compiled under R. & T. Code, Sec. 327.  
Effective Roll Year 2006–2007

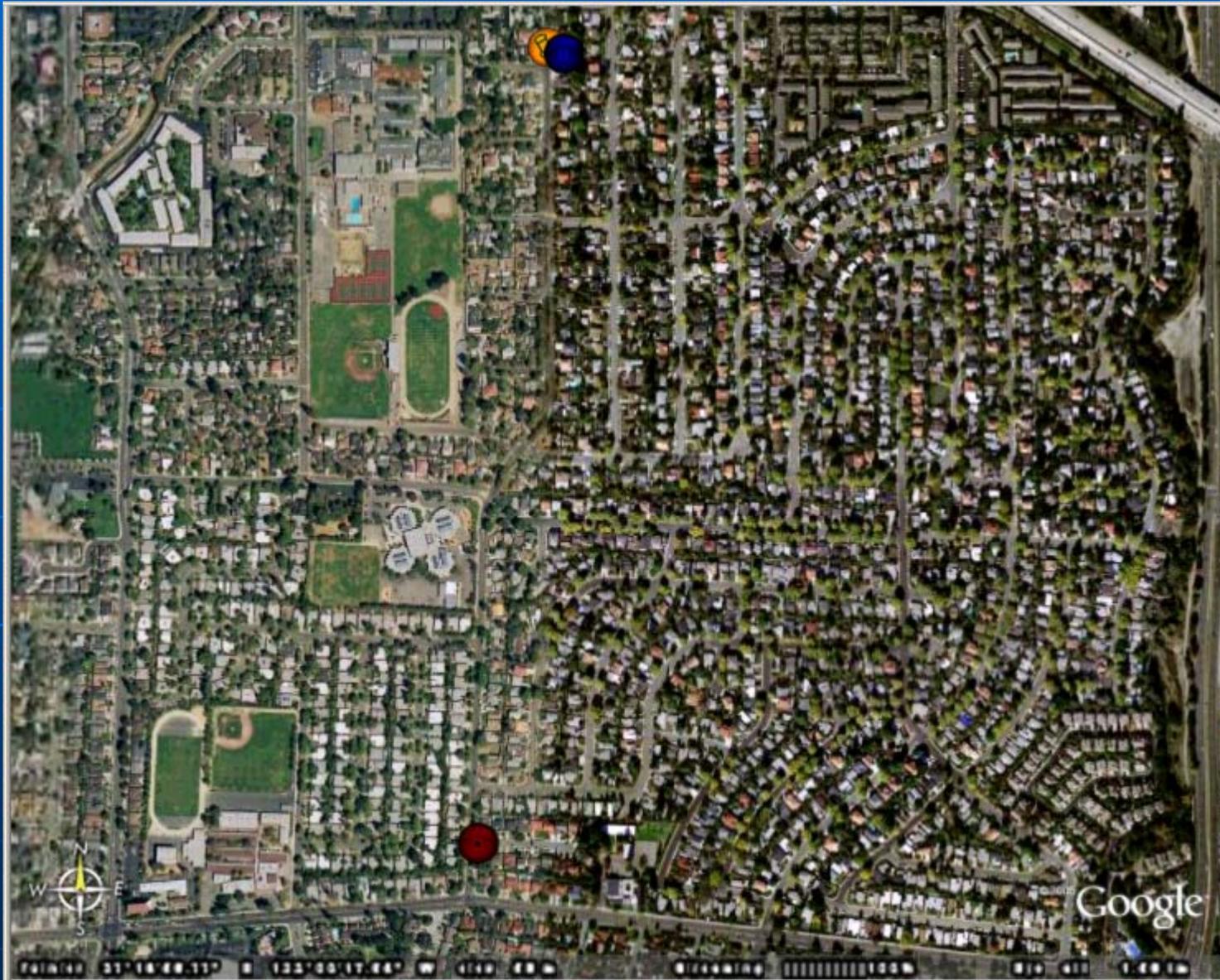
# High Density



Google  
geocode:  
orange dot

Parcel  
geocode:  
blue dot

# Hi Density



Google  
geocode:  
orange dot

Parcel  
geocode:  
blue dot

Streetfile  
geocode:  
red dot

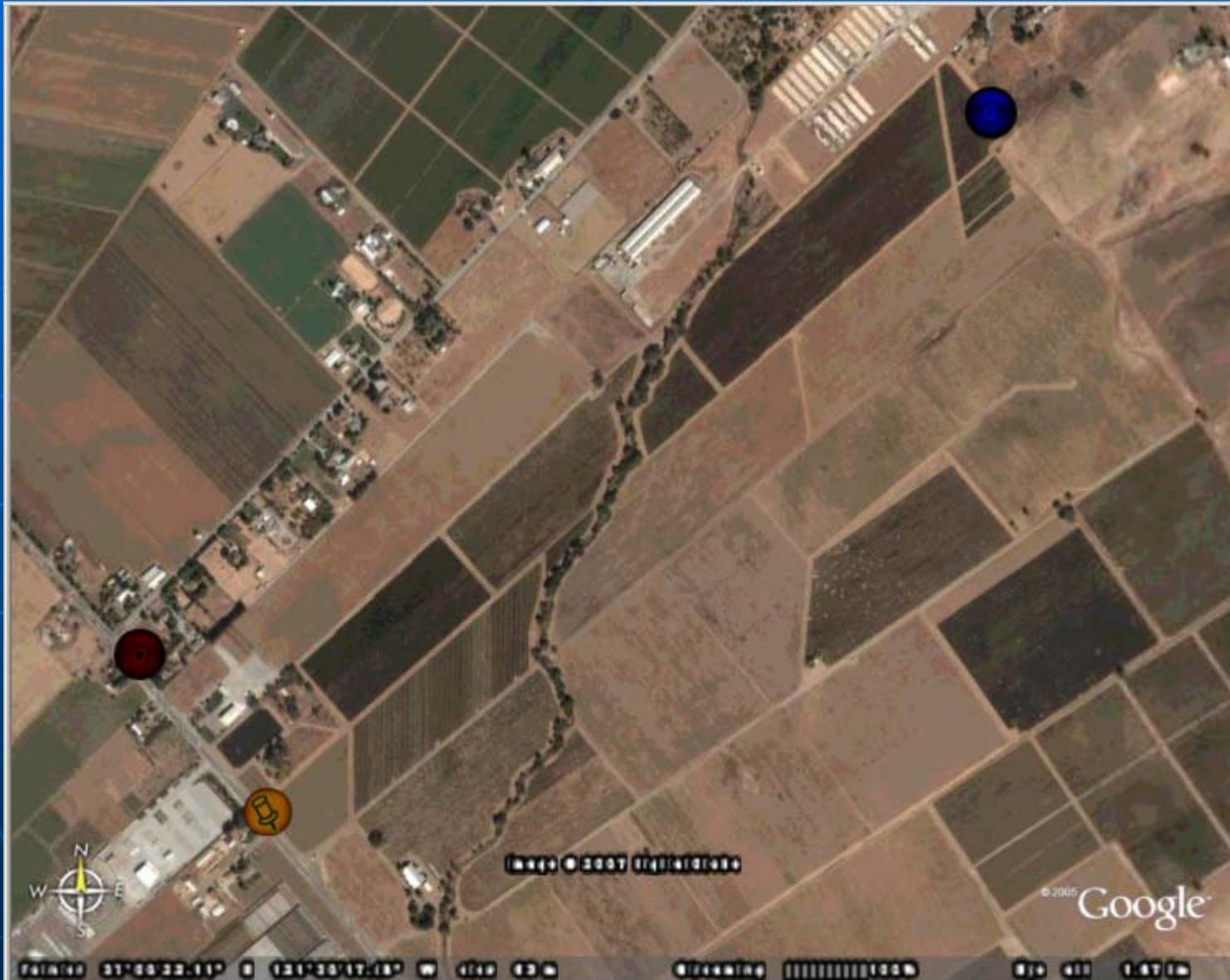
Distance from red dot to blue dot = 1141 meters

# Middle Density Areas



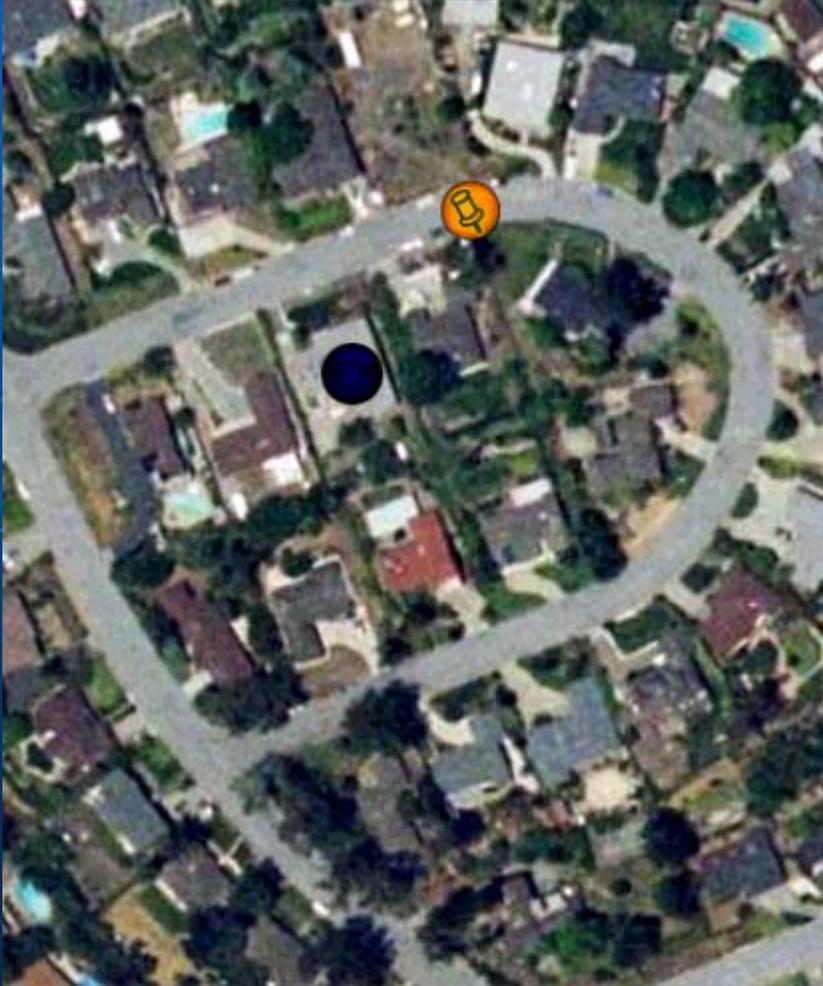
Distance from red dot to blue dot = 660 meters.

# Low Density Areas



Distance from red dot to blue dot = 1708 meters

# Matched only with EHIB parcel

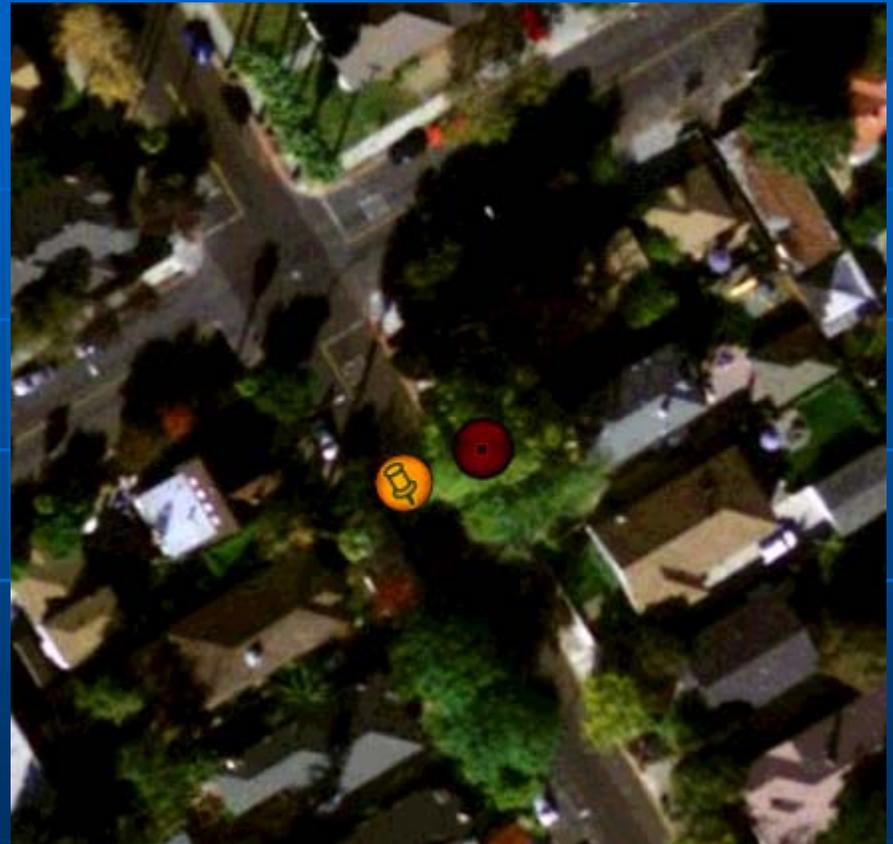


- One had an assessor's map and the EHIB parcel was correct
- Four had apparent typos in the address that the EHIB geocoder parsed into a corrected address

Distance between blue and orange dots = 52 meters

# Matched only in Streetfile

- Three had no assessor's map, so unclear
- Two were in a high density area where the address looked accurate



Distance between red and orange dots = 9.6 meters

# Study that shows bias by geocoding method

Zandbergen 2007

Streetfile geocoding found to consistently over-estimate the number of potentially exposed children at distances up to 250 meters (Orange Co, FL)

# Bias decreases as buffer increases

Adapted from Table 2 of Zandbergen 2007

**Bias and error in determining children at-risk based on proximity to major roads in Orange County, Florida.**

Buffer Radius	Parcels yes/ Street yes	Parcels yes/ Street no	Parcel no/ Street yes	Parcel no/ Street no	Sensitivity (%) <sup>1</sup>	Specificity (%) <sup>2</sup>
50 meters	<b>151</b>	240	1262	<b>103212</b>	39	99
150 meters	<b>3899</b>	818	1377	<b>98771</b>	83	99
250 meters	<b>9704</b>	1025	1241	<b>92895</b>	90	99
1000 meters	<b>53838</b>	662	776	<b>49589</b>	99	98

<sup>1</sup> percentage of parcel geocoded children residing within the buffer zone that were correctly classified using street geocoding.

<sup>2</sup> percentage of parcel geocoded children residing outside the buffer zone that were correctly classified using street geocoding.

Bold indicates where streetfile and parcel in accord

# Cartographic Confounding

When the factor of interest is geographic, a factor related to the disease that is not distributed randomly across the study area can confound the appearance of the map of the disease

Oliver et al 2005

# Conclusions

- Parcel geocoding more accurate than streetfile geocoding, but lower match rate
- In high density areas average distance between the geocoding methods is lower

# Conclusions (cont.)

- In high density areas, only a small percentage of addresses change exposure category due to a change in geocoding method
- Spatial data accuracy needs to be greater than minimum distance used in analyses.



Colleen Reid

Recent UCB Grad  
MPH program  
(Env Health Sciences)



Edmund Seto

UCB Lecturer &  
Assistant Researcher  
(Env Health Sciences)



**Contact – [rbroadwi@oehha.ca.gov](mailto:rbroadwi@oehha.ca.gov)**