

Ground Truth Tracking of Land Use in SCAG's Parcel Database

Mami, Ramy and Matthew P.H. Taylor

September 2013

University of California, Riverside

Introduction

Parcel-level property records are valuable sources of information, both to the forecasting and tax assessment efforts of the local governments that typically construct and maintain such databases and to research conducted by a wide range of academic interests. Unfortunately, such parcel-level databases are ubiquitously problematic. First, the state of individual properties is constantly in flux, making accuracy and completeness moving targets. Second, the responsibility of collecting land use data is frequently the responsibility of county municipalities. Lacking the resources necessary to achieve the accuracy and sophistication of, say, the U.S. Census Bureau, efforts to maintain a comprehensive, accurate, and up-to-date database are inherently unrealistic.¹ Furthermore, because this process is decentralized, there is often a lack of standardization in the data collected, its level of accuracy and completeness, and the collection process across, if not within, counties, increasing the difficulty of using the data as a basis for rigorous analysis.

This study undertakes a preliminary investigation of the validity of land use data in the parcel-level property records of Southern California. The database in question is maintained by the Southern California Association of Governments (SCAG), an association that incorporates the Imperial, Los Angeles, Orange, Riverside, San Bernardino, and Ventura counties. The Association assembles assessment data from each of the six County Assessor's offices, data on sales transactions from their respective Recorder's Divisions of the County Clerk, and unstandardized land use, zoning, and planned land use data from individual municipalities and planning boards. From its initial construction through its intermittent maintenance the SCAG database has been poorly documented, resulting in a lack of clarity regarding the underlying processes. This creates further uncertainty regarding the quality and consistency of this particular database.

These property records are an important source of inputs for the application of the RELU-TRAN model to the Los Angeles Metropolitan Area, the primary focus of the LA-Plan project. Prior use of the SCAG database in the LA-Plan project has encountered numerous difficulties and uncovered extensive problems in the database. For one, many cells contain zeroes and blanks, and their patterns differ across counties (Gu and Arnott, 2011). Further, the computed mean floor-area ratios do not square with informed experience. Consequently, Goodchild et al (2011) utilize "ground truth tracking" to assess the validity of zeroes and blanks in the database's total floor space variable. This work found that, across the study's samples, 34% of supposedly developed properties were actually vacant. This casts some doubt on the accuracy of parcel land use information recorded in the SCAG database.

Following up on these initial discrepancies, this assessment utilizes "ground truth tracking" and a small random sample of properties in Palm Springs, California to systematically evaluate the land use data in the 2007 SCAG database. As on-site visitation is a time intensive process, the samples used in the analysis are relatively small and statistically representative of neither the Palm Springs area nor the

¹In addition, because these municipalities use such databases for forecasting and tax assessment, they frequently lack the incentives to maintain a level of accuracy that is typically required of more academic research efforts.

encompassing SCAG region. However, the results provide initial feedback regarding the frequency, types, and sources of error in the land use data, providing insight and foundation for more systematic study and adjustment of the SCAG database. The remainder of the report lays out the assessment methodology in detail, followed by a presentation of the study results and, lastly, some concluding discussion and recommendations concerning the use of SCAG's land use data.

Methodology

The 2008 SCAG land use data was assessed in three separate stages, each of which focused on a broad category of property land use – stage 1 looked at parcels categorized as taxable and developed, stage 2 at parcels indicated to be vacant (undeveloped), and stage 3 studied tax-exempt developed parcels of land (i.e. properties owned by government, educational, medical, and other organizations that are tax exempt and developed). Determining a parcel's land use required some simplification of the SCAG database, which utilizes over 100 unique land use codes. We aggregate land uses to just 13 classes (see Appendix A for details), further generalized into the three broad categories used in this study (taxable developed, undeveloped/vacant, and tax-exempt developed).

In the first two stages ground truth tracking was implemented on a small random sample of 100 parcels of the respective land use types. The third stage looked at a smaller sample of 30 sites, 5 each of the 6 classes of tax-exempt property (government, medical, religious, educational, special use, and parks and recreation facilities). The geographic area of study was the model zone (92) associated with the greater Palm Springs area, including the municipalities of Cathedral City, Desert Hot Springs, and White Water. For each parcel assessed, the goal was to determine (1) if the 2008 SCAG database accurately reflects the property's actual usage in 2013, and (2) the nature of any existing discrepancies using ground truth tracking.

The ground truth tracking process uses an ArcGIS dataset, Google Maps, Google Street View, and real estate websites to conduct a preliminary evaluation of the parcels in question. The sequence of the process is as follows: (1) enter the latitude and longitude coordinates in ArcGIS, obtaining the parcel boundary, an initial Google Earth picture (via a GIS overlay), and potentially an address; (2) attempt verify that information using Google Maps; (3) provide further verification using Google Street Views where possible; and lastly, (4) use the real estate websites www.trulia.com, www.realtor.com, and www.zillow.com to verify the address, building type, and built date of any improvements on the property. In many cases, this preliminary investigation is sufficient to either affirmatively confirm the land use that is recorded in the SCAG database or to provide sufficient evidence to contradict and correct the land use on record. As the database in question is from 2008, one would expect that at least some of the parcels being analyzed would have undergone significant changes in development status, if not land use.

In person site visits were warranted for those parcels whose land use remained uncertain after the initial internet evaluation. These site visits were conducted throughout August and September of 2013, providing detailed information on the use and status of the property in question ascertained through

visual inspection and, wherever possible, interviews with property owners, tenants, and neighbors.² The heuristics for determining if a site visit is necessary vary between the three stages:

- **Stage 1** – All purportedly developed parcels where preliminary internet investigation provided evidence contradicting the recorded SCAG land use were visited in person.
- **Stage 2** – Parcels labeled as vacant were visited in person if the preliminary internet investigation indicated that the property had some completed level of development, and if that development could not be thoroughly verified by finding a street address using the real estate websites.
- **Stage 3** – All 30 tax-exempt properties received site visits.

On rare occasion a site visit was not informative because the property in question is privately owned and not street-side, making the assessment methodology unfeasible. However, successful site visits resulted in a file clearly documenting the evidence used to support the investigator’s conclusions regarding the parcels actual 2013 land use. A typical file includes the image of the parcel boundary recorded from the ArcGIS database, the view from Google Maps and/or Google Street View, and a photograph taken by the on-site investigator.³

Figure 1: Example of ArcGIS Parcel Boundary Image



Figure 2: Example of Google Maps’ Street View Image

² The questions asked were aimed at determining the number of families living in the building, the vintage of the building, etc. For privacy reasons, these conversations were not recorded and no link to the transcripts is included in this report.

³ Note that all images obtained from Google Maps or Google Street View had been taken in 2013.



Figure 3: Example of Site Visit Photograph



Assessment Results

Ground truth tracking indicated that, for the samples in question, 85% of all developed parcels had the actual land use accurately reported in the 2008 SCAG database. Similarly, of the vacant parcels and the tax-exempt developed parcels sampled 58% were still vacant and 77% were tax-exempt properties, respectively. The remainder of this section discusses the results in further detail.⁴

Stage 1 – Developed Parcels

Ground-truth tracking indicates that 85% of the taxable developed properties sampled had accurate 2008 land uses – 8% appeared to be vacant properties and the remaining 7% were developed but were determined to have a contradicting land use. Table 1 below shows the accuracy of this sample’s recorded 2008 land use, by land use type. Note that no industrial properties and very few commercial properties were included in the study, so the above numbers are largely a reflection of residential properties. For more detail on the nature of the land use discrepancies see Appendix B.

⁴ Note that, with the exception of 10 discrepancies which were obtained from the Google Maps images, the discrepancies regarding parcel land use that led to site visits were all obtained when the investigator switched to Google Street View and found conflicting images.

A certain amount of ‘switching’ land use is expected given the 5-year time lag between the determination of the 2008 land use code and the time of this study. This is especially true for residential properties that might move between mobile/manufactured homes, single family homes, and small multi-family units (i.e. duplexes, etc.), or for movement between large multi-family structures and other commercial uses. Thus, perhaps a more telling figure is that 92% of the taxable developed sample was indeed developed.

Table 1: Accuracy of Recorded Land Use for Developed Parcels

Land Use	Sample N	N Accurate Land Use	% Accurate Land Use
Single Family	81	75	93%
Multi Family	6	3	50%
Mixed Residential	5	1	20%
Office	0	--	--
Retail	4	3	75%
Other Commercial	0	--	--
Warehousing	0	--	--
Industrial	0	--	--
Mixed Commercial and Industrial	0	--	--
Transportation/Communication/Utilities	3	2	67%
Public	1	1	100%
Other	0	--	--
Total	100	85	85%

Stage 2 – Vacant Parcels

The MRPI land-use classifications denote four distinct SCAG land use types as vacant land – agricultural, open space and recreation, vacant, and under construction. As seen in Table 2, the majority of the sampled vacant properties were listed as under construction. Just 58% of the properties in the sample were verified as being vacant, and the majority of the attrition came from properties listed as ‘under construction’ - the observation that almost half of these properties were now completed prompted further investigation.

Table 2: Accuracy of Recorded Land Use for Developed Parcels

Vacancy Type	Sample N	N Accurate Land Use	% Accurate Land Use
Under Construction	85	45	53%
Recreation	11	10	91%
Agricultural	4	3	75%
Vacant	0	--	--
Total	100	58	58%

While not surprising that some properties under construction in 2008 were completed five years later (that is, after all, the hopeful destiny of properties under construction), the concern is that classifying ‘under construction’ properties as being vacant may be inaccurate and inappropriate. If the 2008 land use classification ‘under construction’ is regularly treating built homes as vacant then reliance on this classification may bias further analysis, a potential problem for the LA-Plan project.

There were two sources of information that point towards some systematic miss-classification of ‘under construction’ properties in the 2008 land use code. First, for those properties that were in fact completed buildings the SCAG record of planned land use was relatively accurate. Table 3 below displays the accuracy of the 2007 SCAG planned land use code for those parcels with a 2008 land use coded as ‘under construction’ that ground-truth tracking revealed as having completed construction, by observed land use. The planned land use was accurate for 78% of the properties in question, indicating that, at least for a sub-sample of the ‘under construction’ properties, there is an alternative and more accurate record of actual land use than the 2008 land use field.

Table 3: Accuracy of Recorded Planned Land Use for Properties No Longer ‘Under Construction’

Actual Land Use	N Parcels	N with accurate SCAG planned use	% with accurate SCAG planned use
Indeterminate	1	--	--
Road	1	0	0%
Single Family	34	28	78%
Church	1	1	100%
Park and Ride	2	1	50%
Mult-family	1	1	100%
Total	40	31	78%

Second, if we focus on the 34 single family homes that, according to trulia.com and realtor.com, were verified as being built, 30 of these have records indicating the homes were completed no later than 2005. Furthermore, 28 of these 30 had accurate planned land use codes. Taken together, these two facts suggest that the ‘under construction’ classification includes homes that had completed construction well before 2008. This may be an artifact of the SCAG database’s updating processes, however it is difficult to determine the source of the discrepancy with any degree of confidence.

This issue is less problematic if the entire property record has not yet been updated, meaning the parcel data is consistent and reflects a property that is still under construction (depending upon the prevalence of this misclassification, however, such an error may still have importance). However, if there was only partial updating of the parcel data where some variables were changed to reflect a built home but those used to determine 2008 land use were not changed, then using this land use classification to conduct further analysis may lead to significant bias.

For example, if we focus on the total 2007 value of single family properties we see that this appears to be the case. Table 4 below displays cursory descriptive statistics of the total 2007 value of three subsets of single family properties:

1. single family properties with a land use of single family that were in fact single family residences (from the stage 1 sample);
2. single family properties with a 2008 land use of under construction that were in fact completed single family homes;
3. single family residential properties with a 2008 land use of under construction that, while still under construction, had planned land use of single family residential.

The first group, drawn from Stage 1, is intended as a reference. We would expect that, if the data for the misclassified property records reflected ‘under construction’ properties then the distribution of land values would be similar to that of the properly classified ‘under construction’ properties. This, however, is clearly not the case. The mean and median values of the misclassified properties that were in fact completed single family homes resembles those of properly classified single family homes much more so than single family homes that are actually under construction.

Table 4: Descriptive Statistics for Three Samples of Single Family Homes

	Stage 1 Single Family*	Stage 2 Completed Single Family**	Stage 2 Single Family Under Construction***
N	69	27	28
Minimum	\$17,748	\$7,693	\$3,445
Median	\$256,337	\$339,553	\$70,737
Mean	\$287,465	\$332,772	\$63,248
Maximum	\$768,510	\$536,193	\$230,625

* Six of the properties in the original sample were removed because they had total values of zero; ** one of the properties in the original sample was removed because it had a value of over \$1.2 million; *** two of the properties in the original sample were removed because they had values of \$1 and \$10.

While this investigation is by no means representative of the Palm Springs properties, never mind the SCAG database overall, the results from ground-truth tracking cast some concern about the classification of properties ‘under construction’ as being vacant. This is most significant if these parcels’ data are to be used for further analysis or imputation. This finding warrants further analysis and documentation of the development and accuracy of the lu_08 classification.

Stage 3 – Tax-exempt Parcels

The final stage of this research analyzed 30 developed parcels having tax-exempt 2008 land uses. As seen in Table 5 below, approximately 77% of the sample were confirmed as appropriately having tax-exempt status. With the exception of one empty lot next to a church, all parcels in question meet the definition of being developed. As with the samples of Stage 1 and Stage 2, a detailed list of the attrition can be seen in Appendix B.

Table 5: Accuracy of Tax-Exempt Status by Institution Type

Public Use Type	Sample N	N Accurate Land Use	% Accurate Land Use
Government Offices	4	3	75%
Medical	5	5	100%
Religious	5	5	60%
Special Use	5	2	40%
Educational	5	5	100%
Parks and Recreation	5	3	60%
Other	1	0	0%
Total	30	23	77%

Concluding Remarks

This study provides a preliminary investigation into the accuracy of the recorded land use of parcels in the 2007 SCAG property records database. While the samples draw only from the Palm Springs area and are by no means statistically representative of either Palm Springs or the database, we find that 85% of taxable developed properties sampled have accurate land uses, 58% of vacant properties sampled were verified as such, and 78% of tax-exempt developed properties sampled appear to qualify for tax-exempt status.

Furthermore, the low accuracy rates for vacant properties is driven largely by the inclusion of built single family homes in the 'under construction' land use classification. In light of the prevalence of this scenario it is quite possible that this misclassification is widespread throughout the database. Treatment of properties under construction may differ across counties. This issue warrants further investigation into how different counties treat these properties, the magnitude of the observed misclassification problem, and potential paths for correction.

One potential source of the inaccurate land use data is the origin of the land use classification. It is currently unclear where the 2008 land use field came from – was it a field in the original SCAG database, or was it constructed by researchers as part of the LA-Plan project? More importantly, is there enough information in the property records to identify and reclassify 'under construction' parcels that were actually completed buildings? These unanswered questions merit future investigation.

Appendix A – Mapping of Land Use Codes, SCAG Codes to LA-Plan Codes

The following table is adapted from Table 2 in the appendix of Gu and Arnott (2011). The SCAG land use codes are based upon the Southern California Aerial Land Use Consortium’s 1993 land use classification, as developed by Aerial Information Systems, Inc. The LA-Plan Codes aggregate this classification system into 13 broad land uses suitable for the project’s purposes.

Table A1. Land Use Code Mapping

LA-Plan Land Use Code	LA-Plan Land Use Description	SCAG Land Use Codes
1	Single-Family Residential	1110 Single Family Residential 1111 High-Density Single Family Residential 1112 Low-Density Single Family Residential
2	Multi-Family Residential	1120 Multi-Family Residential 1121 Mixed Multi-Family Residential 1122 Duplexes, Triplexes and 2-or 3-Unit Condominiums and Townhouses 1123 Low-Rise Apartments, Condominiums, and Townhouses 1124 Medium-Rise Apartments and Condominiums 1125 High-Rise Apartments and Condominiums
3	Mixed Residential	1100 Residential Parcels Lacking Detailed Classification 1130 Mobile Homes and Trailer Parks 1131 Trailer Parks and Mobile Home Courts, High-Density 1132 Mobile Home Courts and Subdivisions, Low-Density 1140 Mixed Residential 1150 Rural Residential 1151 Rural Residential, High-Density 1152 Rural Residential, Low-Density
4	Commercial - Office	1210 General Office Use 1211 Low- and Medium-Rise Major Office Use 1212 High-Rise Major Office Use 1213 Skyscrapers
5	Commercial - Retail	1220 Retail Stores and Commercial Services 1221 Regional Shopping Center 1222 Retail Centers (Non-Strip With Contiguous Interconnected Off-Street Parking) 1223 Modern Strip Development 1224 Older Strip Development
6	Commercial – Other	1200 Commercial Lacking Detailed Classification 1230 Other Commercial 1231 Commercial Storage 1232 Commercial Recreation 1233 Hotels and Motels 1234 Attended Pay Public Parking Facilities
7	Public	1240 Public Facilities 1241 Government Offices 1242 Police and Sheriff Stations 1243 Fire Stations 1244 Major Medical Health Care Facilities 1245 Religious Facilities 1246 Other Public Facilities 1247 Non-Attended Public Parking Facilities 1250 Special Use Facilities

		1251 Correctional Facilities 1252 Special Care Facilities 1253 Other Special Use Facilities 1260 Educational Institutions 1261 Pre-Schools/Day Care Centers 1262 Elementary Schools 1263 Junior or Intermediate High Schools 1264 Senior High Schools 1265 Colleges and Universities 1266 Trade Schools and Professional Training Facilities 1270 Military Installations 1271 Base (Built-up Area) 1272 Vacant Area 1273 Air Field 1274 Former Base (Built-up Area) 1275 Former Base Vacant Area 1276 Former Base Air Field
8	Warehousing	1340 Wholesaling and Warehousing
9	Industrial	1300 Industrial Lacking Detailed Classification 1310 Light Industrial 1311 Manufacturing, Assembly, and Industrial Services 1312 Motion Picture and Television Studio Lots 1313 Packing Houses and Grain Elevators 1314 Research and Development 1320 Heavy Industrial 1321 Manufacturing 1322 Petroleum Refining and Processing 1323 Open Storage 1324 Major Metal Processing 1325 Chemical Processing 1330 Extraction 1331 Mineral Extraction - Other Than Oil and Gas 1332 Mineral Extraction - Oil and Gas
10	Transportation, Communication, and Utilities	1400 Transportation, Communications, and Utilities 1410 Transportation 1411 Airports 1412 Railroads 1413 Freeways and Major Roads 1414 Park-and-Ride Lots 1415 Bus Terminals and Yards 1416 Truck Terminals 1417 Harbor Facilities 1418 Navigation Aids 1420 Communication Facilities 1430 Utility Facilities 1431 Electrical Power Facilities 1432 Solid Waste Disposal Facilities 1433 Liquid Waste Disposal Facilities 1434 Water Storage Facilities 1435 Natural Gas and Petroleum Facilities 1436 Water Transfer Facilities 1437 Improved Flood Waterways and Structures 1438 Mixed Utilities 1440 Maintenance Yards 1450 Mixed Transportation 1460 Mixed Transportation and Utility
11	Mixed	1500 Mixed Commercial and Industrial 1600 Mixed Urban
12	Vacant	1700 Under Construction 1800 Open Space and Recreation

		1810 Golf Courses 1820 Local Parks and Recreation (1990 Database only) 1821 Developed Local Parks and Recreation 1822 Undeveloped Local Parks and Recreation 1830 Regional Parks and Recreation (1990 Database only) 1831 Developed Regional Parks and Recreation 1832 Undeveloped Regional Parks and Recreation 1840 Cemeteries 1850 Wildlife Preserves and Sanctuaries 1860 Specimen Gardens and Arboreta 1870 Beach Parks 1880 Other Open Space and Recreation 2000 Agriculture Lacking Detailed Classification 2100 Cropland and Improved Pasture Land 2110 Irrigated Cropland and Improved Pasture Land 2120 Non-Irrigated Cropland and Improved Pasture Land 2200 Orchards and Vineyards 2300 Nurseries 2400 Dairy, Intensive Livestock, and Associated Facilities 2500 Poultry Operations 2600 Other Agriculture 2700 Horse Ranches 3000 Vacant 3100 Vacant Undifferentiated 3200 Abandoned Orchards and Vineyards 3300 Vacant With Limited Improvements 3400 Beaches (Vacant)
13	Other	4000 Water 4100 Water, Undifferentiated 4200 Harbor Water Facilities 4300 Marina Water Facilities 4400 Water Within a Military Installation 4500 Area of Inundation (High Water) (1990 Database only) 9999 Missing/Not Included/Other

Appendix B – Detailed Attrition of Land Use Accuracy

This appendix provides additional information on the qualitative nature of land use changes and/or discrepancies observed during the three stages of ground-truth tracking. Note that, due to the 5-year time lag between the determination of the 2008 land use and the 2013 ground-truth tracking, it is not clear if the differences between the recorded and observed land uses are due to a change in the properties use within the last five years or to incomplete and/or inaccurate updates in the SCAG database.

Stage 1 – Of the 100 non-vacant, non-tax-exempt properties sampled:

- 81 single family residential properties were sampled, 75 of which were confirmed as such - 2 had changed to alternative residential use and 4 were now vacant;
- 6 multi-family residential properties were sampled, 3 were verified as such - 1 was converted to single family residential use, 1 to commercial use, and 1 was now a vacant property;
- 5 mixed residential properties were sampled, 1 was confirmed as such - 2 mobile homes had been changed to permanent single family residences and 2 rural residences were now vacant;
- 4 retail commercial properties were sampled, 3 were confirmed as such – 1 was vacant;
- 3 transportation/utility properties were sampled, 2 were as listed - 1 airport is now a single family residence.

Stage 2 – Of the 100 vacant properties sampled:

- 4 vacant agricultural properties were sampled, 3 of which were confirmed as such – 1 was in fact a mobile/manufactured home built in 1965 (verified by real estate website). All four of these properties had planned land uses of mixed residential;
- 11 vacant recreation properties were sampled, 10 of which were indeed vacant - 1 did not meet the working definition of vacant, but was in fact a single family residential built in 2004. However, 3 of the vacant properties were classified as vacant for the wrong reason – they were currently under construction and not recreation properties (note that all 4 of these parcels had planned land use codes of single family residences);
- 85 vacant properties were listed as ‘under construction,’ 45 of which were verified as still under construction - 1 property was a road, 1 was a completed church, 1 was a multi-family residence, 2 were completed ‘park and ride’ lots, 34 properties were completed single family residences, and 1 could not be determined.

Stage 3 – Of the 30 tax-exempt properties sampled:

- 5 parks and recreation properties were sampled, 3 of which were parks - 2 were determined to be privately owned tennis courts;
- 5 educational properties were sampled, 4 of which were indeed schools or school buildings - 1 was a school parking lot;
- 5 special use properties were sampled, 2 of which were vacant special use facility land - 2 were single family residential properties and 1 was a multi-family residence;
- 1 other public property was confirmed to be a single family residence;
- 5 religious institution properties were sampled, 3 of which were confirmed as religious institution structures - 1 was a church parking lot and 1 was a vacant parcel next to a church;
- 5 medical facility properties were sampled, 2 of which were actual medical buildings - 1 was a church next to a medical building, 1 was a road on a medical campus, and the last was a medical facility parking structure;
- 4 government properties were sampled, 3 of which were confirmed as government buildings - 1 was a parking lot for a casino.

Works Cited

Goodchild, Michael et al. *SCAG Parcel Database Validation Report on Accuracy of Total Floor Space Per Parcel by Ground Truth Trekking*. Riverside, CA: Center for Sustainable Suburban Development, 2011.

Gu, Yizhen, and Richard Arnott. *Blanks and Zeroes in the SCAG Parcel Database*. Riverside, CA: Center for Sustainable Suburban Development, 2011.