Comparing California Agricultural Land Use Datasets: Coverage, Accuracy, and Pitfalls

Abstract

Given the extensive agricultural sector in California, knowing what crops are planted where is a necessary perquisite for many research subjects including: water use, environmental pesticide loading, agricultural sediment runoff, crop-climate interactions, and public health environmental toxicity. Unfortunately, no perfect data set exists. All publicly available agricultural land use data sets that are available to researchers have their own set of limitations and errors. This study compare; the USDA Cropland Data Layer, California Department of Water Resources' land use surveys, county crop reports, and two different acreage summations derived from California Department of Pesticide Regulation's Pesticide Use Reports (PUR). Data sets were compared statewide and by county for each crop and for all crops. A statistical analysis comparing the two PUR summations pinpointed counties and crops with large data collection errors. The results of this study can be used by other researchers to guide their choice of land use data and how to minimize errors from their chosen dataset.

Overview

California is home to an extensive agricultural sector with over 400 commodities that cover approximately 27% of the state's land area, uses around 40% of the available water, and generates over $36 billion a year. The environmental, economic, and social impact of California's agricultural sector cannot be overestimated. In order to model past, present, and future impacts to the economy, environment, growers, and neighbors, it is imperative to have an accurate accounting of how many acres of what commodities are grown where. During the last few decades, environmental research and subsequent policy decisions have become more and more dependent on computer models which span spatial scales from a single farmer's field, to the watershed, to global circulation models. But models are only as good as their input data, thus ensuring that the available California agricultural land use datasets are accurate is of essential.

There are four major datasets of varying spatial resolution and time coverage that track California's agricultural land use: California Department of Pesticide Regulation's (CDPR) Pesticide Use Reports (PUR)[1], the United States Department of Agriculture (USDA) National Agricultural Statistics Service's (NASS) Cropland Data Layer (CDL)[2] [3] [4], county agricultural commissioner crop reports compiled by NASS[5], and California Department of Water Resources (DWR)[6] land use survey. Each source claims a high degree of accuracy in representing how many acres of a given crop exist at their declared spatial and time delineations. Yet for many crop/county/year combinations, claimed totals diverged wildly between

the datasets, and sometimes even within a dataset. This paper seeks to detangle the disagreements to determine which should be given more credence over the others for a given crop and county.

California's Pesticide Use Report (PUR) database is the most comprehensive and complete database of its kind in the world, having tracked every commercial pesticide application in the state since 1990. This public dataset has been utilized by researchers worldwide for topics ranging from land use modeling to epidemiology with around 500 peer reviewed papers citing the database[7]. Because the PUR records contain spatial resolution of crops down to the square mile and time resolution of pesticide applications down to the hour, it makes for a very attractive dataset to use in agro-environmental modeling, both to model land use and to model the fate and transport of applied chemicals. Unfortunately, several county/crop combinations have acreage totals orders of magnitude off from other datasets or even common sense.

The results for many county/crop combinations show a fairly strong agreement between PUR and the crop reports. But there are several major pitfalls that must be taken into account before using the data in modeling or decision making. Truck crop (fresh annual vegetables) PUR acreage in the central coast (Monterey to Ventura) went from a strong agreement with the crop reports to orders of magnitude greater than the crop reports, indicating a dramatic shift for the worse in how the counties collected and entered the data. Low value crops like Sudan grass and crops with minimal pest pressure like avocados show much lower PUR acres compared to the county crop reports due to many fields not being treated during the year. Land use studies that consider these crops should take that into account and thus avoid only relying on PUR acre planted data.

The CDL is only useful for field crops, rice, and some deciduous tree crops. It is poor for delineating vineyards, truck crops, or many grain crops. Even the crops that perform well, often only attain that high level of correlation in the last couple years of data collection. The hit and miss nature of the CDL means that researchers should double check CDL acreage totals against other data sources for the crops and counties being studied before using it in their research.

Materials & Methods

Datasets

California Department of Pesticide Regulation's (DPR) Pesticide Use Reports (PUR)

Since 1990, every application of an agricultural chemical, aside from fertilizer, has been recorded in a publicly available database. Non-agricultural, commercial, pesticide and herbicide applications are also recorded, but at a county level spatial resolution rather than a Public Land Survey System section (~1mi^2) level and a monthly temporal resolution rather than a sub-day resolution. In addition to the type and amount of product applied, each row of the database also includes, among other things, the acres treated in that application and the total acres planted in the field. Each grower is identified by a unique ID and each field is identified by a, unique to the grower, ID. Since there is a legal mandate to collect this information, there is no evidence of even minimal noncompliance, and most crops, even organics, are

treated with something other than water and fertilizer during the season, these reports should produce an accurate representation of what was planted where down to a one square mile spatial resolution, and a sub-hour time resolution.

Each pesticide use report is filed by the grower with the county where their crop is located. The county then sends the reports to DPR. Until recently, all grower reports were submitted on paper, with county personnel transcribing the report into digital form and bulk sending them to the state. A centralized web platform is being phased in. Hopefully, in the future, all growers will submit their reports digitally so as to eliminate the transcription step which represents another layer of possible error. DPR receives the reports throughout the year then processes them with an extensive quality assurance data analysis in order to spot and rectify as many data errors as possible. If necessary, DPR will notify the county that they need to recheck a given report and resubmit. Once DPR is satisfied with the data quality, they publicly release a year's worth of data in one large compressed file. Small excerpts of the full public dataset can be extracted on a year-to-year basis using DPR's California Pesticide Information Portal[8]. The lag between data generation and public availability of the data can stretch to years, but DPR strives to release the previous year's data as promptly as possible.

The first few years of data collection were plagued with teething pains resulting in some highly questionable results. Therefore, this analysis only includes reports from 1995 to 2015, the last year of publicly released data as of the time of this study.

United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL)

The CDL is a 30m resolution, raster data layer generated by the USDA for the contiguous, continental United States on an annual basis. These rasters were created by an automated classification algorithm using satellite imagery from a multitude of sources including Landsat 8, the UK Disaster Monitoring Constellation, and the following USGS data layers: National Elevation Dataset, National Land Cover Dataset Imperviousness, National Land Cover Dataset Tree Canopy. Training and validation utilized the USDA Common Land Unit Data, the USGS National Land Cover Dataset, and vineyard locations identified by E.&J. Gallo winery.

CDL layers for California are available starting with the 2007 growing season, continuing to the present. National data is available from 2008 onwards and some states have layers for years going back to as early as 1999 (North Dakota goes back to 1997). Layers are generally released promptly at the end of the year, allowing for timely analysis of the previous season.

Luman and Tweddale[9] found a high statewide accuracy for the CDL 2007 Illinois data set. But since corn and soybeans account for over half of Illinois total surface area, their results might not be applicable to California's high crop diversity. Yang et al[10] were able to increase the accuracy of the CDL for Texas corn and sorghum using a multi-level algorithm that utilized other land use layers like roads, waterways, and the public land survey system grid. These techniques were not applied in this study and given the extreme year to year variation for many county/crop summations, it is unlikely that this methodology would be as effective in California.

USDA NASS County Crop Reports, compiled by the California County Agricultural Commissioners

This dataset is an annual, county wide summation of how much of what crops were grown, how much was harvested, the yield per acre, and value of the harvested products. The counties use a combination of production agriculture permits they have issued, surveys administered to all growers, and water use data in order to compile their statistics. Due to the close working relationship between growers and the county agriculture department, the final crop report totals should be highly accurate.

Each county usually releases their crop report during the summer of the following year. Once all counties have released their crop reports, NASS compiles them all into a common format and publicly publishes a statewide summary. By adding a time lag on top of the time required for the tardiest county to release their data, NASS county crop report summaries can take over a year to be published.

California Department of Water Resources (DWR) Land Use Survey

Every year, starting in 1986, DWR produces a vector data layer of land use for 1-10 counties with a focus on agricultural land use. Individual field boundaries are delineated using a combination of aerial photos, satellite imagery, and GPS ground mapping. DWR staff personally visits over 95% of developed agricultural areas within each survey area. The crop of each field is identified using two level naming scheme where the primary class represents the general crop type (field, truck, tree, vineyard, etc) while the secondary class represents the specific crop (oranges, almonds, avocado, etc). It is also possible to note multiple cropping of the same field.

Because of the labor intensive mapping, this dataset should be very accurate spatially, both in total acres planted and where those acres are located within the county. Unfortunately, because only a handful of counties are surveyed per year, many counties have only been surveyed once over the last couple of decades, thus it is impossible to use this dataset for time series studies or to study recent trends.

Table 1: Dataset comparison

	CDPR PUR	CDL	Crop Reports	DWR Surveys
Spatial Resolution	$1mi^2$	$30m^2$	County	< 1ft
Time Resolution	Sub-hour	Annual	Annual	Annual
New Data Lag	Year+	Weeks to Months	~6 months	Year+
Coverage	Statewide	Statewide	Statewide	1-10 counties/year

Methodology

All datasets were imported into a PostgreSQL database utilizing the PostGIS spatial extensions for both vector and raster data.

The CDL was summarized for each year by first extracting the crop related pixels, converting the pixels to vectors, merging the vectors by crop type, splitting the merged vectors by county boundaries, then finally summing the area by county.

In order to summarize the PUR, first each distinct combination of grower ID, site location ID, site code (aka crop), county, and year were extracted. This should represent each distinct field for each year. For each of these distinct fields, the average declared acres planted and maximum declared acres treated were calculated. Occasionally, declared acres planted changes from one report to another either by transcription error, or grower misunderstanding. Usually, enough applications happen over a season such that erroneous values can be mitigated against through averaging. CDPR has an internally developed, non-public table that applies a far more rigorous method to identify distinct fields and what their most accurate acres planted should be. While this process does identify and fix many errors, when summarizing at the annual county level, the difference between a naive average and the CDPR fields table is often minuscule. The theory behind collecting both the average declared acres planted and the maximum acres treated is that those providing, collecting, and paying for this data really care about how much product was applied to how many acres. There is a greater chance for erroneous acres planted data to slip through compared to pounds product applied or acres treated. Since most crops that receive treatment of any sort during the season will have the whole field treated at once with at least one of the products, taking the maximum value for acres treated should also result in an accurate portrayal of how many acres were actually planted.

DWR vectors were simply summed by area for the county and year in which the survey took place. Crop reports were processed even less as they were only summarized by DWR primary and secondary classes along with the PUR and CDL datasets.

Each dataset had a different naming and numbering scheme to identify a specific crop. The DWR dataset uses a pair of two-character IDs that specify the general crop class (truck crops, field crops, tree crops, vineyards, etc), then the specific crop type (asparagus, barley, citrus, etc). There are twenty-three primary classes and 175 individual combinations of primary and secondary classes. Not all these are agriculture products, for instance classes also exist for urban and native land uses. The county crop reports use a five or six digit commodity code that correlates with a crop name. There are 428 distinct commodity codes restricted to purely agricultural products including both plant and animal commodities. The CDL has 89 distinct values for each raster cell, and like DWR, includes classes for non-agricultural land uses. The PUR uses a 2-5 digit integer site code for each record. The vast majority of the 494 distinct site codes are for agricultural products, but also include classes like 'structural pest control' and 'landscape maintenance'.

The DWR naming scheme was chosen as the lingua franca of all datasets because it had one of the fewest number of classes and had inherent flexibility to summarize by even more general commodity classes. For the CCR, CDL, and PUR datasets, each distinct class was hand assigned to a DWR primary and secondary class. These lookup tables were imported into the database then used to summarize each dataset at the annual, county level using DWR's naming scheme. This county/year time and space discretization was chosen to match the resolution of the coarsest dataset: the county crop reports.

The final summarization table was then graphed and analyzed. Only crops that had over a thousand acres harvested during at least one year in the Crop Reports were surveyed. When only a few fields are present, small mistakes can result in large errors. Also, from both an environmental and economic

standpoint, a few hundred acres of anything is little more than a rounding error. Therefore, crop/county combinations that average in the hundreds of acres per year were discarded prior to analysis. Because the focus of this analysis is in improving the quality of data sources over time, emphasis will be placed on crops whose acre totals diverge over time verses those that converge.

Results were summarized statewide by crop, then countywide for all crops, and finally for each county/crop combination. In addition to graphical plots of the above summarizations over time, a statistical trend analysis was performed on the difference between paired datasets. While identifying time consistent differences between the datasets is important, these differences can usually be easily explained and rectified with a simple scaling factor. More intriguing are cases where datasets diverge or converge over time and thus point to either a breakdown in record keeping practices, an improvement in data consistency, or in the case of PUR & crop reports a change in pesticide usage rates.

In order to determine if the datasets are consistent with each other over time, first the residuals were compiled by finding the difference on an annual basis for all county/crop pairs for three data pairings: PUR acres planted minus PUR acres treated, county crop reports minus PUR acres planted, and county crop reports minus PUR acres treated. The CDL was ignored due to the low number of years available and the excessive year-to-year variability across crops and counties.

Next, an augmented[11] Dickey-Fuller test[12] was performed on the residuals to extract the county-crop pairs that statistically couldn't have a unit root rejected (ie were not stationary over time) at the 90%, 95%, & 99% confidence intervals. A five year lag was used. These non-stationary residual series then had a line fit to them using simple linear regression. The slope of the fitted line is used to identify if the series is converging or diverging and as a metric to quantify the magnitude of change over time.

TODO: KERN? DWR?

Results

Ideally, all datasets should match for every crop for every year statewide. Some commodities actually come close to this ideal, with cotton being the best example (Figure 1). Any of the datasets could be used for cotton with confidence at their native time and space resolutions. This accuracy is likely due to cotton's unique spectral signature combined with its highly organized growing community and need to treat every acre every year with at least one agricultural chemical.

Other crops that had a least some sort of visibly recognizable correlation between datasets at a statewide level include grapefruit, lemons, rice, and sugar beets (Figure 2, Figure 3, Figure 4, Figure 5). While these crops tracked each other well over time, there was still a significant discrepancy in acreage between them and, in the case of lemons and pears, this discrepancy is increasing over the last few years.

Summing all crops for a given county can indicate how well the county performs their data collection in aggregate. An exceptionally good example of a county accurately reporting their acreage to all data collectors is Fresno County (Figure 6). Aside from an odd spike in 2011, all data sources track each other

fairly well. One advantage Fresno has is that the large number of crops, growers, and acres in the county helps filter out errors from individual fields or even problems with specific specialty crops. Counties with tens of thousands of total cropland or less tend to have a wider spread between PUR average acre planted, PUR max acres treated, and the crop report harvested acres. Other counties that show an acceptable visual correlation between PUR and the CCR over all crops include: Butte, Colusa, Imperial, Glenn, Kings, Madera, Merced, Sacramento, Riverside, Tulare, Sutter, Yuba, and Yolo.

Counties whose data does not track each other well tend to either be counties with a small amount of cropland acreage where the PUR totals oscillate wildly from year to year or counties where the PUR average acres planted diverges from the PUR max acres treated in a consistently widening trajectory. An example of the former is Inyo county which never goes over 5,000 acres a year for all crops (less than most individual crops in Fresno). Errors between datasets for these types of counties is less worrisome because there is a minimal amount of acreage with an associated minimal amount of applied chemicals and therefore will not generate enough revenue or pollution to be worth the effort to model.

More worrisome are counties that display a huge, and growing, chasm between the two main PUR metrics. Some divergence between maximum acres treated and average acres planted can be explained away by recognizing that the assumption that all acres will be treated at some point may not hold for all fields and all crops. But when the spread enlarges to many multiples of the total CCR acres, something is amiss. Monterey county is the worst offender in this category. PUR max acres treated tracks the CCR quite well, but the average acre planted and the internal field table summations wind up being up to six million acres greater. Since the total acreage of the whole county is less than 2.5 million acres and irrigated land is only 220,000 acres, something is definitely amiss. Santa Barbara displays similar behavior to Monterey, but to a lesser degree and starting later in time.

Divergence between the two PUR metrics indicates a problem with how the records are initially recorded or a severe amount of individual erroneous records. It might be possible for a given crop to generally only be spot treated during a growing season which could explain some discrepancy between the two PUR statistics, but even then, the two lines should track well over time. For many crops, especially vegetable truck crops, there is a huge and growing discrepancy. The worst offenders statewide are artichokes, broccoli, Brussels sprouts, cabbage, carrots, cauliflower, lettuce, onions & garlic, peppers, spinach, and strawberries. Of these, lettuce is the worst (Figure 7). These also happen to be the crops predominantly grown by Monterey (Figure 8), Santa Barbara (Figure 9), and Ventura (Figure 10) which points to the county being the problem, not the crop.

Another curious discrepancy is when PUR data diverges strongly from the County Crop Reports. Since the county agricultural commissioner is responsible for collecting and disseminating both datasets, they should align much better when compared to datasets produced by other agencies. The crop with the greatest divergence between PUR and crop reports is Sudan grass (Figure 11) in which the crop report acres are consistently much higher than the PUR acres. Imperial county accounts for the majority of Sudan grass acreage in the state. Through conversations with Margo Sanchez, the Imperial county deputy agricultural commissioner in charge of data collection, the discrepancy can partially be explained by the fact that Sudan grass is a very low value crop and as such, there is little economic incentive to

spend money to treat the field with anything. Imperial county estimates that over 50% of the Sudan acreage each year is not treated and thus would not show up in the PUR database. Additionally, multiple cuts of the same field could be recorded as that many multiples of the field's planted acreage in the CCR's “harvested_acres” column.

A similar, but less striking, example is Avocados where up until 2011 there was a large, but consistent, spread between the crop report acres, the PUR average acre planted, and the PUR maximum acres treated (Figure 12). The DPR internal field table tracks the PUR average acre planted almost exactly. Through conversations with sources who have worked the Avocado ranches of San Diego for many years, most farmers only occasionally applied Glyphosate for weed control. Of those that do apply Glyphosate, many will purchase Round-Up from the local hardware store in order to avoid the paperwork overhead of filing a pesticide use report.

The last major error class is between the Cropland Data Layer and the other data sources. Due to its timely updates, annual time resolution, 30m spatial resolution, and continuous coverage, the USDA Cropland Data Layer is very enticing for modeling tasks that depend on knowing the spatial distribution of agricultural land use. Unfortunately, for many crops, its totals are inconsistent from year to year compared to a consistent PUR and county crop report trend over the same time period. Additionally, the totals themselves are often far different than the other data sources.

Using DWR's initial class breakdown of crop types and summarizing statewide, the CDL performs fairly well for deciduous fruits and nuts (D: Figure 13), field crops (F: Figure 14), and rice (R: Figure 15). Even then, the first year recorded for rice and the first couple of years for deciduous crops are much different than trend set by other CDL years and the other data sources. Wild swings of acreage for orchard crops from year to year are highly unlikely given the lifespan of the crops and the capital required to plant new acreage. The other major crop categories had unusable results: grain & hay crops (G: Figure 16), truck crops (T: Figure 17), citrus & subtropical crops (C: Figure 18), and vineyards (V: Figure 19). That vineyards are included is surprising given that the texture of a vineyard on satellite imagery is quite distinct and very different from other land uses. The USDA also received training data from California grape growers, so theoretically, the classification algorithm should perform better on this crop than others that were trained on crops outside California. Before using the CDL for crop analysis, it is imperative to check that crop's county and/or state total over time against the CCR or PUR to make sure that the CDL totals are within reason.

Trends in the residuals

Table 1 includes the ten highest slopes of (PUR acre planted – PUR acre treated) residuals by crop statewide. Crops with positive slopes are showing a worsening of PUR data entry errors while negative slopes indicate that the two PUR metrics are converging. Truck crops, lettuce especially, dominate the list of worsening crops while cotton, wheat, and rice have the greatest improvement. The residual plot of "Flowers, Nursery & Christmas Tree Farms" shows that data points for 1995-1997 are throwing off the linear interpolation with the remaining years being more or less consistent with a slope close to zero.

Table 2 lists the slope of PUR residuals for all crops within each county. Monterey stands out as a very problematic county with a slope over forty times as large as the next highest county. Around 70% of the counties with a significant residual trend are getting better over time.

Most of the statewide, truck crop, intra-PUR acre discrepancies can be attributed to Monterey. Starting around 2004, the acres planted verses acres treated diverged by millions of acres. Through conversations with the office of Monterey county's Agricultural Commissioner’s Office, one explanation was that the county has changed how it defines the 'site_loc_id' field in the Pesticide Use Reports: instead of identifying a specific field that grows a specific crop, instead it defines a 'Ranch' which is a large physical area where multiple, different crops can be grown throughout the season. While this may make things easier for the grower and the county with regards to data collection, it makes using the resulting data extremely hard for the end user.

In order to divine the true cause of the Monterey discrepancy, a table was created with columns for year, grower_id, and site_loc_id as the keys that defined a field and columns of statistics for each field including minimum, maximum, average, and standard deviation of acres planted and treated along with a list of crops that were associated with that field (available as appendix document monterey_field_stats.xlsx).

Starting around 2005, when the divergence between PUR and the CCR started for Monterey crops, there appeared many fields which had recorded acres_planted (both min and max) at an absurdly high level, often larger than a section’s 640 acres. A visual inspection of satellite imagery for every section in the Salinas valley with at least one associated PUR record identified zero fields that were close to a section in size. This includes fields that obviously had multiple crops present and would thus qualify as a 'Ranch' under Monterey's vague definition.

Even stranger is when a grower has multiple, distinct, site_loc_id fields and each field is larger than a section. In 2013, grower ID 27132700164 had 126 distinct site_loc_id fields, each with a declared planted acreage of 660ac. The crops grown in these fields were the usual suspects: spinach, lettuce, and broccoli. The max acre treated for each site_loc_id ranged from 7ac to 35ac with most around 10ac.

It seems like in many cases, Monterey is setting the acres planted to be equal the size of the section that the fields reside in. Because of the geography of the Salinas Valley, no grouping of visually distinct fields falls within a section delineation like central valley farm land does. Given that, Monterey’s official explanation that the acres planted represented the whole 'Ranch' the grower operated rather than individual fields within that ranch seems rather suspect. Many other growers had all their fields listed with identical large acreages that were greater than the sum total of the maximum acres treated for each site_loc_id but less than a full section in size.

Worse still, the same intra-PUR divergence of truck crop acres is showing up in Santa Barbara post 2011 (Figure 20). If Monterey's antipathy towards recording accurate data spreads to other counties beyond the central coast, the whole PUR database might become useless for land use modeling. Any argument that proper record keeping is too time consuming or difficult is refuted by Fresno county (Figure 21) which not only grows more acres of truck crops than any of the aforementioned counties, but also has

minimal difference over time between the various PUR acreage metrics. Furthermore, prior to 2004 Monterey was able to record acreage accurately. Since total farmland acreage has not changed significantly, the workload to accurately record planted acres would not have increased.

Conclusions

While the county crop reports are the most accurate accounting of California crop acreage, it is limited spatially to the county level and temporally to the annual level. Many scientific questions cannot be answered or modeled using a data set this coarse, so researchers must turn to other available datasets. All of these alternative datasets have caveats for proper use.

For studies that need sub-county spatial resolution and sub-annual temporal resolution with full statewide coverage, the PUR data set would be a good choice with the caveat that central coast truck crops are very problematic. Until all counties are required to report accurate data for all PUR data fields, when studying central coast truck crops, researchers should use the maximum acres treated per field (defined as the unique combination of grower_id, site_loc_id, and site_code for each year/county). If researching a commodity that has low economic value like Sudan grass, or minimal pest pressure like Avocados, researchers should not totally rely on PUR data for an acreage total as many acres of that commodity might not be treated with a regulated chemical during the growing season. Lastly, individual counties and crops should have their time series analyzed before making any conclusions based on single year deviations from the trend line. For instance, a naïve summation of PUR acres planted for Fresno in 2011 and Merced in 2006 & 2011 results in excessively high acre totals compared to neighboring years and other data sets for those years. These spikes are not present in the PUR acres treated totals.

The CDL should only be used in California for field crops, rice, and some deciduous tree crops. It should not be used for delineating vineyards, truck crops, or grain crops. The CDL has been improving over time for many crops; future releases might converge on other data sources due to tweaks in the classifying tool chain. But due to the historically bad performance of the CDL over a wide range of crops and counties, researchers should compare the CDL data they are using with other data sets before relying on it being correct.

Figures and Tables

Figure 1: Statewide cotton acreage

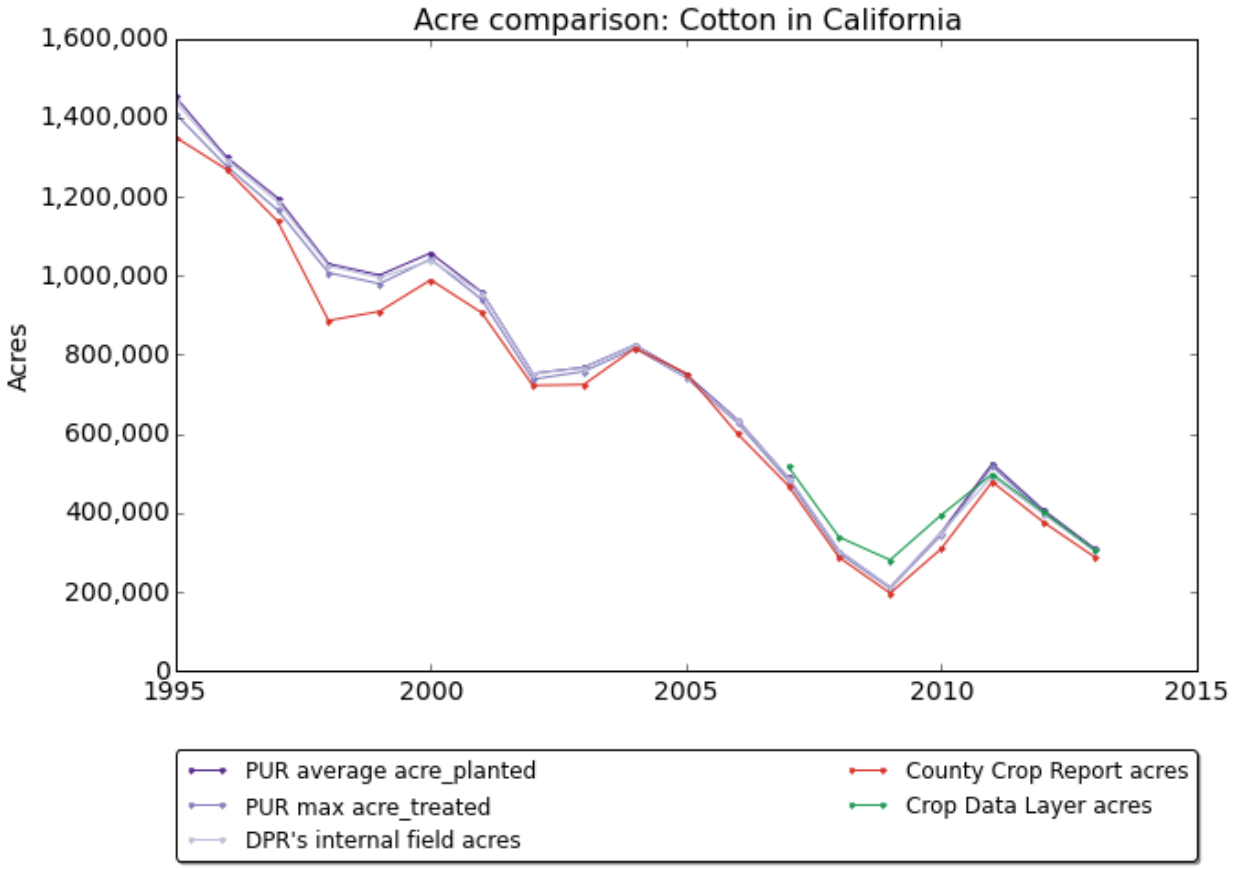

Figure 2: State wide grapefruit acreage

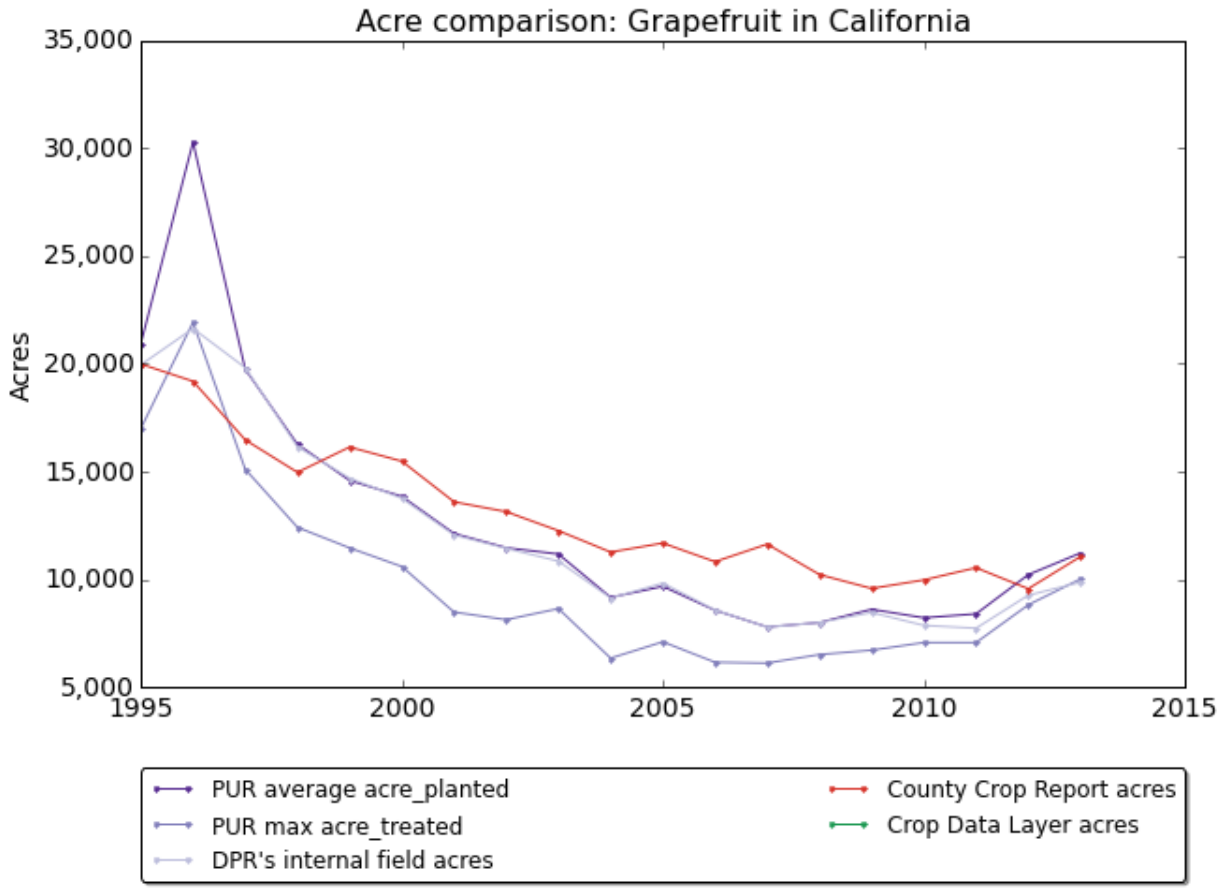
Acre comparison: Grapefruit in California
Acres
35,000
30,000
25,000
20,000
15,000
10,000
5,000
1995
2000
2005
2010
2015
PUR average acre_planted
PUR max acre_treated
DPR's internal field acres
County Crop Report acres
Crop Data Layer acres

Figure 3: Statewide lemon acreage

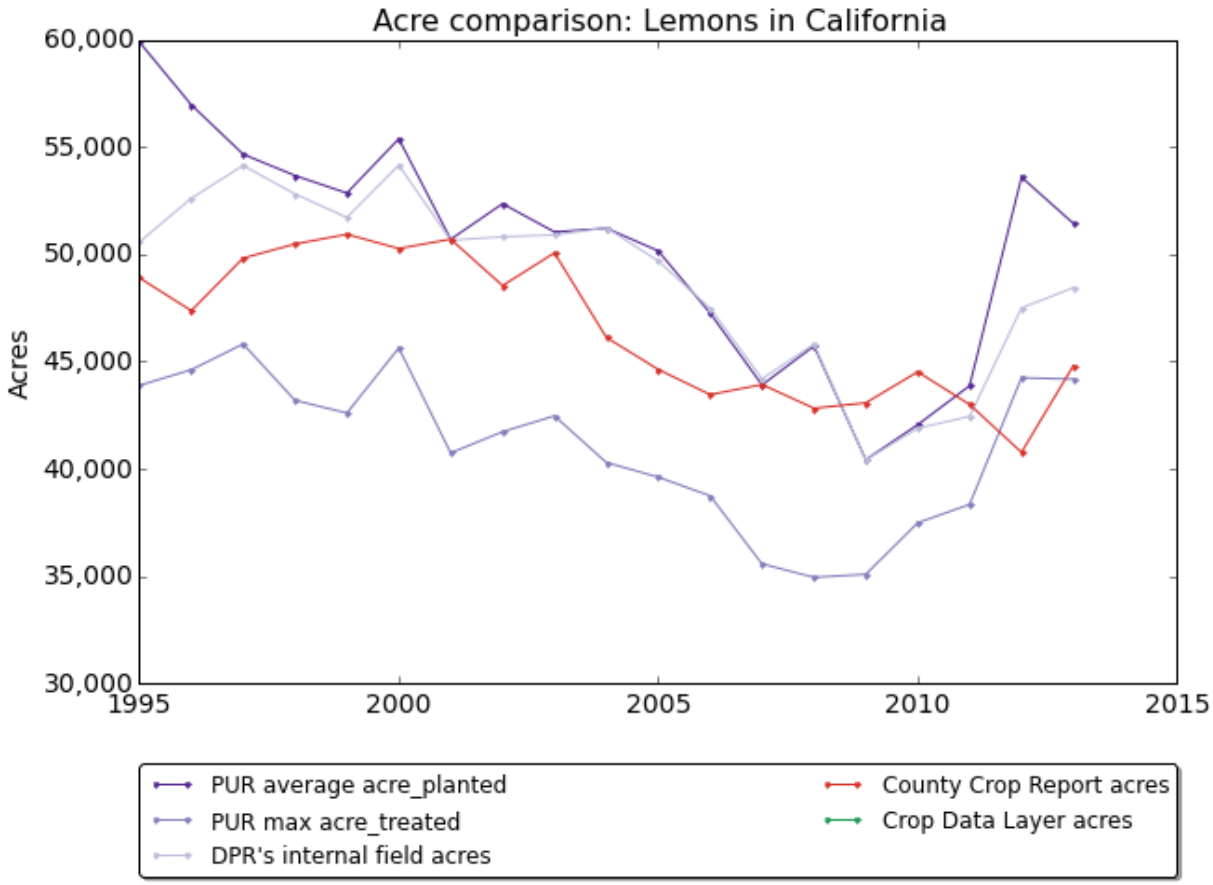

Figure 4: Statewide rice agreage

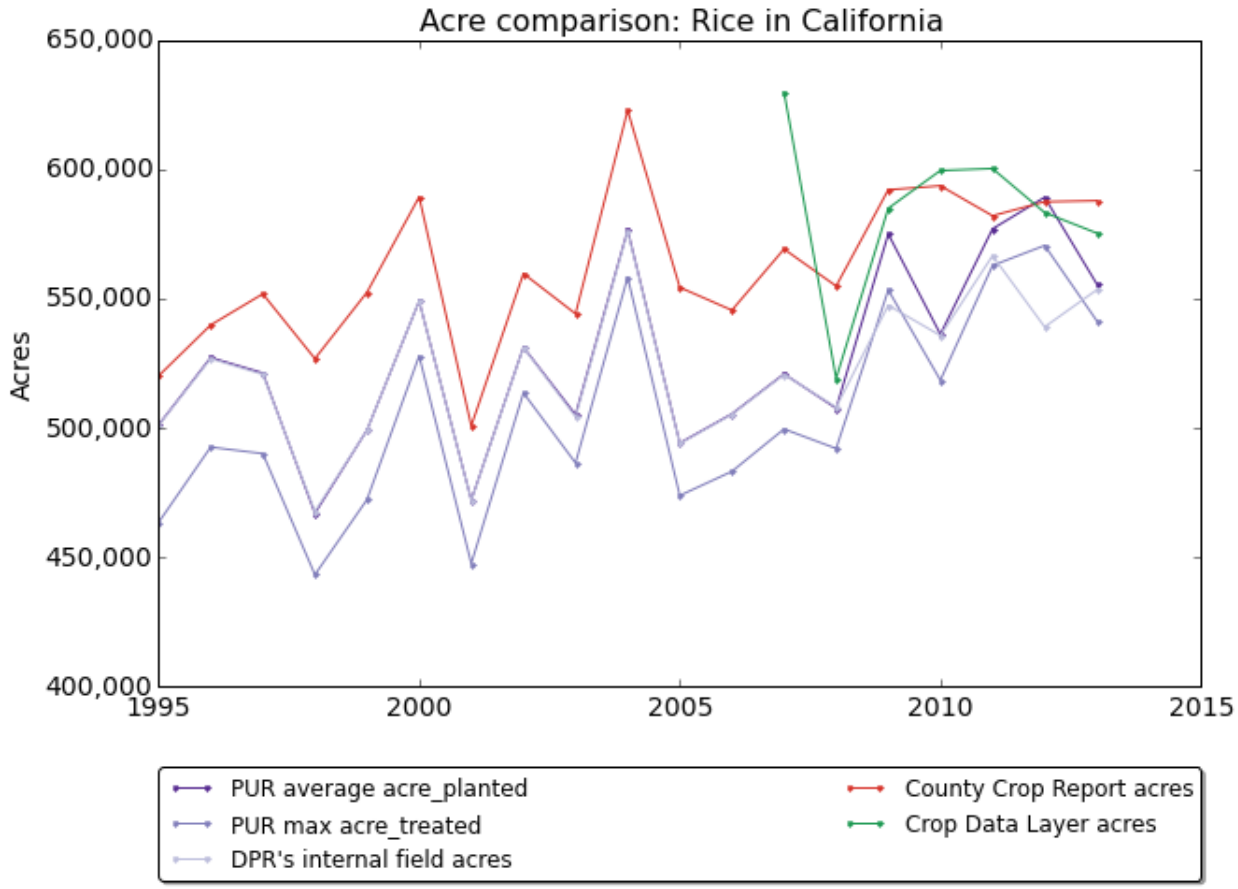

Figure 5: Statewide sugar beet acreage

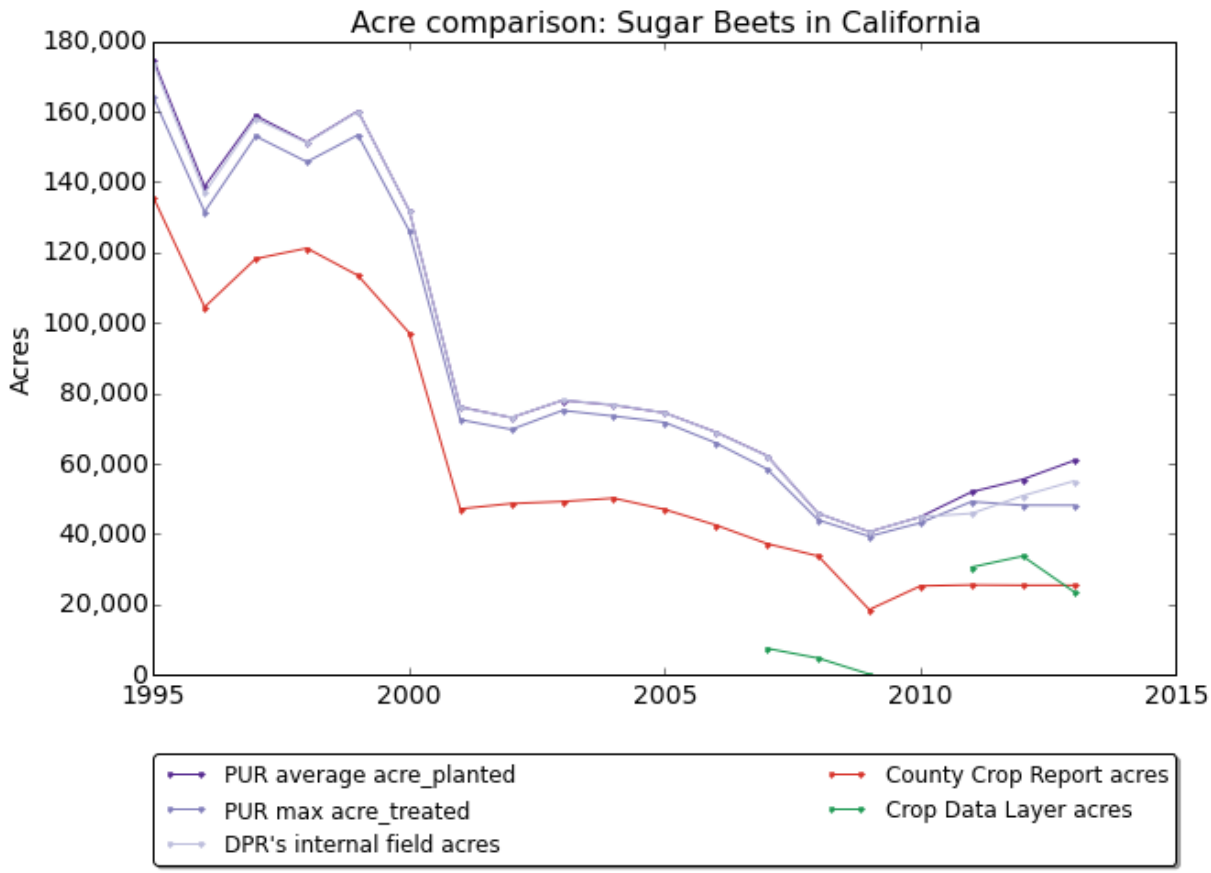

Figure 6: All cropland acres in Fresno county

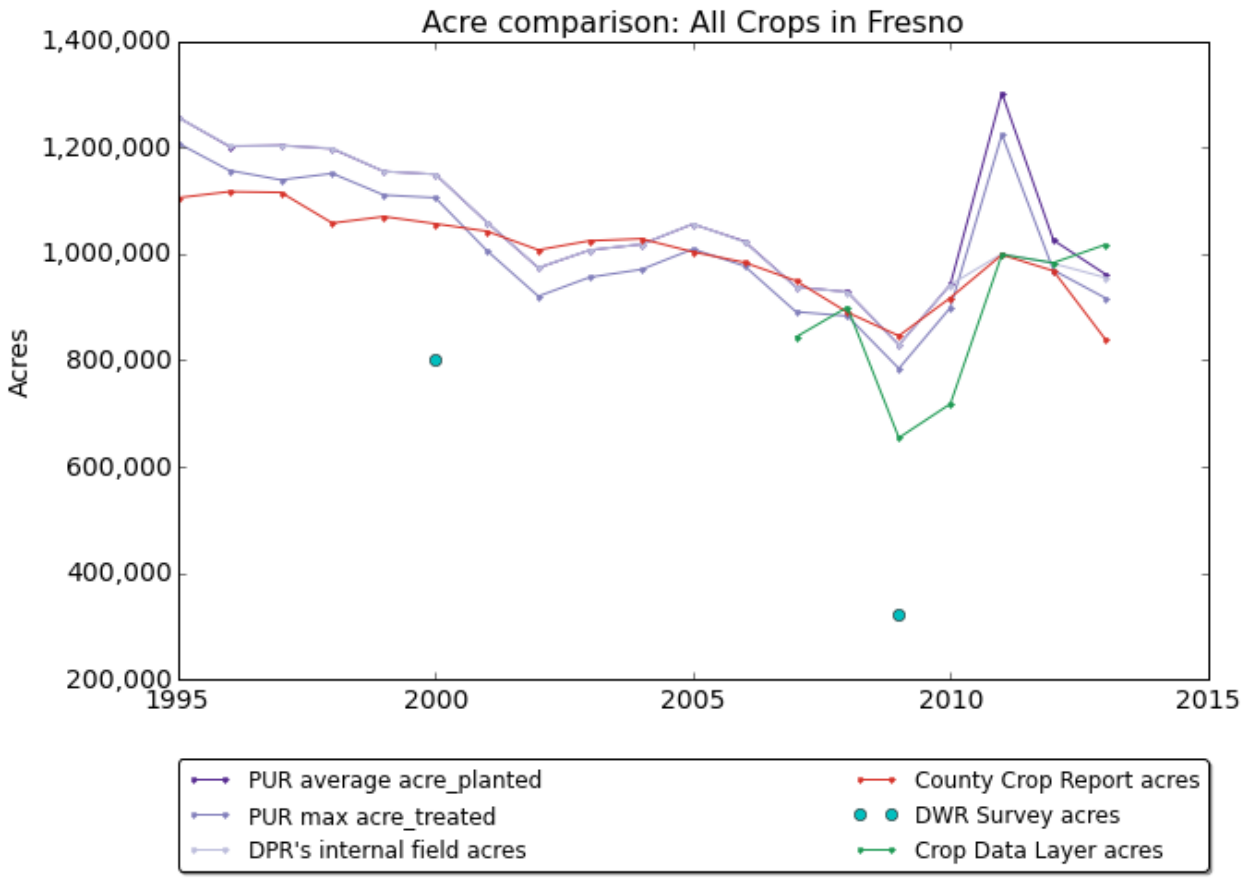

Figure 7: Statewide lettuce acreage totals

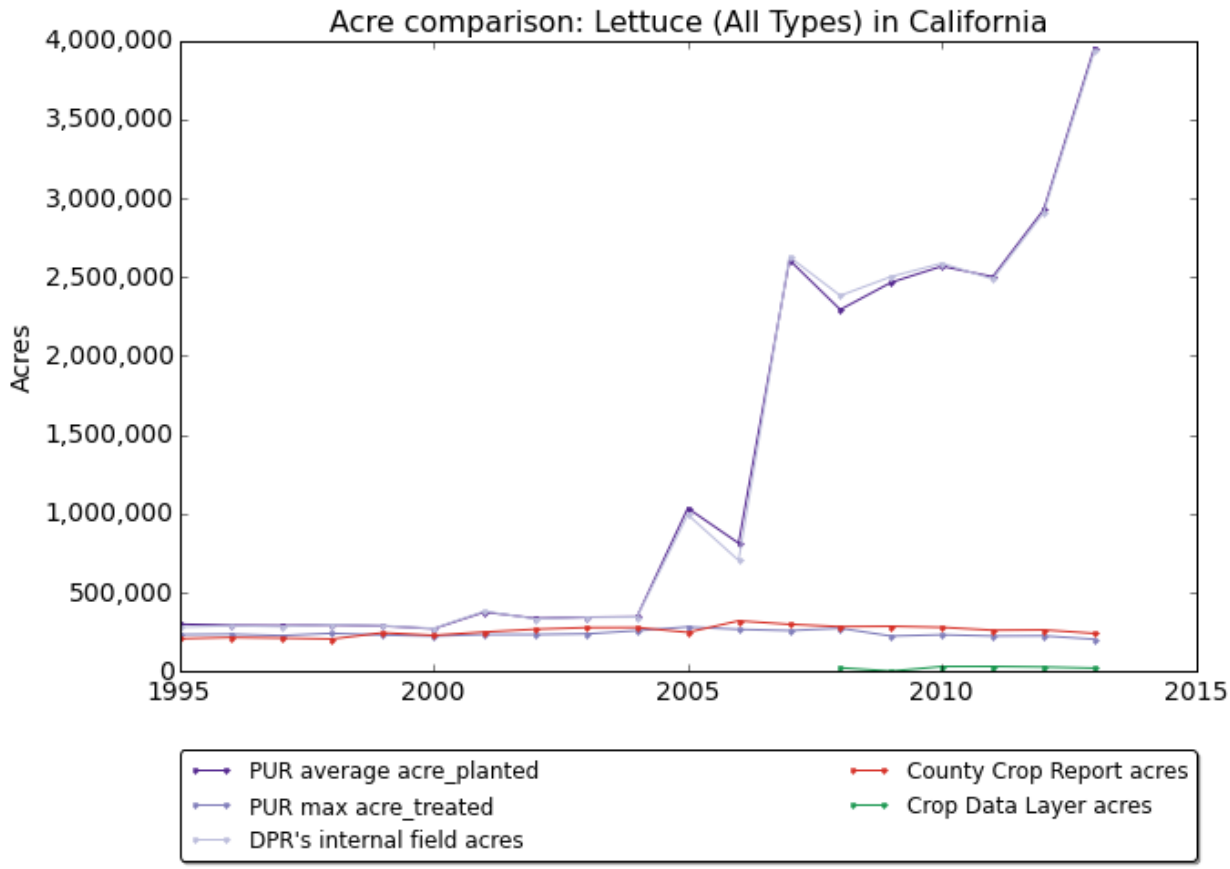

Figure 8: Acre totals for Monterey County for all crops

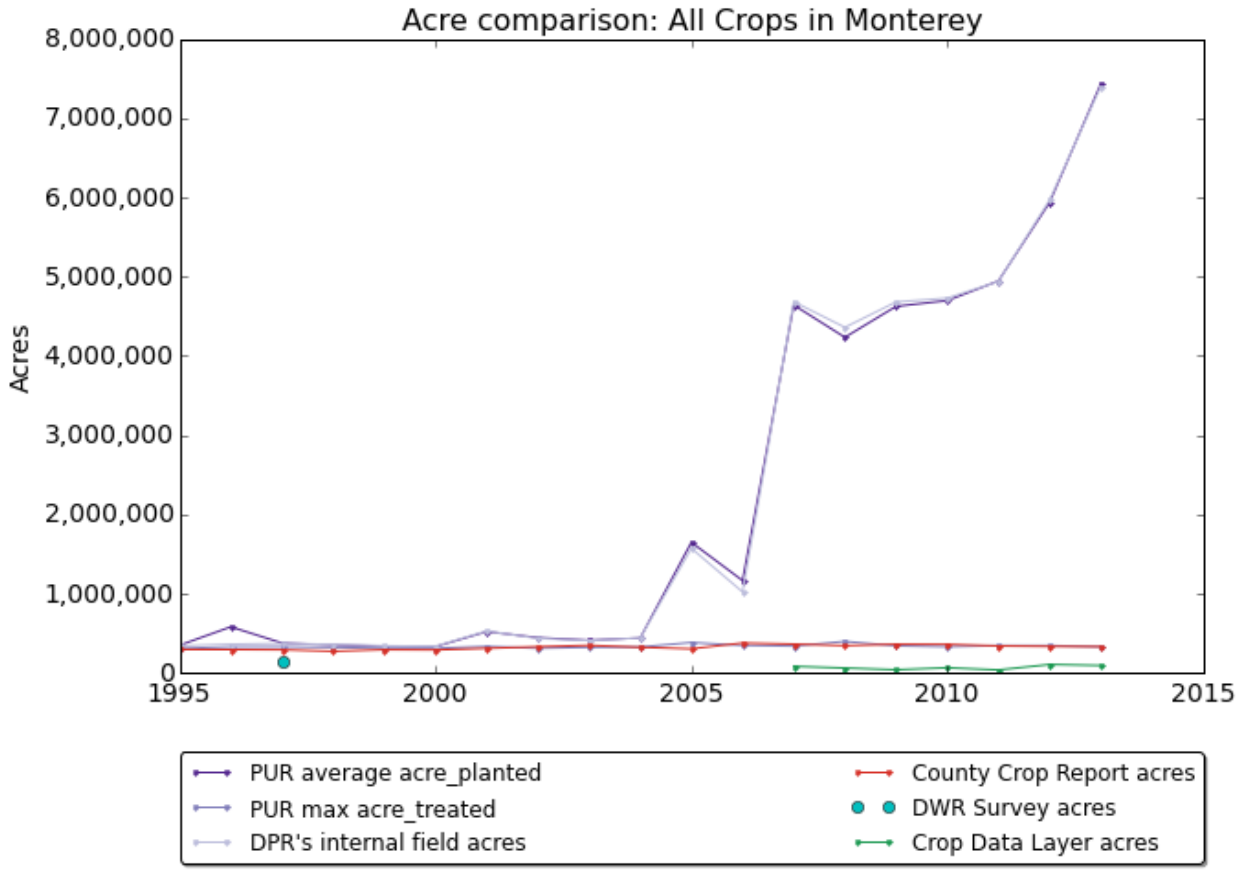

Figure 9: Acre totals for Santa Barbara County for all crops

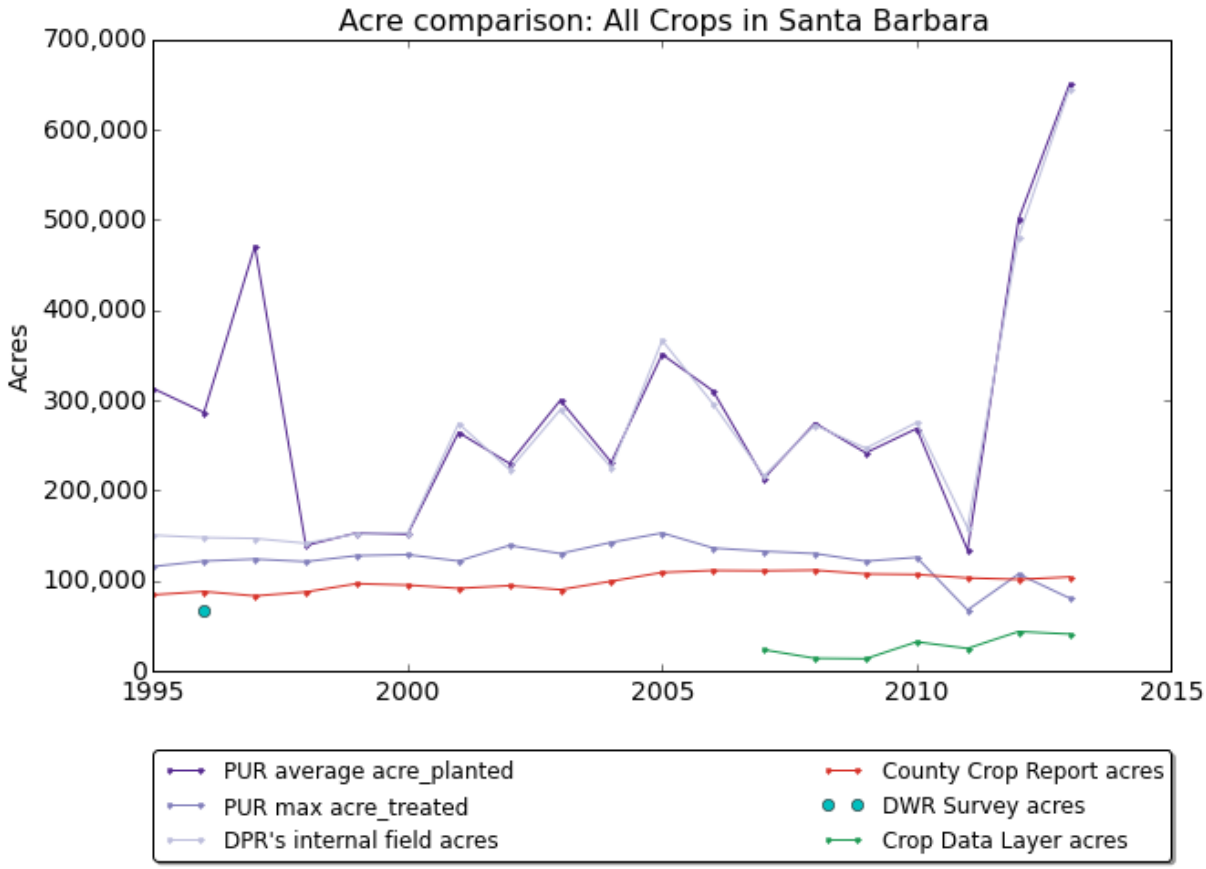

Figure 10: Acre totals for Ventura County for all crops

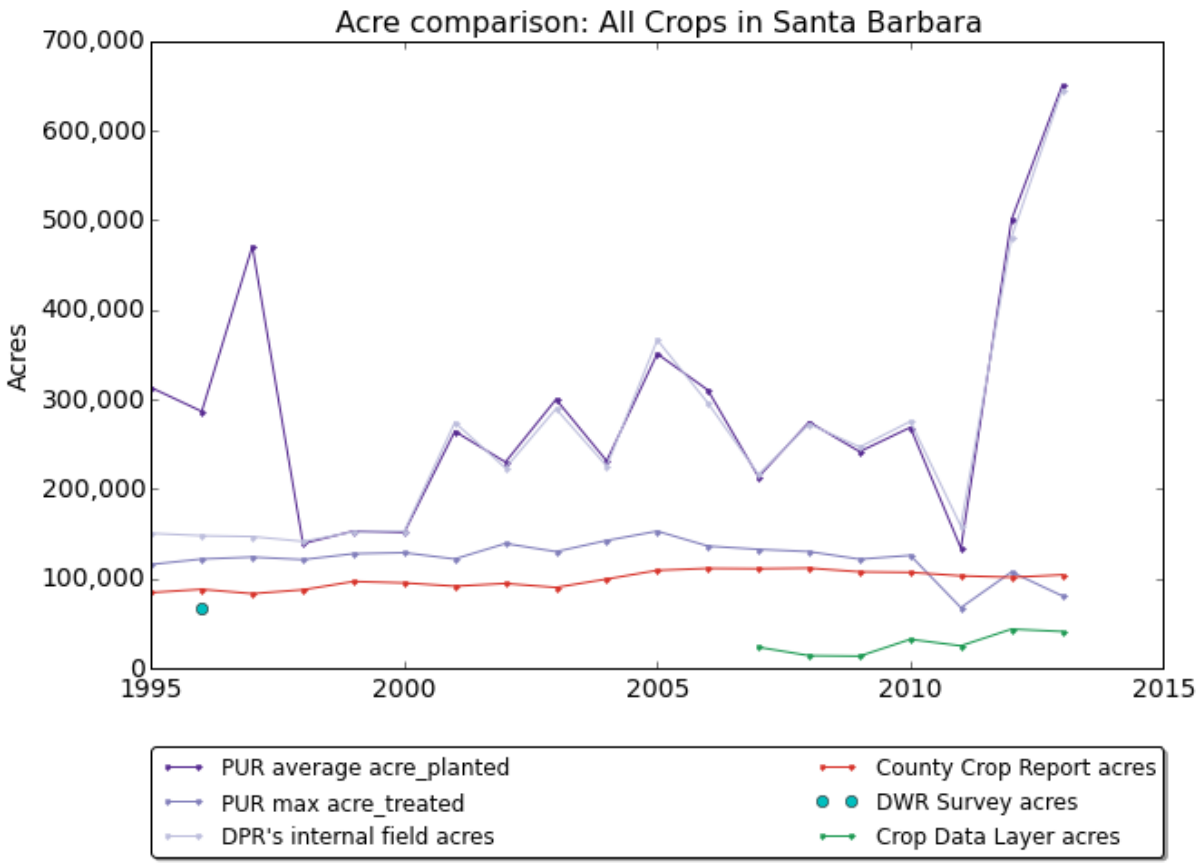

Figure 11: Statewide Sudan grass acreage

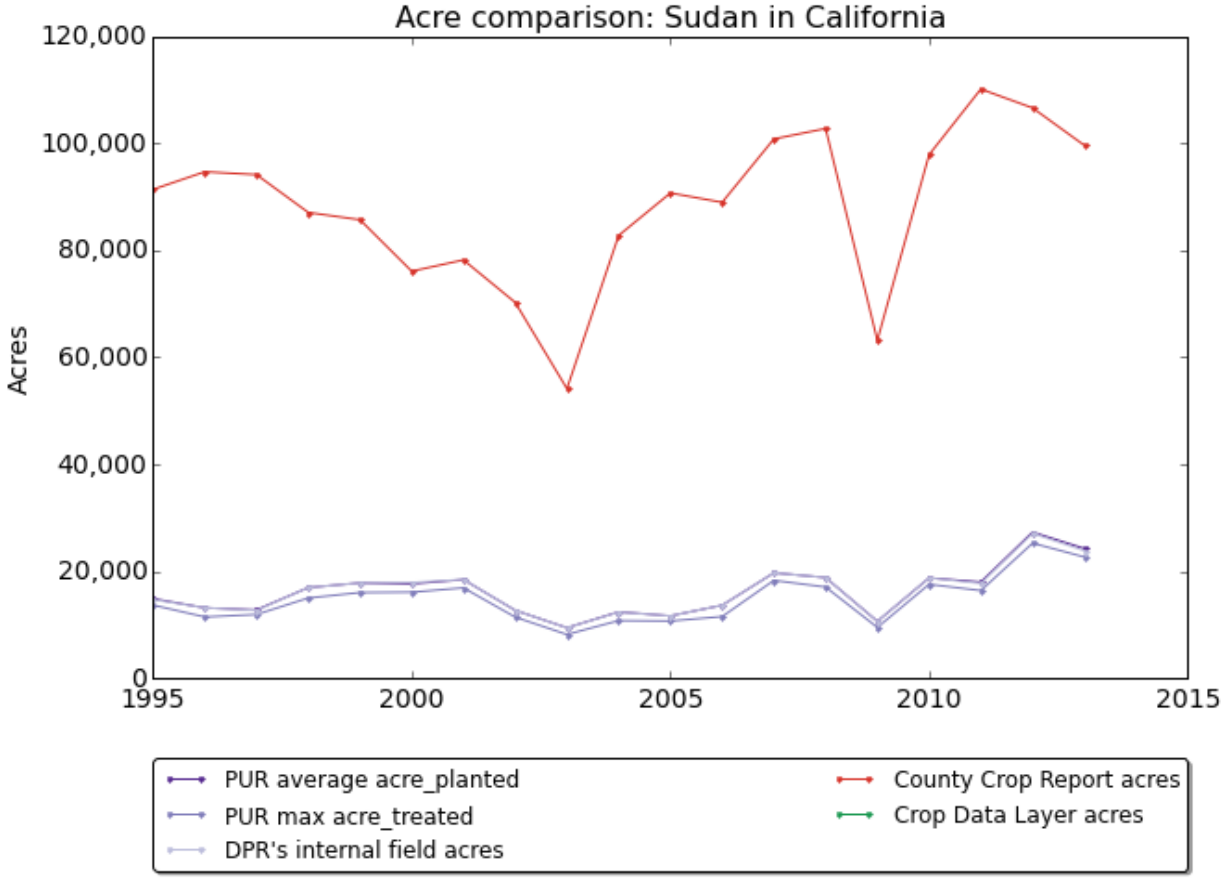

Figure 12: Statewide Avocado acreage

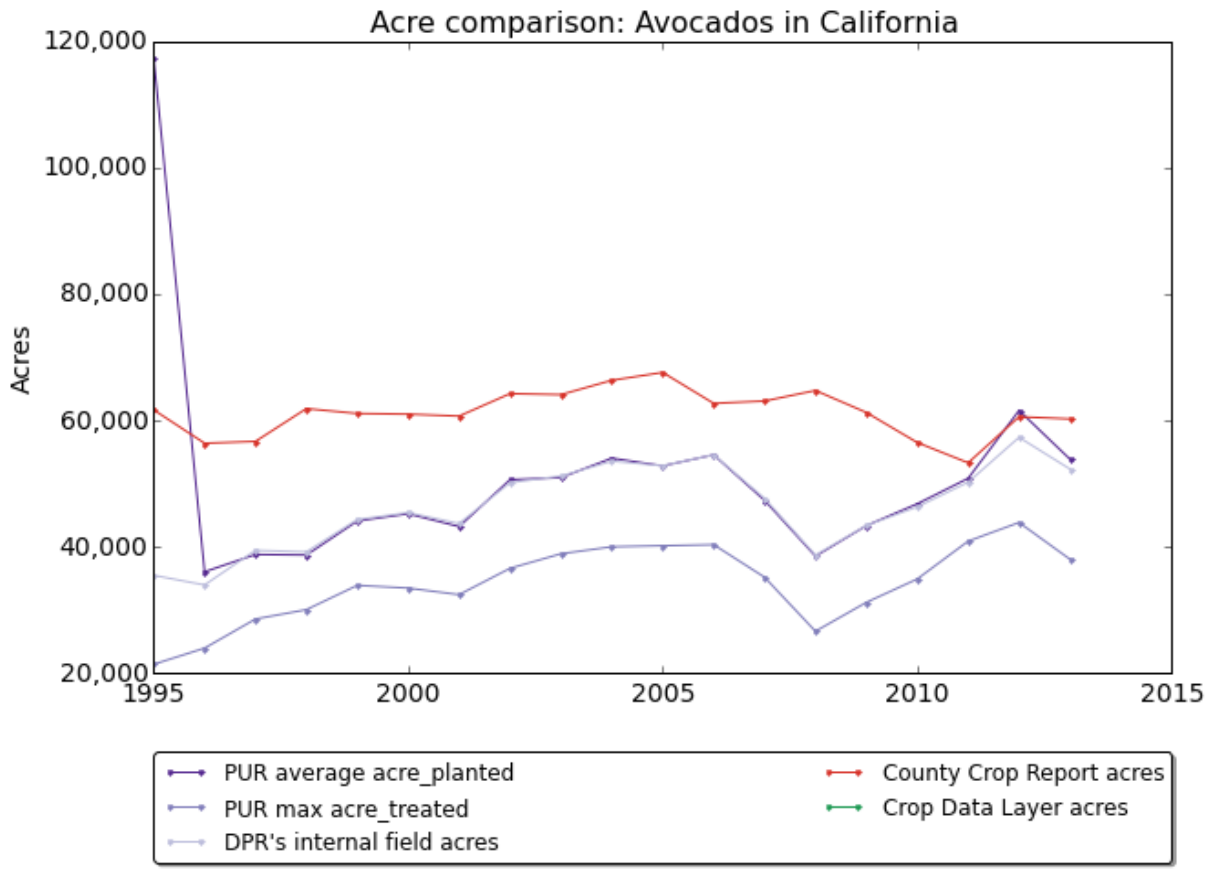

Figure 13: Statewide deciduous fruits & nuts (D) acreage

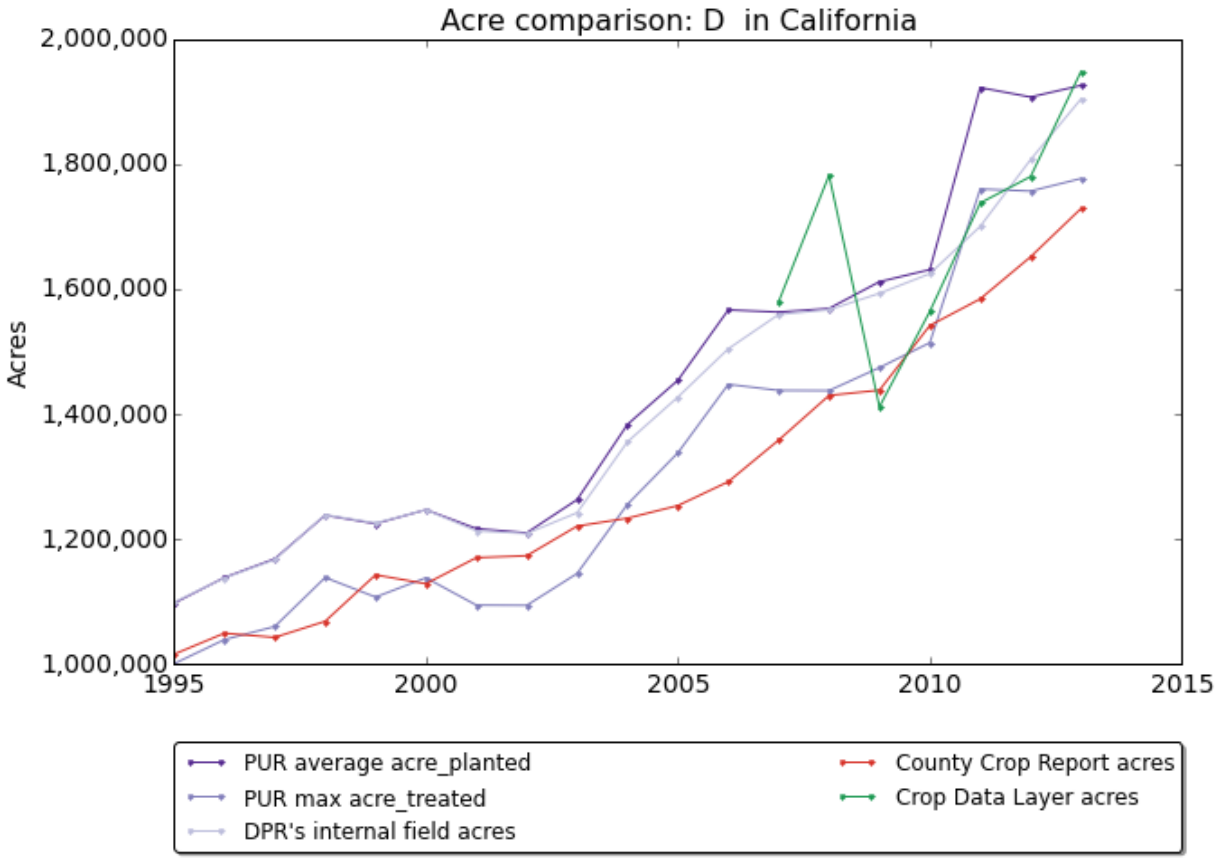

Figure 14: Statewide field crop (F) acreage

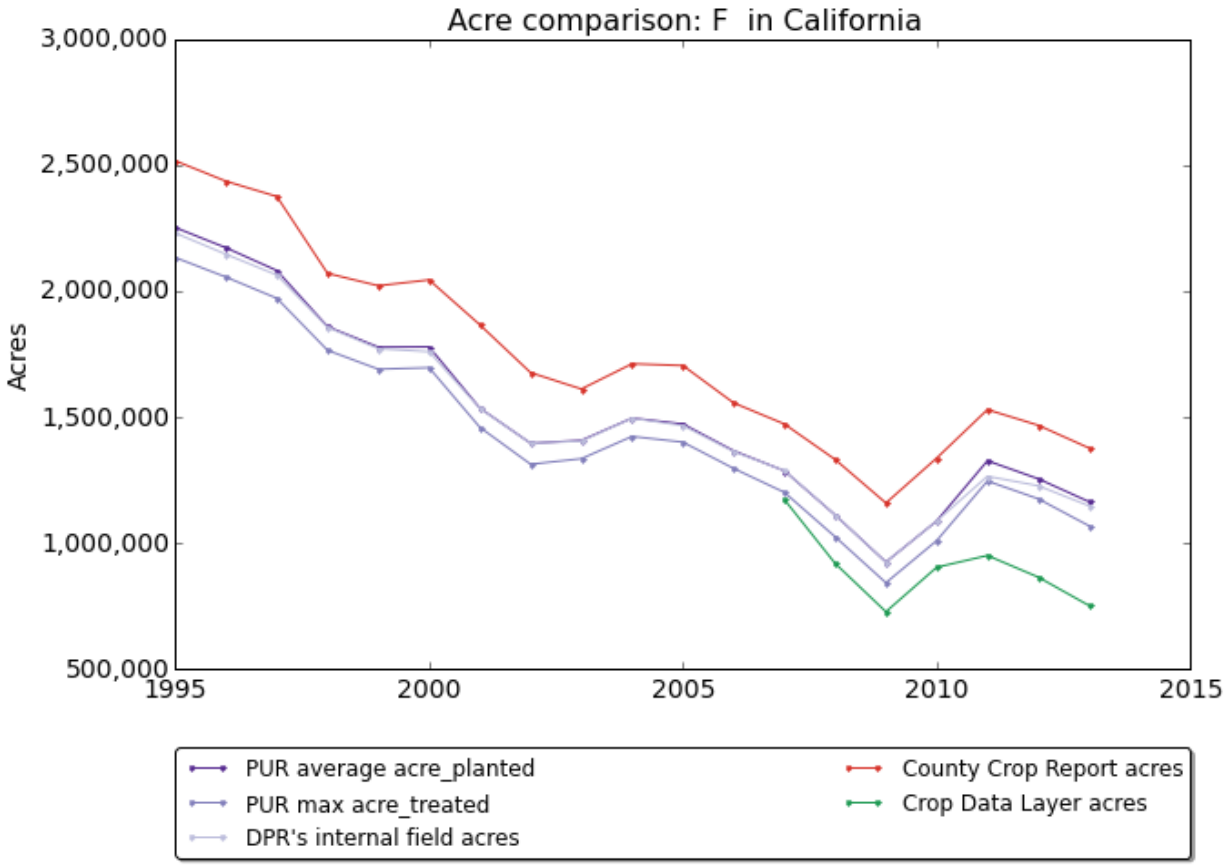

Figure 15: Statewide rice (R) acreage

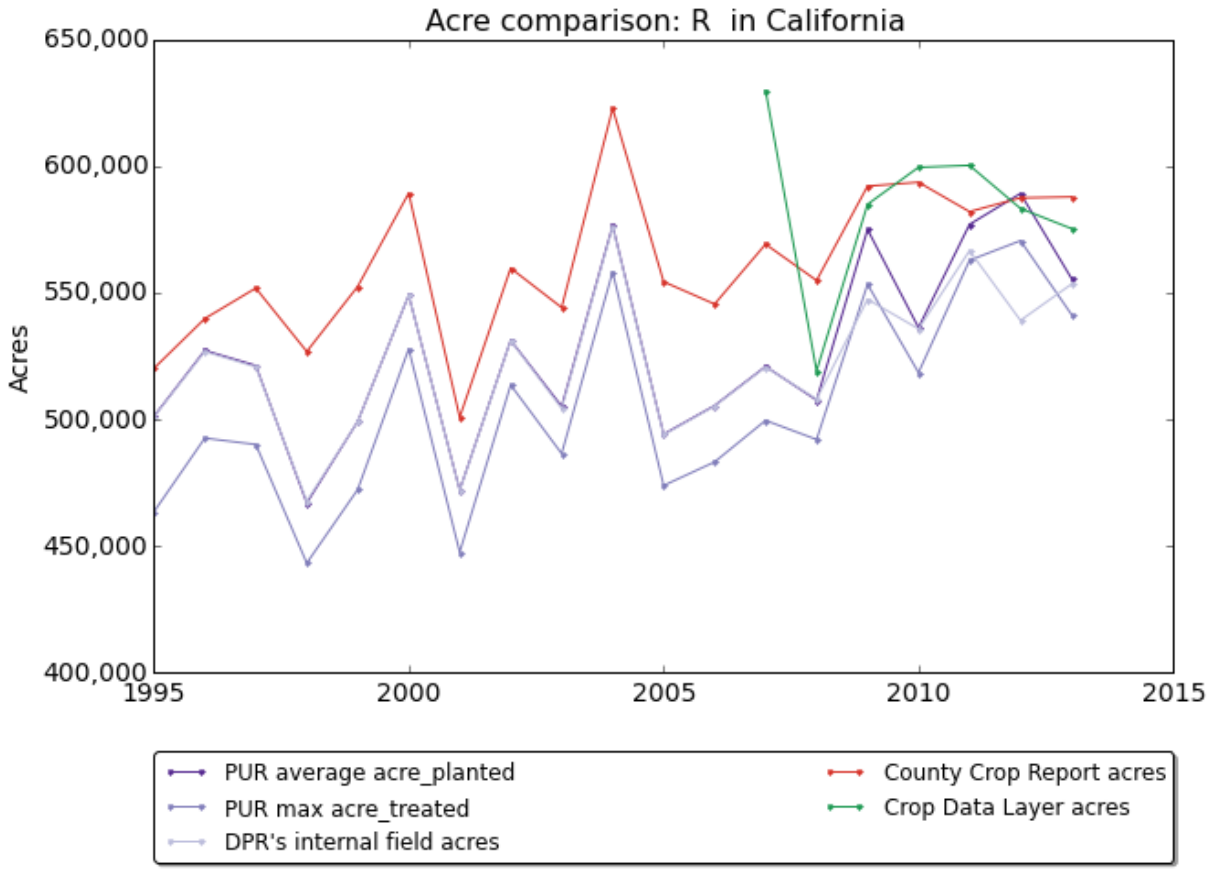

Figure 16: Statewide grain and hay crop (G) acreage

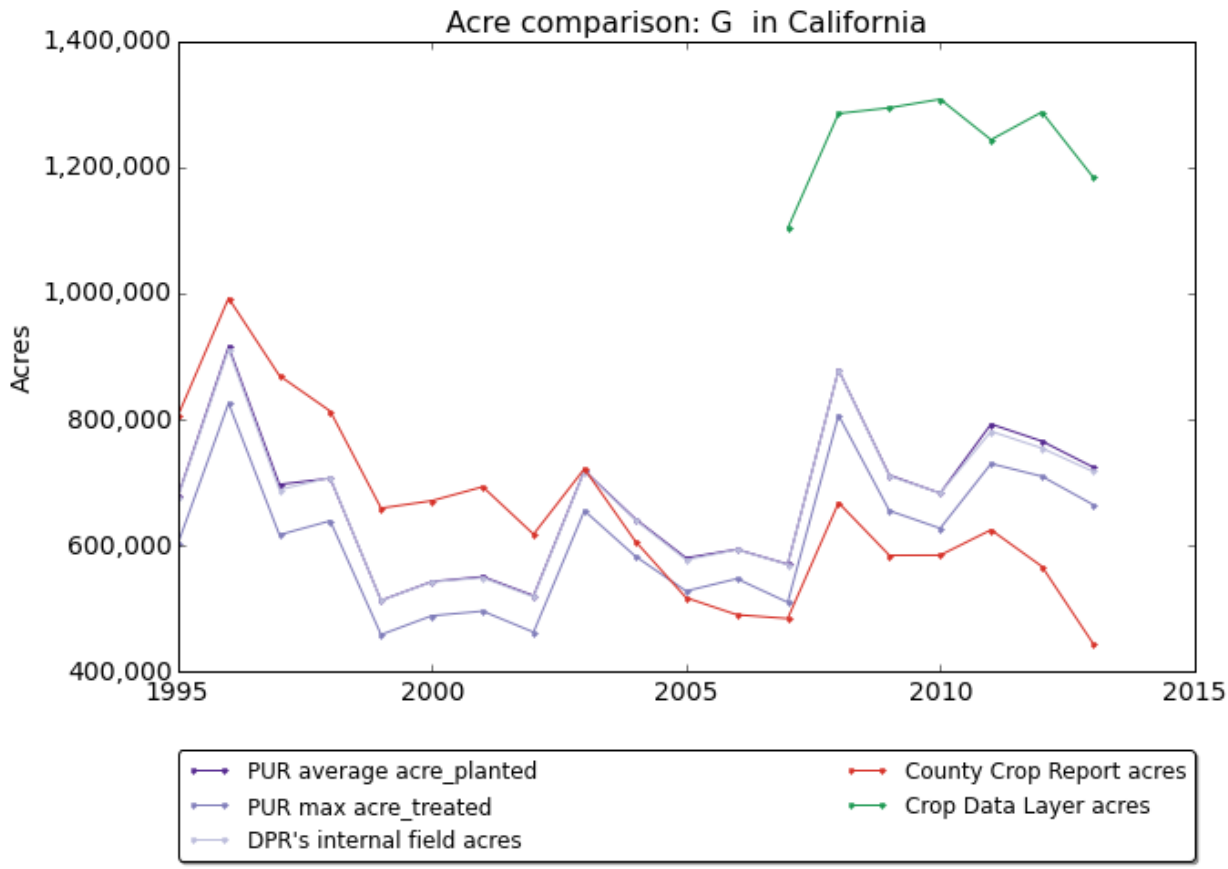

Figure 17: Statewide truck crop (T) acreage

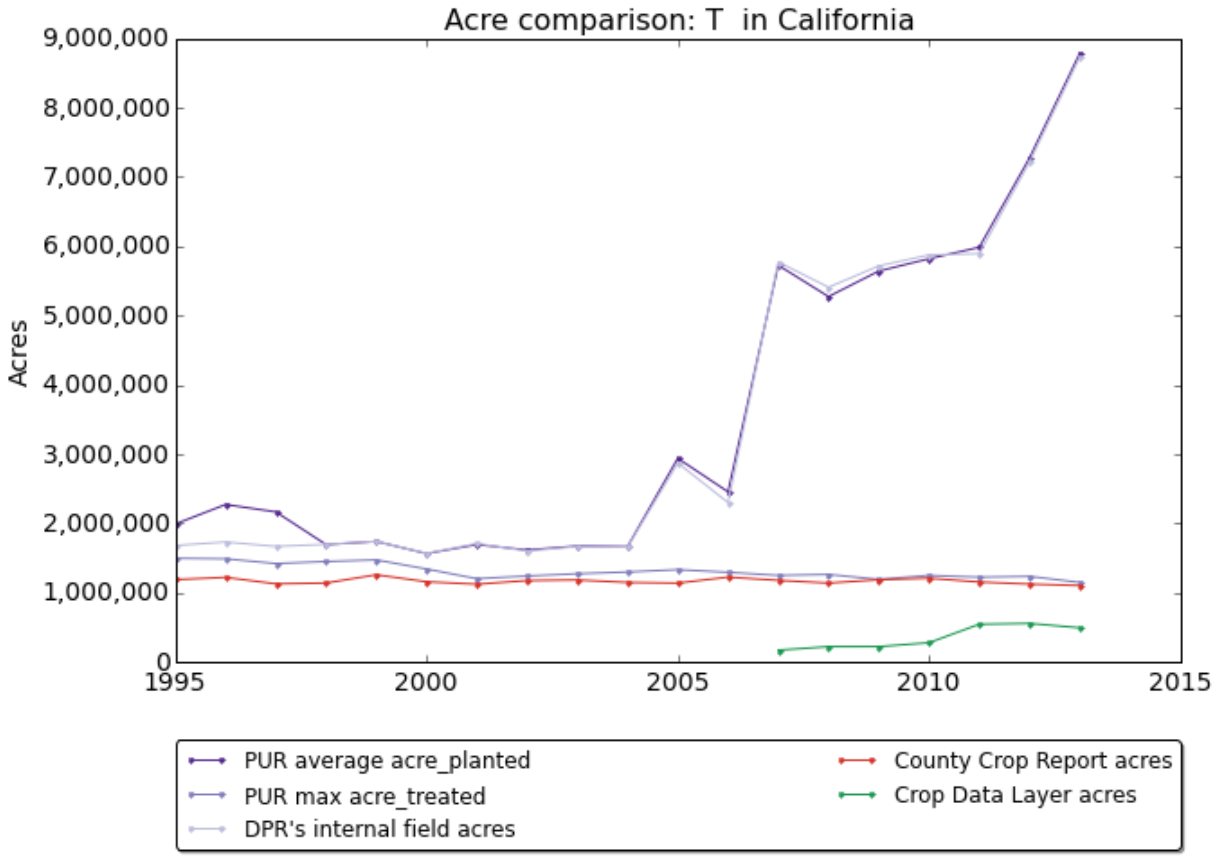

Figure 18: Statewide citrus & subtropical (C) acreage

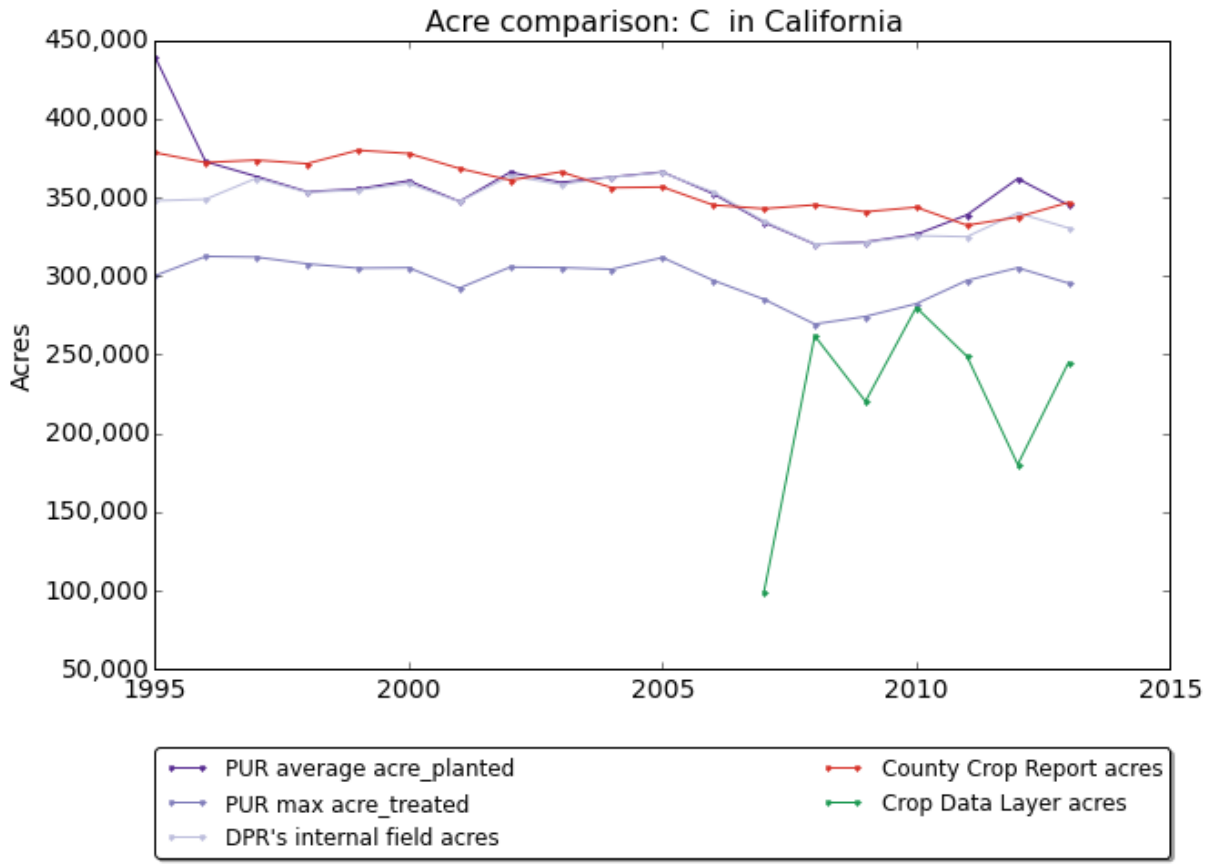

Figure 19: Statewide vineyard (V) acreage

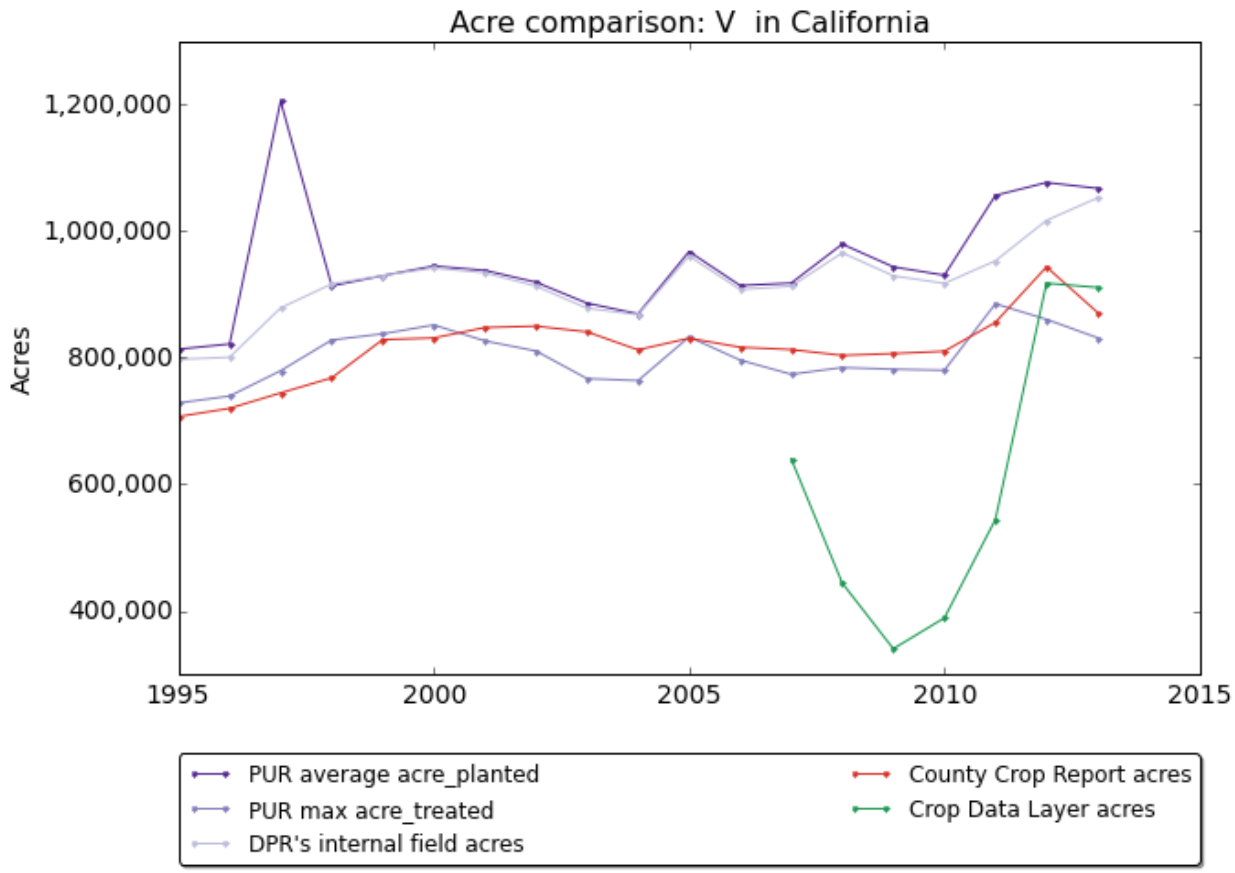

Figure 20: Santa Barbara truck (T) crop acreage

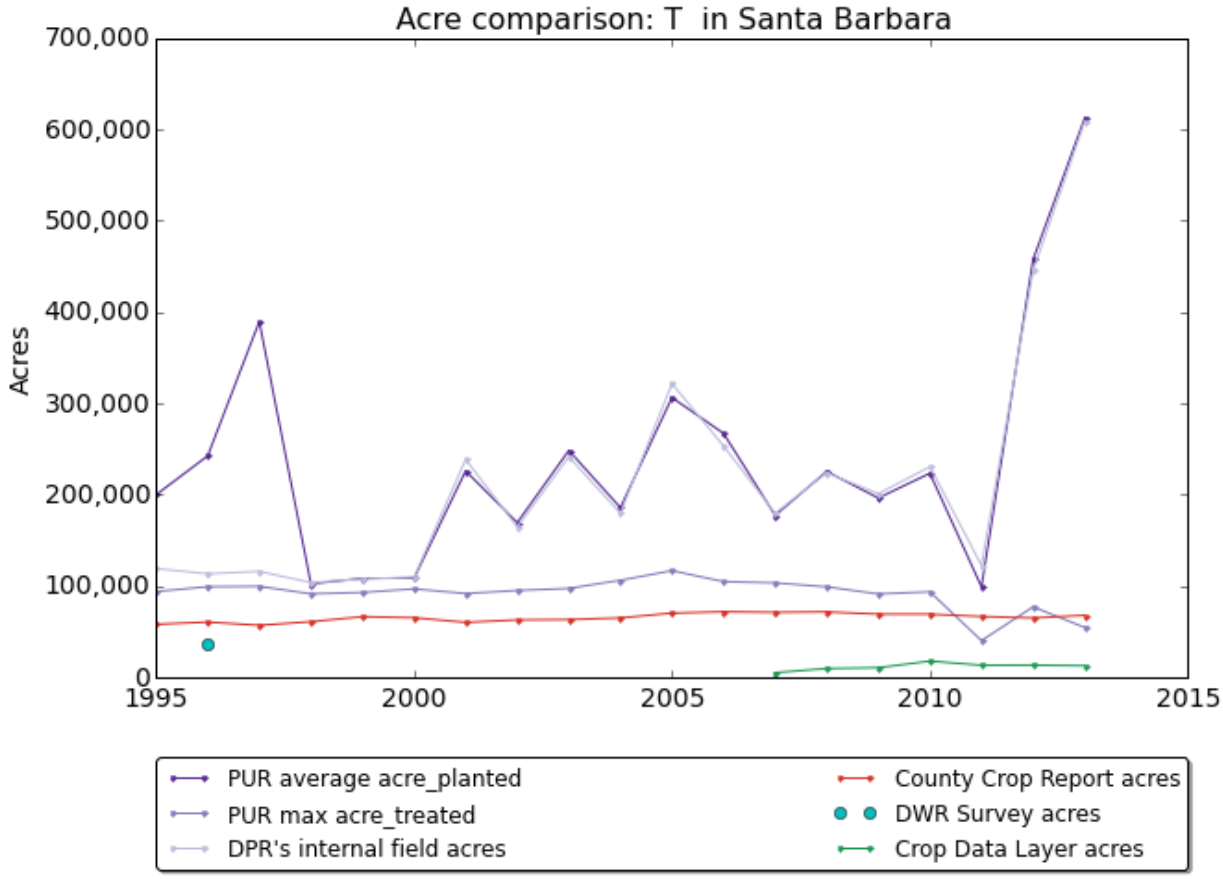

Figure 21: Fresno truck crop (T) acreage

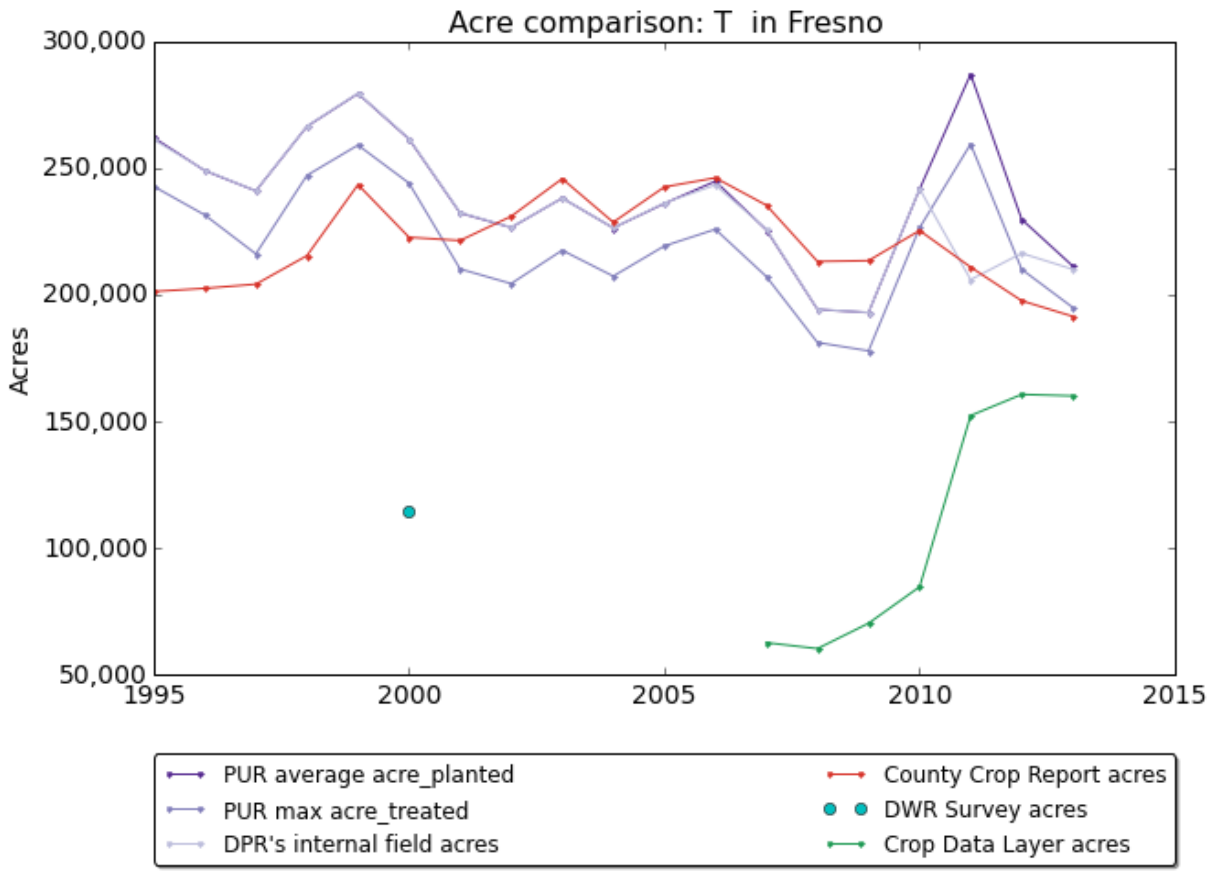

Table 2: Largest and smallest slopes of (PUR acre planted - PUR acre treated) residuals

Commodity	Slope of Residual
All Crops	371356
T	370278
Lettuce (All Types)	193827
Broccoli	66439
Spinach	47722
Cauliflower	22659
Celery	16923
Cabbage	7295
Peas	6157
Carrots	5124
…	…
Safflower	-306
Lemons	-331
Tomatoes (Processing)	-382
Corn (Field & Sweet)	-834
R	-949
Rice	-949
Wheat	-1074
F	-1436
Cotton	-1630
Flowers, Nursery & Christmas Tree Farms	-12632

Table 3: Slope of the residuals for all counties across all crops

County	Slope of Residual
Monterey	378231
Santa Barbara	8645
Kern	469
Stanislaus	360
San Benito	307
San Diego	205
Santa Cruz	114
Placer	90
Glenn	86
Siskiyou	40
Lake	33
Humboldt	14
Alameda	13
Del Norte	12
Nevada	0
Tuolumne	0
Tulare	-2
Marin	-5
Amador	-7
Yolo	-23
Modoc	-23
El Dorado	-26
Lassen	-30
Madera	-45
Mariposa	-58
San Bernardino	-64
Sacramento	-82
Tehama	-104
Solano	-150
Orange	-171
Sutter	-174
Los Angeles	-193
Contra Costa	-244
Santa Clara	-253
Butte	-356

Imperial	-371
Riverside	-665
Kings	-976
San Mateo	-1072
San Joaquin	-1113
Sonoma	-2110
Ventura	-5707

References

1. Pesticide Use Reporting. at <http://www.cdpr.ca.gov/docs/pur/purmain.htm>
2. USDA, N. The 2009 Cropland Data Layer, November 2010. *Wash. DC US Dep. Agric. Natl. Agric. Stat. Serv.* (2010).
3. Mueller, R., Boryan, C. & Seffrin, R. Data partnership synergy: The Cropland Data Layer. in *2009 17th International Conference on Geoinformatics* 1–6 (2009). doi:10.1109/GEOINFORMATICS.2009.5293489
4. Boryan, C., Yang, Z., Mueller, R. & Craig, M. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto Int.* **26,** 341–358 (2011).
5. NASS - California Publications and Press Releases. at <http://www.nass.usda.gov/Statistics_by_State/California/Publications/AgComm/Detail/>
6. California Land & Water Use. at <http://www.water.ca.gov/landwateruse/lusrvymain.cfm>
7. AGIS:Publications:Journals. at <http://agis.ucdavis.edu/publications/journals.html>
8. California Pesticide Information Portal. at <http://calpip.cdpr.ca.gov/main.cfm>
9. Luman, D. & Tweddale, T. Assessment and potential of the 2007 USDA-NASS Cropland data layer for statewide annual land cover applications. *Ill. Nat. Hist. Surv. Tech. Rep.* **49,** (2008).
10. Yang, Y., Wilson, L. T. & Wang, J. Reconciling field size distributions of the US NASS (National Agricultural Statistics Service) cropland data. *Comput. Electron. Agric.* **109,** 232–246 (2014).
11. Banerjee, A., Dolado, J. J., Galbraith, J. W. & Hendry, D. *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. (Oxford University Press, 1993). at <http://www.oxfordscholarship.com/view/10.1093/0198288107.001.0001/acprof-9780198288107>

12. Said, S. E. & Dickey, D. A. Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika* **71,** 599–607 (1984).